Can Patients with Dementia Be Identified in Primary Care Electronic Medical Records Using Natural Language Processing?

Laura C. Maclagan[1] · Mohamed Abdalla[2] · Daniel A. Harris[3] ·
Therese A. Stukel[1,4,5] · Branson Chen[1] · Elisa Candido[1] · Richard H. Swartz[1,5,6] ·
Andrea Iaboni[7,8] · R. Liisa Jaakkimainen[1,4,5,9] · Susan E. Bronskill[1,3,4,5,10]

Abstract

Dementia and mild cognitive impairment can be underrecognized in primary care practice and research. Free-text fields in electronic medical records (EMRs) are a rich source of information which might support increased detection and enable a better understanding of populations at risk of dementia. We used natural language processing (NLP) to identify dementia-related features in EMRs and compared the performance of supervised machine learning models to classify patients with dementia. We assembled a cohort of primary care patients aged 66 + years in Ontario, Canada, from EMR notes collected until December 2016: 526 with dementia and 44,148 without dementia. We identified dementia-related features by applying published lists, clinician input, and NLP with word embeddings to free-text progress and consult notes and organized features into thematic groups. Using machine learning models, we compared the performance of features to detect dementia, overall and during time periods relative to dementia case ascertainment in health administrative databases. Over 900 dementia-related features were identified and grouped into eight themes (including symptoms, social, function, cognition). Using notes from all time periods, LASSO had the best performance (F1 score: 77.2%, sensitivity: 71.5%, specificity: 99.8%). Model performance was poor when notes written before case ascertainment were included (F1 score: 14.4%, sensitivity: 8.3%, specificity 99.9%) but improved as later notes were added. While similar models may eventually improve recognition of cognitive issues and dementia in primary care EMRs, our findings suggest that further research is needed to identify which additional EMR components might be useful to promote early detection of dementia.

Extended author information available on the last page of the article

1 Introduction

Dementia is characterized by a broad set of symptoms related to memory loss, language difficulties, diminished executive function, and other behavioral and functional impairments that are beyond what is considered normal aging [1]. Studies have documented significant increases in the prevalence of dementia over the past several decades[1, 2], and this condition is associated with substantial health care costs and caregiver burden [1]. The early detection and treatment of dementia may aid in clinical management, enable interventions, and help individuals plan for future care needs [3–7]. Previous work has shown that dementia and mild cognitive impairment are often underrecognized in primary care [8, 9], with between 40 and 80% of dementia cases going undetected [7, 10–12]. Limited knowledge and time, diagnostic uncertainty, and stigma are considered barriers to diagnosing dementia in primary care settings [6, 7, 13]. Machine learning methods provide a promising avenue to increase the detection of dementia in primary care, potentially enabling a better understanding of populations at risk of dementia.

Prior studies have used machine learning methods as a strategy to improve dementia detection for clinicians and researchers [14–16], often leveraging neuropsychological assessments, brain imaging data, genetic profiles, and clinical and lifestyle factors as their primary inputs [14–18]. However, these data may not be available in all research contexts, can be costly to obtain, and are time-consuming to collect. The wealth of information stored in electronic medical records (EMRs) provides opportunities to enable the early detection of signs and symptoms of dementia, to assist primary care providers in recommending individuals for further diagnostic testing and treatment [19], and to support capacity planning for the needs of persons living with dementia. EMRs combined with natural language processing (NLP) techniques present an opportunity to utilize routinely collected clinical notes, which contain rich longitudinal histories on patients, as an emerging and efficient solution [20, 21].

Prior studies have used NLP, free-text clinical notes, and machine learning models to describe, predict, and/or classify geriatric syndromes [20, 22], frailty [23], multiple sclerosis [24], severe mental illness [25], dementia [26–28], and other chronic diseases [29]. Among the studies that have used NLP for dementia risk prediction, some have focused on more narrowly defined features of the disease (e.g., cognitive symptoms) [28], while others have combined patients across a number of diverse clinical settings (where substantial patient heterogeneity is to be expected) [27]. Although most patients with dementia receive care in the primary care setting, limited research has used NLP and machine learning in this patient population.

Given the need to develop tools to enhance dementia detection in primary care, our objectives were to (1) use NLP to identify features commonly associated with dementia in primary care EMR clinical notes and (2) evaluate the performance of machine learning models to classify patients with dementia based on these features.

2 Materials and Methods

2.1 Data Sources

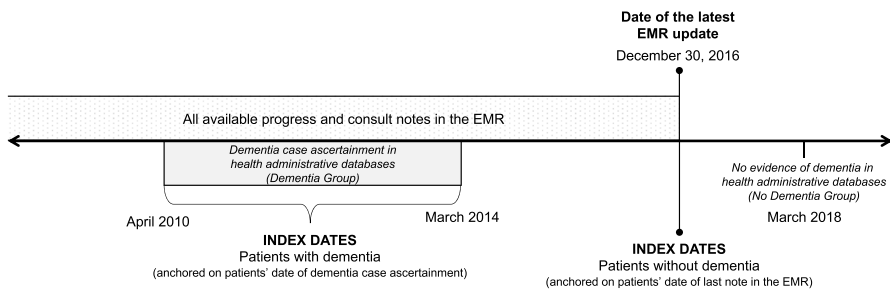#### 2.1.1 Electronic Medical Records

We used the Electronic Medical Records Primary Care (EMRPC) database which includes a convenience sample of community-based primary care physicians in Ontario, Canada [30]. The EMRPC database links health administrative data with EMR data from over 350 family physicians who care for approximately 400,000 patients. The demographic composition of this primary care population is representative of the broader primary care patient population of Ontario [31]. EMRPC includes primary care records extracted until December 30, 2016. Different types of free-text clinical notes are available in EMRPC: progress notes, which are recorded by the primary care physician at each visit and include detailed information about the concerns and issues discussed; and consult notes, which contain information from specialist consultations. Disease reference standards have also been abstracted from EMRPC and used to validate health administrative data algorithms for various medical conditions, including ischemic heart disease and Alzheimer's disease and related dementias [32–34]. For this study, EMRPC was used to identify the study cohort and abstract dementia-related features in free-text clinical notes using NLP.

