

Original Paper

Use of Patient-Reported Symptoms from an Online Symptom Tracking Tool for Dementia Severity Staging: Development and Validation of a Machine Learning Approach

Aaqib Shehzad^{1*}, BDS, MHI; Kenneth Rockwood^{1,2,3*}, MD; Justin Stanley^{1*}, BEng; Taylor Dunn^{1*}, MSc; Susan E Howlett^{1,3,4*}, PhD

¹DGI Clinical Inc, Halifax, NS, Canada

²Geriatric Medicine Research Unit, Nova Scotia Health Authority, Halifax, NS, Canada

³Division of Geriatric Medicine, Dalhousie University, Halifax, NS, Canada

⁴Department of Pharmacology, Dalhousie University, Halifax, NS, Canada

* all authors contributed equally

Corresponding Author:

Kenneth Rockwood, MD

Geriatric Medicine Research Unit

Nova Scotia Health Authority

5955 Veterans' Memorial Lane

Halifax, NS, B3H 2E9

Canada

Phone: 1 9024738687

Fax: 1 9024731050

Email: research@dgiclinical.com

Abstract

Background: SymptomGuide Dementia (DGI Clinical Inc) is a publicly available online symptom tracking tool to support caregivers of persons living with dementia. The value of such data are enhanced when the specific dementia stage is identified.

Objective: We aimed to develop a supervised machine learning algorithm to classify dementia stages based on tracked symptoms.

Methods: We employed clinical data from 717 people from 3 sources: (1) a memory clinic; (2) long-term care; and (3) an open-label trial of donepezil in vascular and mixed dementia (VASPECT). Symptoms were captured with SymptomGuide Dementia. A clinician classified participants into 4 groups using either the Functional Assessment Staging Test or the Global Deterioration Scale as mild cognitive impairment, mild dementia, moderate dementia, or severe dementia. Individualized symptom profiles from the pooled data were used to train machine learning models to predict dementia severity. Models trained with 6 different machine learning algorithms were compared using nested cross-validation to identify the best performing model. Model performance was assessed using measures of balanced accuracy, precision, recall, Cohen κ , area under the receiver operating characteristic curve (AUROC), and area under the precision-recall curve (AUPRC). The best performing algorithm was used to train a model optimized for balanced accuracy.

Results: The study population was mostly female (424/717, 59.1%), older adults (mean 77.3 years, SD 10.6, range 40-100) with mild to moderate dementia (332/717, 46.3%). Age, duration of symptoms, 37 unique dementia symptoms, and 10 symptom-derived variables were used to distinguish dementia stages. A model trained with a support vector machine learning algorithm using a one-versus-rest approach showed the best performance. The correct dementia stage was identified with 83% balanced accuracy (Cohen κ =0.81, AUPRC 0.91, AUROC 0.96). The best performance was seen when classifying severe dementia (AUROC 0.99).

Conclusions: A supervised machine learning algorithm exhibited excellent performance in identifying dementia stages based on dementia symptoms reported in an online environment. This novel dementia staging algorithm can be used to describe dementia stage based on user-reported symptoms. This type of symptom recording offers real-world data that reflect important symptoms in people with dementia.

(*J Med Internet Res* 2020;22(11):e20840) doi: [10.2196/20840](https://doi.org/10.2196/20840)

KEYWORDS

dementia stage; Alzheimer disease; mild cognitive impairment; machine learning

Introduction

Background

People living with dementia experience a variety of symptoms. These symptoms cross several domains beyond cognition, including executive function (eg, planning [1]), behavior (eg, agitation [2]), and physical manifestations (eg, mobility [3]). This heterogeneity of symptoms is further increased by changes in daily occurrence and manifestation. Furthermore, these combinations can vary both between people and within people across time [4-6]. This variability can be informative. Hallucinations, for example, have been reported in all stages of Alzheimer disease but most commonly at later stages [7]. In contrast, in people with Lewy body dementia, they can be a presenting feature [8]. The complex nature of dementia poses diagnostic and management challenges for health care professionals [9,10]. A key strategy is recognizing patterns, which forms the basis of dementia staging. Pattern recognition can be enhanced by tracking dementia symptoms early in the course of progressive cognitive impairment. This is especially useful when employing an approach that allows common but under-studied symptoms (eg, verbal repetition [11] or misplacing objects [12]), which may nevertheless be informative when assembled in an accessible fashion [5,6] or respond to treatment [13], to be recognized and evaluated.