2.1.2 Health Administrative Data

Due to the single-payer health insurance model in Ontario, population-based health system encounters result in health administrative data that can be linked across a variety of settings. We used a validated health administrative data algorithm based on these data as the reference standard for dementia case ascertainment [34]. Individuals with ≥ 1 hospitalization and/or ≥ 3 physician visits for dementia within 2 years (with a 30-day gap between each) and/or a prescription for a cholinesterase inhibitor were classified as having dementia. The algorithm was validated against clinical review of primary care EMR charts and was found to have excellent performance characteristics (99.1% specificity and 79.3% sensitivity). These databases were also used to describe the sociodemographic, clinical, and health service use characteristics of the cohort (see Supplementary Table 1 for a full list of data sources). These databases were linked using unique encoded identifiers and analyzed at ICES. ICES is an independent, non-profit research institute whose legal status under Ontario's health information privacy law allows it to collect and analyze health care and demographic data, without consent, for health system evaluation and improvement. The use of the data in this project is authorized under section 45 of Ontario's Personal Health Information Protection Act (PHIPA) and does not require review by a Research Ethics Board.

2.2 Study Cohort Population and Dementia Criteria

We identified patients aged 66 years and older in the EMRPC database who were formally enrolled with primary care physicians in Ontario, Canada. Patients aged 66 years and older were included in order to have complete drug information following enrollment in the public drug plan at age 65. This population was linked to health administrative databases to obtain information on dementia over time and other characteristics. Patients were identified as having dementia if they had a case ascertainment date from the health administrative algorithm between April 2010 and March 2014 to allow for a minimum of 2 years of notes in the EMR following dementia case ascertainment. Patients were identified as not having dementia if they had not met criteria for the health administrative algorithm by March 2018. This ensured that patients without dementia were not in the prodromal stages of the disease. The study index date was defined as the date of case ascertainment for patients with dementia and the date of the latest note in the EMR for patients without dementia (with the latest possible index date being December 30, 2016). Covariates were assessed as of the study index date for patients with dementia and as of the latest data update in the EMR for patients without dementia. See Fig. 1 for a schematic of the study design. To be eligible for the study cohort, patients were required to have at least two visits to a primary care physician recorded in the EMR in the 2 years prior to their study index date (to ensure they would have sufficient information in the EMR to enable detection of dementia signs and symptoms).

Fig. 1 Methodological schematic for the development of machine learning models to classify patients with dementia in primary care electronic medical records in Ontario, Canada. All electronic medical record (EMR) progress and consult notes were obtained from participating primary care practices in Ontario, Canada, between April 1, 2016, and December 30, 2016 (i.e., date of latest EMR update). Linked health administrative databases were used to identify patients with dementia between April 2010 and March 2014. All patients who met criteria for dementia during this period, based on a validated health administrative data algorithm, where included in the dementia group. The date patients met full algorithmic criteria for dementia represented their study index date. Patients without a history of dementia in the health administrative data as of March 2018 were included as the non-dementia comparison group. The study index date for patients without dementia was the date of their last note in the EMR. As such, the latest possible study index date for patients without dementia was December 30, 2016; however, it could also occur any time prior

2.3 Patient Characteristics

Patient characteristics included age, sex, neighborhood income quintile, residence in a long-term care facility (i.e., nursing home), and chronic conditions. Neighbourhood income quintile measures the relative income of neighbourhoods and was assigned using Statistics Canada's Postal Code Conversion File Plus (PCCF + Version 7A) [35]. Health administrative data algorithms were used to identify whether patients had one or more of 17 chronic conditions (see Supplementary Table 2 for a full list of conditions and diagnostic codes) [36, 37]. For the year prior to study index, we identified use of physician services including family physician visits and visits to dementia specialists (i.e., neurologists, geriatricians, and psychiatrists) and dispensed medications commonly associated with dementia (i.e., cholinesterase inhibitors, antipsychotics, antidepressants, and benzodiazepines).

2.4 Natural Language Processing Approach to Develop an EMR-Based Feature List for Dementia

An initial set of features representing the signs and symptoms of dementia was created from published lists [38–40] and expert clinical input (neurologist (RS), family physician (LJ), and geriatric psychiatrist (AI)). Features included dementia symptoms, measures of psychosocial well-being, and other aspects of dementia presentation and care such as types of family caregivers and social supports. For both the feature list and the clinical notes in the EMR, we used pre-processing steps to improve standardization and matching of the feature list and free-text. Please see Technical Appendix 1 for details. For each feature, the 5, 10, and 15 nearest words were identified in a word embedding space [41] trained on all available progress and consult notes based on cosine distance to measure distance between words [42]. This expansion of the feature list was intended to identify misspellings and relevant, but potentially missed, words by capturing closely related words in vector representations. The 10 nearest words were chosen to maximize the number of related words while ensuring clinical relevance of words added. Previous work has shown that word embeddings trained on clinical notes can find more semantically similar words than methods using pre-trained word vectors and that this results in identification of words closer to expert human judgement and improved performance on most tasks [43]. As a final step, we applied stemming. In order to summarize the features and investigate which showed the best performance in classifying patients with dementia, we manually categorized features into eight thematic groups based on their relationship to dementia care (symptoms, social, function, cognition, dementia medication, other medication, health system use, and other). LM categorized words according to accepted clinical and diagnostic criteria and descriptions of cognitive and behavioral presentation in dementia. The final categorizations were reviewed by members of the author team with relevant clinical expertise (RS, LJ, and AI).