Requirement for Dementia Staging Tools

To allow individual applicability, any treatment approach must consider the person's dementia stage [14]. Several clinician-facilitated dementia tools allow face-to-face staging, including the Global Deterioration Scale (GDS) [15], the Functional Assessment Staging Test (FAST) [16], the Dependence Scale [17], and the Clinical Dementia Rating Scale Sum of Boxes [18]. Defining dementia from unadjudicated online encounters (ie, where people living with dementia symptoms or their care partners track their symptoms in a web-based tool) is an important challenge that could improve both early detection and treatment evaluation [19]. Even so, dementia staging from solely online interactions has rarely been explored [20-22].

Online symptom tracking tools are common ways to help health care professionals understand dementia symptoms. They can also be valuable as education tools. SymptomGuide Dementia (DGI Clinical Inc) is an online dementia symptom tracking tool that provides a library of common and distressing symptoms. It serves as an educational tool and allows a user to identify symptoms of concern and track their change over time [5,23]. Earlier, we developed an algorithm to stage dementia severity into 4 levels of cognitive impairment for use with SymptomGuide Dementia or other similar databases using clinician-staged symptom profiles of 320 people [24]. Here, we aimed to develop a new staging algorithm using machine learning techniques with training data from a larger and more diverse set of clinical data and to validate this approach with well-established clinical dementia staging tools.

Methods

Participants and Procedure

Data for this study were obtained from a tertiary care memory clinic in Halifax, Nova Scotia, Canada from 2007 to 2013 as well as data from a study in long-term care, and an open-label trial of donepezil in vascular and mixed dementia (VASPECT) clinical trial [25,26]. Data from patients and family members (care partners) were collected using SymptomGuide Dementia in its electronic (web-based) or paper format. In addition, participants (N=717) underwent standard clinical assessments, including staging of dementia with one of two clinical tests, the GDS or the FAST. Both GDS and FAST have excellent reliability and validity [16,27]. Additionally, FAST stages have been shown to be concordant with GDS stages, and a correlation of 0.9 has been observed between them [28]. The GDS and FAST scores were interpreted as follows: a score of 3 indicated mild cognitive impairment, a score of 4 indicated mild dementia, a score of 5 indicated moderate dementia, and a score of 6 indicated severe dementia. These stages were used as target variables for classification prediction. All 4 stages were treated as discrete; therefore, discriminative models were used to perform the classification task. Only data collected at baseline (first visit) for each participant were prepared and used to train the models.

A web-based symptom tracking tool aimed to support caregivers of persons living with dementia, SymptomGuide Dementia, was used for data capture and storage for data obtained from the 3 sources. The symptoms can be either selected from an existing library of standardized symptoms or created by the caregiver. For each of the standardized symptoms, several plain-language descriptors are present. These provided another submenu for selection by the user. For each symptom selected, users were asked to indicate the frequency of the symptom and rank all the symptoms from most to least important. Users were also asked to input demographic information (eg, age and gender) and health-related information (eg, duration since first symptom), which was attached to their symptom profiles. Symptom information for each participant in the 3 sources was coded in the same format as represented in the online database. We, therefore, refer to *participants* when describing their characteristics and *user profiles* in relation to the representation of their symptoms.

Data Preparation

Users who did not select at least one symptom from the existing library of standardized symptoms were excluded from the analysis. Any patient age reported as less than 40 years was replaced with the group average for the respective stage. This was done with the assumption that the survey question was misinterpreted, and the reported age was the care partner's age not the age of the participant with dementia. Each symptom was represented by the ratio of descriptors selected for that symptom to the total number of descriptors selected across all symptoms by the participant. In addition to individual

symptoms, the ratio of selected descriptors and ratio of reported frequency of all symptoms were grouped into the following 5 domains: Behavioral Function, Cognitive Function, Daily Function, Executive Function, and Physical Manifestations for each participant [29]. Finally, age and the duration of symptoms were also included as features (variables used for prediction). All features were continuous except for the duration of symptoms which was treated as categorical data (I don't know, 1-3 months, 3-6 months, 6-12 months, 1-2 years, 1 year or more). Of the 56 symptoms in the standardized menu in SymptomGuide Dementia, 37 symptoms were selected to be included in the final set of features. This was accomplished by pruning symptoms based on a minimum occurrence of at least 15 times. As before in the algorithm developed with 320 users [24], here we maintained the 4 common clinical classifications of mild cognitive impairment, and mild dementia, moderate dementia, and severe dementia. Of the 717 users, a majority (332, 46.3%) were clinically staged as having mild dementia with the FAST or GDS, 133 (18.5%) as having mild cognitive impairment, 138 (19.2%) as having moderate dementia, and 114 (15.8%) as having severe dementia.

Since the different dementia stages were not equally represented in the data set, the minority stages (eg, mild cognitive impairment, moderate dementia, and severe dementia) were oversampled and the most represented stage (mild dementia) was undersampled in the machine learning pipeline. Oversampling was done with the borderline variant of the synthetic minority oversampling technique algorithm with a target of increasing the minority stage sizes by approximately 1.45 times their original size [30]. Undersampling was done with the neighborhood cleaning rule algorithm that focuses on

data cleaning rather than data reduction. This technique has been previously shown to improve identification of minority classes in machine learning [31].

Building the Model

The models were adjudicated and iterated using measures of balanced accuracy, precision (also known as positive predictive value), recall (also known as sensitivity), Cohen κ , area under the receiver operating characteristics curve (AUROC), and area under the precision-recall curve (AUPRC). Balanced accuracy in this study was the average of individual accuracy for each stage [32]. In a balanced data set, this score would represent the accuracy. Data were stratified by stage and randomly split with 70% of the data used as a training data set ($n=502$) and 30% of the data used as a test data set ($n=215$) for validation.

The use of a single set of data to conduct both model selection and model training can lead to overfitting and selection bias [33]. To address this, we used a nested cross-validation approach as described in Figure 1. The average inner cross-validation estimates of the primary selection criterion were maximized by selecting optimal hyperparameters from a range of possible values. The inner and outer cross-validation loops were repeated 3 times to account for variance arising from choice of data set splits [34,35]. We used 5-fold cross validation for both the inner and the outer loops. We used balanced accuracy here as the primary selection criterion for the hyperparameter tuning in the inner loop. Balanced accuracy was also used for the outer loop to provide a measure of model performance. The following machine learning algorithms were used to train models: support vector machine, k-nearest neighbor, random forest, neural network, logistic regression, stochastic gradient boosting.

Figure 1. Pseudocode representation of the nested cross-validation procedure used during model selection trials.

Algorithm 1: Stratified K-Fold Nested Cross-Validation with Grid Search

Require: K_1 , K_2 , where K_1 is number of outer folds and K_2 inner folds

Require: D , dataset containing input features X and output feature y

Require: P_{sets} , set of hyper-parameters with different values

Require: M , a single estimator, model.

Split D into stratified K_1 folds

for $i = 1$ to K_1 splits **do**

 Set fold i as D_i^{test} , and remaining folds as D_i^{train}

 Split D_i^{train} into stratified K_2 folds

for $j = 1$ to K_2 **do**

 Set fold j as D_j^{test} , and remaining folds as D_j^{train}

foreach p in $SequentialSample(P_{sets})$ **do**

 Train M on D_j^{train} with hyperparameter set p

 Compute test error E_j^{test} for M with D_j^{test}

end

end

 Select optimal hyperparameter set p^* from P_{sets} , where E_j^{test} is best

 Train M with D_i^{train} , using p^*

 Compute test error E_i^{test} for M with D_i^{test}

end

The best performing algorithm was then trained on the complete training data set with nested cross-validated hyperparameter tuning to model the data. To understand model performance and guard against overfitting, the model was tested against the test data set to obtain performance estimates of following metrics: weighted precision, weighted recall, balanced accuracy, Cohen κ , AUROC, and AUPRC. The final model was further assessed with a permutation test, which measured the likelihood of obtaining the observed accuracy by chance. This was done by repeating the classification (training and testing) procedure 200 times after randomly shuffling the data and permuting the labels in each iteration. The scores obtained with the permuted data were compared with the scores from the original data. We computed the probability of obtaining a score with permuted data that was better than with original data. Obtaining a small probability value rejects the null hypothesis that our model performed better than random chance and that the model had learned a real relationship between our data and dementia stages [36]. In other words, this process estimates how likely it is to obtain the observed classification performance on the test set by chance [37].