2.5 Machine Learning Models to Classify Patients with Dementia Based on EMR Features

We compared the performance of several supervised machine learning models to classify patients with dementia using the EMR-based feature list. These included penalized logistic regression with both LASSO and ridge regularization, gradient boosted decision trees, and multi-layer perceptron models (MLP classifier, i.e., neural network). Complete methodological details regarding hyperparameter tuning and cross-validation can be found in Technical Appendix 1 and Supplementary Table 3. In order to compare the performance of the classifiers described above to a baseline approach, we also implemented a simple bag-of-words model. A term frequency inverse document frequency (TF-IDF) vectorizer was used to convert text (from the clinical notes and pre-processing steps described above) into vectors which were then passed into basic logistic regression model.

For the primary analysis, model performance was assessed using all available clinical notes and then consult and progress notes separately. To optimize their clinical utility, we trained the machine learning models to maximize the F1 score (defined as the harmonic mean of positive predictive value (PPV) and sensitivity, with values ranging from 0 to 1 (indicating perfect PPV and sensitivity)) because we aimed to prioritize sensitivity and PPV (i.e., minimizing false negatives and false positives). The F1 score can also be a helpful indicator of model performance when classes show size imbalance (i.e., much larger number of patients without dementia than patients with dementia in the present study). Predictions from the trained models were applied to the test dataset, where patients predicted to have a greater than 50% probability of dementia were classified as a potential case. Classification performance (e.g., sensitivity, specificity, PPV, negative predictive value (NPV), F1 score) and calibration (i.e., integrated calibration index (ICI), E50, E90, and EMax) were averaged across five test-folds. E50 and E90 represent the median and 90th percentile of the difference between observed and predicted probabilities, while Emax represents the maximum difference [44]. The model(s) with the highest average F1 score on the test dataset was considered to perform the best.

To compare feature performance across thematic groups, the best performing model from the primary analysis was tested using features belonging to each thematic group separately. To assess how the timing of notes was related to model performance, we used the best performing model from the primary analysis and clinical notes written during time periods before (from index to 7, 14, 30, 60, 120, 365, and 730 days prior) and after (from the first note in the EMR to 0, 7, 14, 30, 60, 120, 365, and 730 days after index) patients' index dates (Supplementary Fig. 1). For patients without dementia, all notes were included when assessing notes after index since the index date was the date of the last note in the EMR. Analyses were conducted in Python version 3.6 (including model functions from the sci-kit learn library) and SAS version 9.4.

3 Results

3.1 Patient Characteristics

We identified 526 patients with dementia and 44,148 patients without dementia in the EMRPC database during the study period who met the study inclusion criteria (Table 1). Although patients with dementia were older (80.3 years vs. 74.6 years) compared to those without dementia, sex and rural residence were similar across the two groups. Patients with dementia were more likely to have five or more comorbidities (11.6% vs. 7.8%) and to be prescribed ten or more medications (7.6% vs. 3.8%).

The study data included more than 5.6 million clinical progress and consult notes (112,053 among patients with dementia and 5,551,274 among patients without dementia) (Supplementary Table 4). Patients had a mean of 51.5 (standard deviation [SD] = 41.3) consult and 75.3 (SD = 61.0) progress notes. Patients with dementia had higher numbers of consult and progress notes (dementia: consult mean = 82.9 [SD = 53.4]; progress mean = 130.2 [SD = 86.5]; without dementia: consult mean = 51.2 [SD = 41.0]; progress mean 74.7 [SD = 60.4]).

3.2 List of Dementia-Related Features in EMR Clinical Progress and Consult Notes

The final list of features included 910 words associated with dementia and dementia symptoms. All words, by theme, are included in Supplementary Table 5. The list included a broad set of features related to dementia symptoms and care including dementia and other medications, social issues, and health service use. The largest numbers of features were categorized as cognition, medication-other (i.e., medications not specifically indicated for dementia but which may be used in patients with dementia), and symptoms (including behavioural and physical symptoms).

3.3 Model Performance for Classifying Patients with Dementia

Compared to the simple bag of words of approach, our pre-defined feature list showed better performance and was therefore used throughout. For example, the average F1 score from the bag of words LASSO model was 0.14 compared to an F1 score of 0.71 from the same model using our pre-defined feature list. Models using all consult and progress notes in the EMR showed superior performance compared to models using consult or progress notes alone (Table 2). LASSO had the highest F1 score of 77% (sensitivity = 71%, specificity = 100%, PPV = 85%, NPV = 100%). LASSO using progress notes alone showed modestly lower performance. Models using only consult notes performed poorly, with most F1 scores being below 50%. Neural networks were generally the poorest performing models. Based on the ICI, the LASSO models showed the best calibration.

Table 1 Baseline characteristics of patients aged 66 years and older included in a primary care electronic medical record database in Ontario, Canada, between 2010 and 2016, by dementia status (N=44,674)