All data processing, analysis, and visualization were performed using Python (version 3.6; 64-bit) libraries (numpy, version 1.18.1; scipy, version 1.4.1; matplotlib, version 3.1.2; pandas, version 0.25.3) [38-42]. Classification algorithms were processed and analyzed using scikit-learn (version 0.22.1) and scipy [39,43]. The synthetic minority oversampling technique and neighborhood cleaning rule were implemented using imbalanced-learn (version 0.6.1) [44].

Results

Participants

This study used data from memory clinic (n=420), a long-term care study (n=169), and the VASPECT clinical trial (n=128) for a participant sample that allows 717 user profiles in people with clinical diagnosis and staging (Table 1) [25,26]. The mean participant age was 77.3 years (SD 10.6 years), and 59.1% of the participants were women. The mean FAST score was 4.0 (SD 0.9), and the mean GDS score was 4.8 (SD 1.9). The participants identified a median of 5 symptoms (range 1-27).

Table 1. Descriptive statistics of participants from clinical studies, by data source.

Characteristic	Memory clinic	Long-term care	VASPECT ^a open-label trial	Total
Sample size, n (%)	420 (58.5)	169 (23.5)	128 (18.0)	717 (100)
Age (in years), mean (SD)	74.6 (12.5)	81.0 (19.1)	75.4 (9.2)	77.3 (10.6)
FAST, mean (SD)	4.0 (0.9)	5.3 (1.1)	4.3 (0.5)	4.1 (0.9)
GDS, mean (SD)	4.8 (1.9)	5.2 (1.0)	— ^b	5.2 (1.1)
Reported symptoms, median (range)	5 (1-14)	4 (1-12)	6 (1-27)	5 (1-27)
Sex, n (%)				
Female	228 (54.3)	129 (76.3)	67 (52.3)	424 (59.1)
Male	192 (45.7)	40 (23.7)	61 (47.7)	293 (40.9)
FAST, mean (SD)	4.0 (0.9)	5.3 (1.1)	4.3 (0.5)	4.1 (0.9)
GDS, mean (SD)	4.8 (1.9)	5.2 (1.0)	—	5.2 (1.1)
Reported symptoms, median (range)	5 (1-14)	4 (1-12)	6 (1-27)	5 (1-27)
Reported symptoms by dementia stage, median (range)				
Mild cognitive impairment	3 (1-14)	2 (2-4)	—	4 (1-14)
Mild dementia	5 (1-11)	3 (1-8)	6 (1-24)	5 (1-24)
Moderate dementia	5.5 (1-11)	4 (1-7)	7 (2-27)	4.5 (1-27)
Severe dementia	5 (2-11)	5 (2-12)	10 (7-13)	5 (1-13)
Stage, n (%)				
Mild cognitive impairment	126 (30.0)	7 (4.1)	—	133 (18.5)
Mild dementia	203 (48.3)	33 (19.5)	96 (75)	332 (46.3)
Moderate dementia	58 (13.8)	50 (29.6)	30 (23.4)	138 (19.2)
Severe dementia	33 (7.8)	79 (46.7)	2 (1.5)	114 (15.8)
Age (years), mean (SD)				
Mild cognitive impairment	71.2 (14.3)	87.5 (9.9)	—	73.2 (12.2)
Mild dementia	75.3 (11.8)	81.5 (16.3)	74.7 (8.8)	76.4 (9.8)
Moderate dementia	77.1 (10.1)	83.6 (14.1)	77.5 (9.9)	80 (9.4)
Severe dementia	77.3 (10.7)	78.5 (23)	75 (16.9)	81.2 (9.7)

^aAn open-label trial of donepezil in vascular and mixed dementia.^bNo data from this source.

Symptoms

Table 2 shows the frequency of dementia symptoms reported for user profiles and classified by dementia stage as assessed clinically with the FAST and GDS tools.

Table 2 illustrates the relationship between symptom frequency and clinical dementia stage. There was a sharp increase in the frequency of aggression, wandering, and incontinence in patients with severe dementia. By contrast, symptoms such as memory of recent events, repetitive questioning, and initiative declined with increasing dementia severity.

Table 2. Mean frequency of reported user profile symptoms of clinical study participants by clinically defined dementia stage.