Characteristic, n (%) unless otherwise noted	Dementia[a,b] (n=526)	Without dementia[b] (n=44,148)
Age		
Mean (SD)	80.3 (6.7)	74.6 (7.0)
66–74	106 (20.2%)	25,479 (57.7%)
75–84	268 (51.0%)	13,910 (31.5%)
85+	152 (28.9%)	4759 (10.8%)
Female sex	309 (58.7%)	24,424 (55.3%)
Rural residence	139 (26.4%)	10,647 (24.1%)
Income quintile		
1 (lowest)	106 (20.2%)	7893 (17.9%)
2	96 (18.3%)	8906 (20.2%)
3	82 (15.6%)	8257 (18.7%)
4	116 (22.1%)	7965 (18.0%)
5 (highest)	124 (23.6%)	11,079 (25.1%)
Long-term care resident	12 (2.3%)	225 (0.5%)
Number of chronic conditions (excluding dementia)		
Mean (SD)	2.2 (1.9)	1.8 (1.8)
0–1	205 (39.0%)	21,740 (49.2%)
2	93 (17.7%)	7861 (17.8%)
3	105 (20.0%)	7076 (16.0%)
4	62 (11.8%)	4021 (9.1%)
5+	61 (11.6%)	3450 (7.8%)
Chronic conditions		
Hypertension	283 (53.8%)	18,441 (41.8%)
Cancer	135 (25.7%)	11,549 (26.2%)
Osteoarthritis	116 (22.1%)	9669 (21.9%)
Mood/anxiety disorder	108 (20.5%)	3916 (8.9%)
Diabetes	99 (18.8%)	7013 (15.9%)
Coronary artery disease	51 (9.7%)	3824 (8.7%)
Asthma	49 (9.3%)	3611 (8.2%)
Congestive heart failure	40 (7.6%)	3013 (6.8%)
Chronic obstructive pulmonary disease	35 (6.7%)	2849 (6.5%)
One or more physician visits (1 year prior)		
Family physician visit	523 (99.4%)	40,937 (92.7%)
Dementia specialist visit	130 (24.7%)	3454 (7.8%)
Neurologist visit	46 (8.7%)	2130 (4.8%)
Geriatrician visit	43 (8.2%)	302 (0.7%)
Psychiatrist visit	52 (9.9%)	1261 (2.9%)
Drug therapies[c] dispensed (1 year prior)		
Mean (SD)	4.96 (2.96)	4.05 (2.72)
0–5	347 (66.0%)	35,005 (79.3%)

Table 1 (continued)

Characteristic, *n* (%) unless otherwise noted	Dementia[a,b] (*n*=526)	Without dementia[b] (*n*=44,148)
6–9	139 (26.4%)	7481 (16.9%)
10+	40 (7.6%)	1662 (3.8%)
Antidepressants	127 (24.1%)	5281 (12.0%)
Antipsychotics	27 (5.1%)	514 (1.2%)
Benzodiazepines	44 (8.4%)	2193 (5.0%)
Cholinesterase inhibitors	153 (29.1%)	0 (0.0%)

N Sample size, *SD* standard deviation

[a]Dementia was defined using a validated health administrative data algorithm (reference standard)

[b]Baseline characteristics shown as of the study index date (date of dementia case ascertainment for patients with dementia was between April 1, 2010, and March 31, 2014) and date of latest update in the EMR for patients without dementia (between April 1, 2016, and December 30, 2016)

[c]By unique drug name

Table 2 Performance characteristics of machine learning models to classify patients with dementia using all primary care electronic medical record notes prior to December 30, 2016 in Ontario, Canada, by note type

Model	F1 Score	Sens	Spec	PPV	NPV	E50	E90	EMax	ICI
Consult and progress notes									
LASSO	77.2	71.5	99.8	85.2	99.7	0.002	0.003	0.3757	0.0049
Ridge	72.6	68.0	99.8	78.8	99.6	0.024	0.04	0.1893	0.0273
Gradient boosted	74.1	68.8	99.8	80.5	99.6	-	-	-	-
Neural network	70.4	65.4	99.8	76.4	99.6	0.0011	0.49	50,271.8	113.6
Consult notes only									
LASSO	44.6	32.9	99.9	73.7	99.2	0.0068	0.008	1.4429	0.0166
Ridge	43.1	44.3	99.3	43.0	99.3	0.0131	0.02	0.4545	0.0194
Gradient boosted	51.1	43.9	99.7	64.2	99.3	-	-	-	-
Neural network	49.4	37.4	99.4	49.4	99.3	0.0052	0.028	2380.8	5.25
Progress notes only									
LASSO	70.6	61.2	99.9	84.3	99.5	0.0036	0.005	0.2738	0.007
Ridge	67.3	64.4	99.7	70.8	99.6	0.0118	0.022	0.2313	0.015
Gradient boosted	68.7	63.3	99.8	75.2	99.6	-	-	-	-
Neural network	66.3	58.8	99.8	76.5	99.5	0.0021	0.093	111,708.1	260.03

Key: PPV positive predictive value; *NPV* negative predictive value; *F1 score* harmonic mean of sensitivity and PPV; *ICI* integrated calibration index; "–" value could not be estimated

3.4 Model Performance for Different Thematic Groups of Features

When comparing performance of the best-performing LASSO model across specific categories of features, medication-related features showed the highest performance based on F1 scores (F1 score dementia medications=73%, F1 score other

medications = 71%), followed by health system use (F1 score = 61%), cognition (F1 score = 59%), and symptoms (F1 score = 43%) features. Function, social, and other features showed poor performance with F1 scores < 30% (Table 3).

3.5 Model Performance by Time Prior To and Following Dementia Case Ascertainment

Model performance was poor (F1 score = 14.4%, sensitivity = 8.3%, specificity = 99.9%, PPV = 65.0%, ICI = 0.01; Table 4) when notes written prior to index were used. Model performance gradually improved with the inclusion of notes written for longer time periods following dementia case ascertainment.

4 Discussion

Using primary care EMR data, we developed a list of over 900 dementia-related features that characterized the signs and symptoms of dementia and included both clinical and social factors, which were grouped into eight themes. We used several machine learning models to classify patients with dementia based solely on the presence of these features in free-text clinical notes in the EMR. The LASSO model achieved good performance in classifying patients with dementia using all available EMR notes, while models solely including notes prior to dementia case ascertainment showed weaker performance. Our examination of model performance by feature theme demonstrated that medication (dementia and other), health system use, and cognition features showed the best performance out of the eight themes examined.