Symptom	Mild cognitive impairment, n (%)	Mild dementia, n (%)	Moderate dementia, n (%)	Severe dementia, n (%)
Aggression	1 (0.8)	7 (2.1)	4 (2.9)	32 (28.1)
Anxiety & worry	35 (26.3)	82 (24.7)	23 (16.7)	19 (16.7)
Appetite	0 (0)	28 (8.4)	12 (8.7)	9 (7.9)
Balance	5 (3.8)	7 (2.1)	13 (9.4)	19 (16.7)
Bathing	0 (0)	9 (2.7)	11 (8.0)	6 (5.3)
Delusions & paranoia	4 (3.0)	29 (8.7)	20 (14.5)	19 (16.7)
Disorientation to place	3 (2.3)	28 (8.4)	15 (10.9)	18 (15.8)
Disorientation to time	4 (3.0)	41 (12.3)	25 (18.1)	23 (20.2)
Dressing	1 (0.8)	13 (3.9)	15 (10.9)	7 (6.1)
Eating	0 (0)	5 (1.5)	3 (2.2)	9 (7.9)
Financial management	5 (3.8)	30 (9.0)	11 (8.0)	1 (0.9)
Following instructions	6 (4.5)	12 (3.6)	2 (1.4)	2 (1.8)
Hallucinations	0 (0)	15 (4.5)	11 (8.0)	12 (10.5)
Hobbies & games	2 (1.5)	17 (5.1)	5 (3.6)	0 (0)
Household chores	4 (3.0)	39 (11.7)	13 (9.4)	5 (4.4)
Incontinence	1 (0.8)	6 (1.8)	6 (4.3)	23 (20.2)
Insight	4 (3.0)	19 (5.7)	6 (4.3)	2 (1.8)
Interest initiative	46 (34.6)	148 (44.6)	39 (28.3)	19 (16.7)
Irritability frustration	19 (14.3)	92 (27.7)	24 (17.4)	28 (24.6)
Judgment	13 (9.8)	42 (12.7)	24 (17.4)	25 (21.9)
Language difficulty	18 (13.5)	59 (17.8)	15 (10.9)	21 (18.4)
Low mood	3 (2.3)	44 (13.3)	13 (9.4)	6 (5.3)
Meal preparation cooking	10 (7.5)	57 (17.2)	19 (13.8)	2 (1.8)
Memory for names faces	2 (1.5)	29 (8.7)	24 (17.4)	12 (10.5)
Memory of future events	0 (0)	37 (11.1)	16 (11.6)	1 (0.9)
Memory of past events	8 (6.0)	33 (9.9)	21 (15.2)	13 (11.4)
Memory of recent events	100 (75.2)	233 (70.2)	81 (58.7)	30 (26.3)
Misplacing or losing objects	21 (15.8)	53 (16.0)	11 (8)	5 (4.4)
Mobility	4 (3.0)	14 (4.2)	16 (11.6)	30 (26.3)
Operating gadgets/appliances	8 (6.0)	76 (22.9)	24 (17.4)	4 (3.5)
Personal care hygiene	7 (5.3)	26 (7.8)	36 (26.1)	28 (24.6)
Physical complaints	2 (1.5)	15 (4.5)	7 (5.1)	4 (3.5)
Repetitive questions stories	51 (38.3)	169 (50.9)	50 (36.2)	18 (15.8)
Shadowing	1 (0.8)	8 (2.4)	8 (5.8)	4 (3.5)
Social interaction/withdrawal	20 (15.0)	64 (19.3)	22 (15.9)	11 (9.6)
Wandering	0 (0)	2 (0.6)	10 (7.2)	29 (25.4)

Model Selection

Six machine learning models were tested on the training data set. [Table 3](#) illustrates the models used and the validation data obtained for each model in terms of accuracy, precision, and recall when predicting dementia stage. The table also indicates

values for the Cohen κ , which measures the agreement between the dementia stage predicted by the model and the dementia stage as determined clinically. At the end of the model selection process, the model trained with a support vector machine was selected as the best performing model when used with the training data set ([Table 3](#)).

Table 3. Performance of candidate models with the training data set.

Model	Balanced accuracy, mean (SD)	Precision (weighted), mean (SD)	Recall (weighted), mean (SD)	Cohen κ , mean (SD)
Support vector machine	0.73 (0.07)	0.75 (0.07)	0.75 (0.06)	0.65 (0.09)
k-nearest neighbor	0.72 (0.08)	0.73 (0.07)	0.72 (0.07)	0.62 (0.10)
Random forest	0.70 (0.07)	0.74 (0.08)	0.73 (0.07)	0.62 (0.09)
Neural network	0.66 (0.10)	0.67 (0.11)	0.66 (0.09)	0.54 (0.13)
Logistic regression	0.65 (0.08)	0.66 (0.08)	0.66 (0.08)	0.53 (0.10)
Gradient boosting	0.68 (0.07)	0.70 (0.07)	0.70 (0.07)	0.58 (0.10)

Next, the support vector machine was trained and optimized with a nested cross-validated grid search on the complete training set. The final trained model was used with the test data set to obtain performance metrics for this new data subset (balanced accuracy 0.85; AUROC 0.96, weighted precision 0.87; weighted recall 0.86; AUPRC 0.91), indicating excellent model performance.