While many studies have reported on models to identify chronic diseases in structured data fields of EMRs (including using International Classification of Diseases codes) [34, 45], relatively few have developed models using EMR free-text fields [28, 29]. Despite EMRs being a longitudinal source of rich patient-level information

Table 3 Performance characteristics of LASSO to classify patients with dementia using all primary care electronic medical record progress and consult notes prior to December 30, 2016 in Ontario, Canada, by feature theme

Category of features	F1 Score	Sens	Spec	PPV	NPV
Cognition	59.0	49.4	99.8	74.4	99.4
Function	28.6	19.6	99.8	56.4	99.1
Health system use	61.2	52.3	99.8	74.9	99.4
Dementia medication	73.2	62.6	99.9	91.2	99.5
Other medication	70.6	58.9	99.9	88.9	99.5
Social	26.3	17.7	99.8	54.1	99.0
Symptoms	43.4	34.6	99.7	63.0	99.2
Other	20.1	13.1	99.8	49.0	99.0

Key: Sens sensitivity; *Spec* specificity; *PPV* positive predictive value; *NPV* negative predictive value; *F1 score* harmonic mean of sensitivity and PPV

Table 4 Performance characteristics of LASSO to classify patients with dementia using primary care electronic medical record notes written before and after dementia case ascertainment in Ontario, Canada

Days relative to index[*]	F1 score	Sens	Spec	PPV	NPV	E50	E90	EMax	ICI
Before index date									
−730	12.2	6.9	99.9	54.6	98.9	0.008	0.01	0.56	0.01
−365	18.8	11.5	99.9	55.4	99.0	0.007	0.01	0.49	0.01
−120	26.6	18.3	99.8	54.2	99.0	–	–	–	–
−60	28.6	19.9	99.8	54.9	99.0	–	–	–	–
−30	20.1	13.2	99.8	52.0	98.7	–	–	–	–
−14	23.1	15.0	99.8	57.6	98.6	–	–	–	–
−7	29.4	19.3	99.9	70.2	98.6	0.003	0.01	0.998	0.01
After index date									
0	14.4	8.3	99.9	65.0	98.9	0.008	0.01	0.44	0.01
+7	34.5	24.3	99.8	60.0	99.1	0.01	0.01	0.26	0.01
+14	34.4	24.0	99.8	62.2	99.1	0.01	0.01	0.23	0.01
+30	40.7	31.4	99.8	61.7	99.2	0.01	0.01	0.26	0.01
+60	46.5	36.7	99.8	67.6	99.3	0.01	0.01	0.23	0.01
+120	54.6	44.7	99.8	71.8	99.3	0.001	0.003	0.21	0.004
+365	62.8	53.8	99.8	77.7	99.5	0.002	0.003	0.21	0.005
+730	71.8	66.2	99.8	78.8	99.6	0.001	0.003	0.23	0.004

Key: Sens sensitivity; *Spec* specificity; *PPV* positive predictive value; *NPV* negative predictive value; *F1 score* harmonic mean of sensitivity and PPV; "–" value could not be estimated

[*]Patients' index dates represent the date of dementia case ascertainment for patients with dementia or the date of the latest note in the EMR for patients without dementia. For dates prior to index, only notes taken from the prespecified date to the index date were included (e.g., 730 days prior to the index date to but not including the index date). For dates after index, all notes from the first note in the EMR to the prespecified date after the index date were included (e.g., first note in the EMR to 730 days after the index date)

in primary care settings, fairly little work has focused on the identification of neurological disorders and mental health conditions using machine learning techniques [20, 24, 27, 29, 46]. We used NLP with word embeddings to capture similar words and improve the comprehensiveness of our feature list. Compared to previous studies [38–40], our feature list included a broader set of features related to dementia symptoms and care including dementia and other medications, social issues, and health service use. This feature list may be used in future studies to enable detection of dementia in clinical notes.

Chen et al. designed a model to identify dementia and geriatric syndromes among patients 65 years and older enrolled in a US health maintenance organization using free-text electronic health records and NLP [20]. Their conditional random fields model showed good performance for identifying dementia (F1 = 77%, PPV = 63%, sensitivity = 100%) and relatively improved performance for identifying geriatric syndromes (macroaverage F1 = 83%, microaverage F1 = 85%). Bullard et al. used a bag-of-words model and logistic regression to identify individuals with Alzheimer's disease and mild cognitive impairment using data collected in the Alzheimer's Disease Neuroimaging

Initiative (ADNI) [19]. The performance of the model for Alzheimer's disease was relatively low (sensitivity = 63%, PPV = 65%) and modestly improved for mild cognitive impairment (sensitivity = 80%, PPV = 78%). As with our primary analysis, these studies did not examine clinical notes written *prior to* when a patient was diagnosed with dementia, which may, in part, explain their generally good performance.

Similar to our secondary analysis that investigated the timing of clinical notes on model performance, Hane et al. used unstructured clinical notes from EMRs to predict the risk of Alzheimer's and related dementias (including mild cognitive impairment) using gradient boosted models fit to a series of decision trees in the time leading up to diagnosis [27]. Compared to structured data alone (e.g., diagnosis codes), the models including clinical notes showed superior performance within the year of diagnosis (PPV = 68%, specificity = 99%, sensitivity = 68%). However, the performance of the model decreased with increasing time prior to diagnosis (3 years, specificity = 95%, sensitivity = 30%), indicating difficulties with making longer-term predictions. However, there is some evidence showing that cognitive symptoms written in clinical notes at hospital discharge were associated with an increased rate of dementia diagnosis up to 8 years prior [28].

Our findings that features related to medication, cognition, and health system use showed the best performance may help inform future work to improve the performance of machine learning models for dementia detection. EMR-based models for dementia detection may be useful to clinicians in screening patients for dementia and related signs and symptoms in settings where rapid identification of patients at risk may help to inform clinical care (e.g., physicians assessing a new patient, patients receiving care in the emergency department, patients receiving care in a multidisciplinary setting where many providers contribute to the EMR). However, the poor performance of the machine learning model prior to dementia case ascertainment suggests that improved documentation and additional sources of information may be needed to improve dementia detection (e.g., incorporating additional features of dementia symptoms, considering changes in symptoms over time).

Strengths of our study include the development of a broader list of features which encompasses factors that indicate early signs and symptoms of dementia ranging from impacts on cognition, activities of daily living, social supports, medications, and health system use. This feature list builds on previous work and demonstrates the potential for NLP methods to detect dementia in EMR data. Our study sample included a large sample of patients and was conducted in the primary care setting where patients are likely to first present with dementia.

Several limitations are worthy of consideration. First, the use of a health administrative data algorithm as the reference standard represents a "silver" standard as it is not a validated clinical diagnosis, and some patients may have been misclassified due to imperfect algorithm performance. For example, it is possible that some true cases of dementia were correctly identified as having dementia by the machine learning model(s) but labeled as a false-positive due to misclassification of the reference standard. However, data availability prevented a formal comparison to clinically validated assessments of dementia. As such, future work should further compare algorithm performance against a clinically reviewed sample of known cases of dementia. In addition, we did not have information on severity of cognitive impairment which may have

helped to inform the ability of the model to predict severity and stage of dementia. We were unable to distinguish between dementia subtypes, which may cause heterogeneity in the signs and symptoms of dementia and may impair the predictive ability of our model. Moreover, physician and practice variation may result in differences in where, when, and how early signs and symptoms of dementia are recorded in primary care EMRs. We also included features associated with dementia medications in our classification models, and as medication status was a component of the administrative and within-EMR dementia algorithms, this may have inflated the performance of features related to dementia medications. Last, although word embeddings can help to address challenges with misspelling and identify semantically similar words, they do not capture more complex phrases of words which may have relevance for dementia classification.

5 Conclusions

In the present study, we developed a feature list of signs and symptoms of dementia and applied this to classify patients with dementia in primary care EMRs using machine learning models. The LASSO model achieved good performance in classifying patients with dementia using all free-text progress and consult notes in the EMR, but relatively poor performance when using only notes collected prior to dementia case ascertainment. Although similar models may eventually improve the recognition of cognitive issues and dementia for primary care providers and for capacity planning across health systems, additional work is needed to refine and improve performance. Future research should focus on identifying which additional components of the EMR would enable earlier detection of dementia in order to maximize clinical utility.

Data Availability The dataset from this study is held securely in coded form at ICES. While legal data sharing agreements between ICES and data providers (e.g., healthcare organizations and government) prohibit ICES from making the dataset publicly available, access may be granted to those who meet prespecified criteria for confidential access, available at www.ices.on.ca/DAS (email: das@ices.on.ca).

Code Availability The full dataset creation plan and underlying analytic code are available from the authors upon request, understanding that the computer programs may rely upon coding templates or macros that are unique to ICES and are therefore either inaccessible or may require modification.

Declarations

Disclaimer The analyses, conclusions, opinions and statements expressed herein are solely those of the authors and do not reflect those of the funding or data sources; no endorsement is intended or should be inferred.

Conflict of Interest The authors declare no competing interests.

References

1. (2020) 2020 Alzheimer's disease facts and figures. Alzheimers Dement. https://doi.org/10.1002/alz.12068
2. Nichols E, Szoeke CEI, Vollset SE, Abbasi N, Abd-Allah F, Abdela J, . . . Murray CJL (2019) Global, regional, and national burden of Alzheimer's disease and other dementias, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet Neurol;18(1):88–106. https://doi.org/10.1016/S1474-4422(18)30403-4
3. Prince M BR, Ferri C. World Alzheimer Report 2011: the benefits of early diagnosis and intervention. London: Alzheimer's Disease International 2011. https://www.alzint.org/u/WorldAlzheimerReport2011.pdf. Accessed February 8, 2021.
4. Black CM, Fillit H, Xie L, Hu X, Kariburyo MF, Ambegaonkar BM, . . . Khandker RK (2018) Economic burden, mortality, and institutionalization in patients newly diagnosed with Alzheimer's disease. J Alzheimers Dis;61(1):185–93. https://doi.org/10.3233/JAD-170518
5. Rasmussen J, Langerman H (2019) Alzheimer's disease - why we need early diagnosis. Degener Neurol Neuromuscul Dis 9:123–130. https://doi.org/10.2147/DNND.S228939
6. Holzer S, Warner JP, Iliffe S (2013) Diagnosis and management of the patient with suspected dementia in primary care. Drugs Aging 30(9):667–676. https://doi.org/10.1007/s40266-013-0098-4
7. Fox C, Maidment I, Moniz-Cook E, White J, Thyrian JR, Young J, . . . Chew-Graham CA (2013) Optimising primary care for people with dementia. Ment Health Fam Med;10(3):143–51.
8. Valcour VG, Masaki KH, Curb JD, Blanchette PL (2000) The detection of dementia in the primary care setting. Arch Intern Med 160(19):2964–2968. https://doi.org/10.1001/archinte.160.19.2964
9. Mitchell AJ, Meader N, Pentzek M (2011) Clinical recognition of dementia and cognitive impairment in primary care: a meta-analysis of physician accuracy. Acta Psychiatr Scand 124(3):165–183. https://doi.org/10.1111/j.1600-0447.2011.01730.x
10. Boustani M, Callahan CM, Unverzagt FW, Austrom MG, Perkins AJ, Fultz BA, . . . Hendrie HC (2005) Implementing a screening and diagnosis program for dementia in primary care. J Gen Intern Med;20(7):572–7. https://doi.org/10.1111/j.1525-1497.2005.0126.x
11. Connolly A, Gaehl E, Martin H, Morris J, Purandare N (2011) Underdiagnosis of dementia in primary care: variations in the observed prevalence and comparisons to the expected prevalence. Aging Ment Health 15(8):978–984. https://doi.org/10.1080/13607863.2011.596805
12. Bradford A, Kunik ME, Schulz P, Williams SP, Singh H (2009) Missed and delayed diagnosis of dementia in primary care: prevalence and contributing factors. Alzheimer Dis Assoc Disord 23(4):306–314. https://doi.org/10.1097/WAD.0b013e3181a6bebc
13. Parmar J, Dobbs B, McKay R, Kirwan C, Cooper T, Marin A, Gupta N (2014) Diagnosis and management of dementia in primary care: exploratory study. Can Fam Physician 60(5):457–465
14. Goerdten J, Cukic I, Danso SO, Carriere I, Muniz-Terrera G (2019) Statistical methods for dementia risk prediction and recommendations for future work: a systematic review. Alzheimers Dement (N Y) 5:563–569. https://doi.org/10.1016/j.trci.2019.08.001
15. Tang EY, Harrison SL, Errington L, Gordon MF, Visser PJ, Novak G, . . . Stephan BC (2015) Current developments in dementia risk prediction modelling: an updated systematic review. PLoS One;10(9):e0136181. https://doi.org/10.1371/journal.pone.0136181
16. Pellegrini E, Ballerini L, Hernandez M, Chappell FM, Gonzalez-Castro V, Anblagan D, . . . Wardlaw JM (2018) Machine learning of neuroimaging for assisted diagnosis of cognitive impairment and dementia: a systematic review. Alzheimers Dement (Amst);10:519–35. https://doi.org/10.1016/j.dadm.2018.07.004

17. Stephan BC, Kurth T, Matthews FE, Brayne C, Dufouil C (2010) Dementia risk prediction in the population: are screening models accurate? Nat Rev Neurol 6(6):318–326. https://doi.org/10.1038/nrneurol.2010.54

18. Walters K, Hardoon S, Petersen I, Iliffe S, Omar RZ, Nazareth I, Rait G (2016) Predicting dementia risk in primary care: development and validation of the Dementia Risk Score using routinely collected data. BMC Med 14:6. https://doi.org/10.1186/s12916-016-0549-y

19. Bullard J, Alm CO, Liu X, Yu Q, Proano RA. Towards early dementia detection: fusing linguistic and non-linguistic clinical data. Proceedings of the Third Workshop on Computational Linguistics and Clinical Psychology 2016. https://aclanthology.org/W16-0302

20. Chen T, Dredze M, Weiner JP, Hernandez L, Kimura J, Kharrazi H (2019) Extraction of geriatric syndromes from electronic health record clinical notes: assessment of statistical natural language processing methods. JMIR Med Inform 7(1):e13039. https://doi.org/10.2196/13039

21. Ford E, Carroll JA, Smith HE, Scott D, Cassell JA (2016) Extracting information from the text of electronic medical records to improve case detection: a systematic review. J Am Med Inform Assoc 23(5):1007–1015. https://doi.org/10.1093/jamia/ocv180

22. Anzaldi LJ, Davison A, Boyd CM, Leff B, Kharrazi H (2017) Comparing clinician descriptions of frailty and geriatric syndromes using electronic health records: a retrospective cohort study. BMC Geriatr 17(1):248. https://doi.org/10.1186/s12877-017-0645-7

23. Aponte-Hao S, Wong ST, Thandi M, Ronksley P, McBrien K, Lee J, . . . Williamson T (2021) Machine learning for identification of frailty in Canadian primary care practices. Int J Pop D Sci;6(1).

24. Chase HS, Mitrani LR, Lu GG, Fulgieri DJ (2017) Early recognition of multiple sclerosis using natural language processing of the electronic health record. BMC Med Inform Decis Mak 17(1):24. https://doi.org/10.1186/s12911-017-0418-4

25. Jackson RG, Patel R, Jayatilleke N, Kolliakou A, Ball M, Gorrell G, . . . Stewart R (2017) Natural language processing to extract symptoms of severe mental illness from clinical text: the Clinical Record Interactive Search Comprehensive Data Extraction (CRIS-CODE) project. BMJ Open;7(1):e012012. https://doi.org/10.1136/bmjopen-2016-012012

26. Topaz M, Adams V, Wilson P, Woo K, Ryvicker M (2020) Free-text documentation of dementia symptoms in home healthcare: a natural language processing study. Gerontol Geriatr Med 6:2333721420959861. https://doi.org/10.1177/2333721420959861

27. Hane CA, Nori VS, Crown WH, Sanghavi DM, Bleicher P (2020) Predicting onset of dementia using clinical notes and machine learning: case-control study. JMIR Med Inform 8(6):e17819. https://doi.org/10.2196/17819

28. McCoy TH Jr, Han L, Pellegrini AM, Tanzi RE, Berretta S, Perlis RH (2020) Stratifying risk for dementia onset using large-scale electronic health record data: a retrospective cohort study. Alzheimers Dement 16(3):531–540. https://doi.org/10.1016/j.jalz.2019.09.084

29. Sheikhalishahi S, Miotto R, Dudley JT, Lavelli A, Rinaldi F, Osmani V (2019) Natural language processing of clinical notes on chronic diseases: systematic review. JMIR Med Inform 7(2):e12239. https://doi.org/10.2196/12239

30. Tu K, Mitiku TF, Ivers NM, Guo H, Lu H, Jaakkimainen L, . . . Tu JV (2014) Evaluation of electronic medical record administrative data linked database (EMRALD). Am J Manag Care;20(1):e15–21.

31. Tu K, Widdifield J, Young J, Oud W, Ivers NM, Butt DA, . . . Jaakkimainen L (2015) Are family physicians comprehensively using electronic medical records such that the data can be used for secondary purposes? A Canadian perspective. BMC Med Inform Decis Mak;15:67. https://doi.org/10.1186/s12911-015-0195-x

32. Tu K, Wang M, Young J, Green D, Ivers NM, Butt D, . . . Kapral MK (2013) Validity of administrative data for identifying patients who have had a stroke or transient ischemic attack using EMRALD as a reference standard. Can J Cardiol;29(11):1388–94. https://doi.org/10.1016/j.cjca.2013.07.676

33. Tu K, Mitiku T, Lee DS, Guo H, Tu JV (2010) Validation of physician billing and hospitalization data to identify patients with ischemic heart disease using data from the Electronic Medical Record Administrative data Linked Database (EMRALD). Can J Cardiol 26(7):e225–e228. https://doi.org/10.1016/s0828-282x(10)70412-8

34. Jaakkimainen RL, Bronskill SE, Tierney MC, Herrmann N, Green D, Young J, . . . Tu K (2016) Identification of physician-diagnosed Alzheimer's disease and related dementias in population-based administrative data: a validation study using family physicians' electronic medical records. J Alzheimers Dis;54(1):337–49. https://doi.org/10.3233/JAD-160105

35. Statistics Canada. Postal CodeOM Conversion File Plus (PCCF+) Version 6C, Reference Guide: Ottawa, Minister of Industry, 2016. https://www150.statcan.gc.ca/n1/en/catalogue/82F0086X.

36. Mondor L, Maxwell CJ, Hogan DB, Bronskill SE, Gruneir A, Lane NE, Wodchis WP (2017) Multimorbidity and healthcare utilization among home care clients with dementia in Ontario, Canada: a retrospective analysis of a population-based cohort. PLoS Med 14(3):e1002249. https://doi.org/10.1371/journal.pmed.1002249

37. Mondor L, Maxwell CJ, Bronskill SE, Gruneir A, Wodchis WP (2016) The relative impact of chronic conditions and multimorbidity on health-related quality of life in Ontario long-stay home care clients. Qual Life Res 25(10):2619–2632. https://doi.org/10.1007/s11136-016-1281-y

38. Halpern R, Seare J, Tong J, Hartry A, Olaoye A, Aigbogun MS (2019) Using electronic health records to estimate the prevalence of agitation in Alzheimer disease/dementia. Int J Geriatr Psychiatry 34(3):420–431. https://doi.org/10.1002/gps.5030

39. Wang L, Lakin J, Riley C, Korach Z, Frain LN, Zhou L (2018) Disease trajectories and end-of-life care for dementias: latent topic modeling and trend analysis using clinical notes. AMIA Annu Symp Proc 2018:1056–1065

40. Gilmore-Bykovskyi AL, Block LM, Walljasper L, Hill N, Gleason C, Shah MN (2018) Unstructured clinical documentation reflecting cognitive and behavioral dysfunction: toward an EHR-based phenotype for cognitive impairment. J Am Med Inform Assoc 25(9):1206–1212. https://doi.org/10.1093/jamia/ocy070

41. Wang B, Wang A, Chen F, Wang Y, Kuo C-CJ (2019) Evaluating word embedding models: methods and experimental results. APSIPA Transactions on Signal and Information Processing;8(E19). https://doi.org/10.1017/ATSIP.2019.12

42. Khattak FK, Jeblee S, Pou-Prom C, Abdalla M, Meaney C, Rudzicz F (2019) A survey of word embeddings for clinical text. J Biomed Inform 4:100057. https://doi.org/10.1016/j.yjbinx.2019.100057

43. Wang Y, Liu S, Afzal N, Rastegar-Mojarad M, Wang L, Shen F, . . . Liu H (2018) A comparison of word embeddings for the biomedical natural language processing. Journal of biomedical informatics;87:12–20.

44. Austin PC, Steyerberg EW (2019) The Integrated Calibration Index (ICI) and related metrics for quantifying the calibration of logistic regression models. Stat Med 38(21):4051–4065

45. Tonelli M, Wiebe N, Fortin M, Guthrie B, Hemmelgarn BR, James MT, . . . For the Alberta Kidney Disease N (2015) Methods for identifying 30 chronic conditions: application to administrative data. BMC Medical Informatics and Decision Making;15(1):31. https://doi.org/10.1186/s12911-015-0155-5

46. Shao Y, Zeng QT, Chen KK, Shutes-David A, Thielke SM, Tsuang DW (2019) Detection of probable dementia cases in undiagnosed patients using structured and unstructured electronic health records. BMC Med Inform Decis Mak 19(1):1–11

Authors and Affiliations

Laura C. Maclagan[1] · Mohamed Abdalla[2] · Daniel A. Harris[3] · Therese A. Stukel[1,4,5] · Branson Chen[1] · Elisa Candido[1] · Richard H. Swartz[1,5,6] · Andrea Iaboni[7,8] · R. Liisa Jaakkimainen[1,4,5,9] · Susan E. Bronskill[1,3,4,5,10]

✉   Susan E. Bronskill
    susan.bronskill@ices.on.ca

[1]   ICES, G1-06, 2075 Bayview Avenue, Toronto M4N 3M5, Canada

[2]   Department of Computer Science, University of Toronto, Toronto, Canada

[3]   Division of Epidemiology, Dalla Lana School of Public Health, University of Toronto, Toronto, Canada

[4]   Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, Canada

[5]   Sunnybrook Research Institute, Sunnybrook Health Sciences Centre, Toronto, Canada

[6]   Department of Medicine (Neurology), Sunnybrook Health Sciences Centre, University of Toronto, Toronto, Canada

[7]   KITE Research Institute, Toronto Rehabilitation Institute, University Health Network, Toronto, Canada

[8]   Department of Psychiatry, University of Toronto, Toronto, Canada

[9]   Department of Family and Community Medicine, University of Toronto, Toronto, Canada

[10] Women's College Research Institute, Women's College Hospital, Toronto, Canada