Final Model Prediction Based on Dementia Stage

The ability of the support vector machine model to predict each of the 4 dementia stages showed excellent precision and recall for all dementia stages (Table 4).

To better demonstrate predictions made across the dementia stages by the model, a confusion matrix is presented in Figure 2.

To determine the relationship between the true positives and false positives identified by the model, receiver operating characteristic curves of the model's output were plotted (Figure 3). The AUROC for the overall model was high (AUROC 0.96). The final model achieved the best results when classifying severe dementia (AUROC 0.98) and mild cognitive impairment (AUROC 0.97).

Table 4. Precision and recall of model prediction by dementia stage.

Stage	Precision	Recall
Mild cognitive impairment	0.85	0.87
Mild dementia	0.82	0.89
Moderate dementia	0.91	0.80
Severe dementia	0.93	0.86

Figure 2. Confusion matrix of the trained model. Each row of the matrix represents the instances of actual dementia stage while each column represents the instances of predicted dementia stage. Counts are colored from the highest cell (darker) to the lowest (lighter). The top-left to bottom-right diagonal cells count correctly predicted dementia stages.

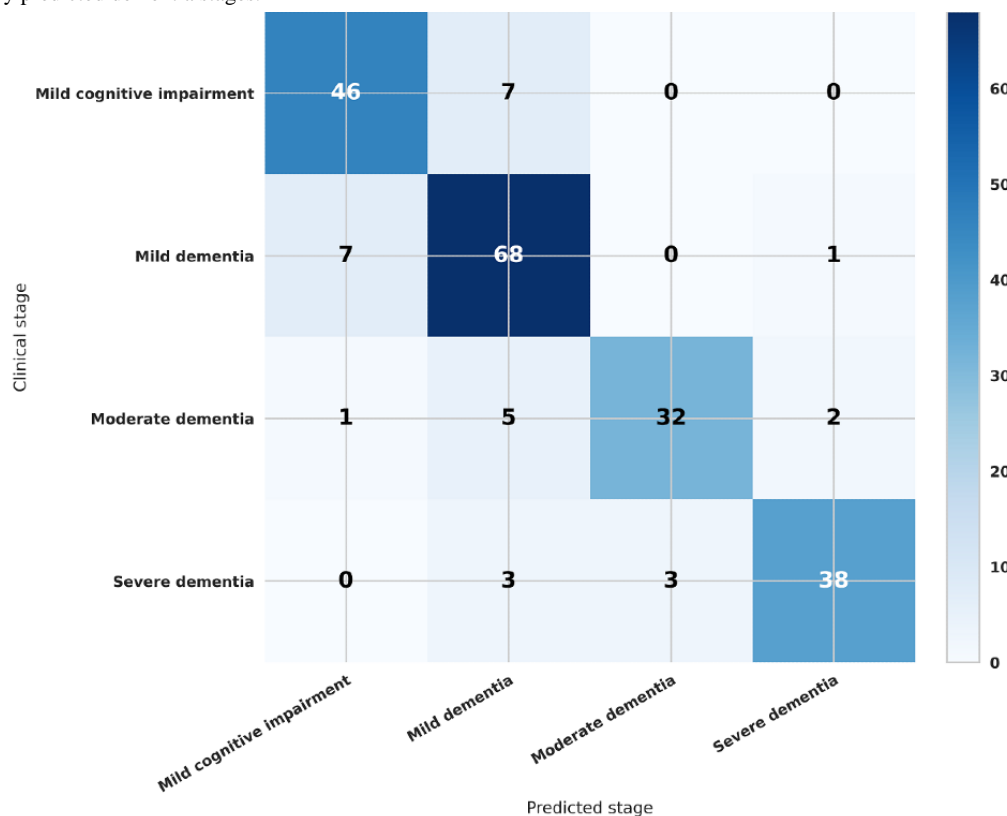
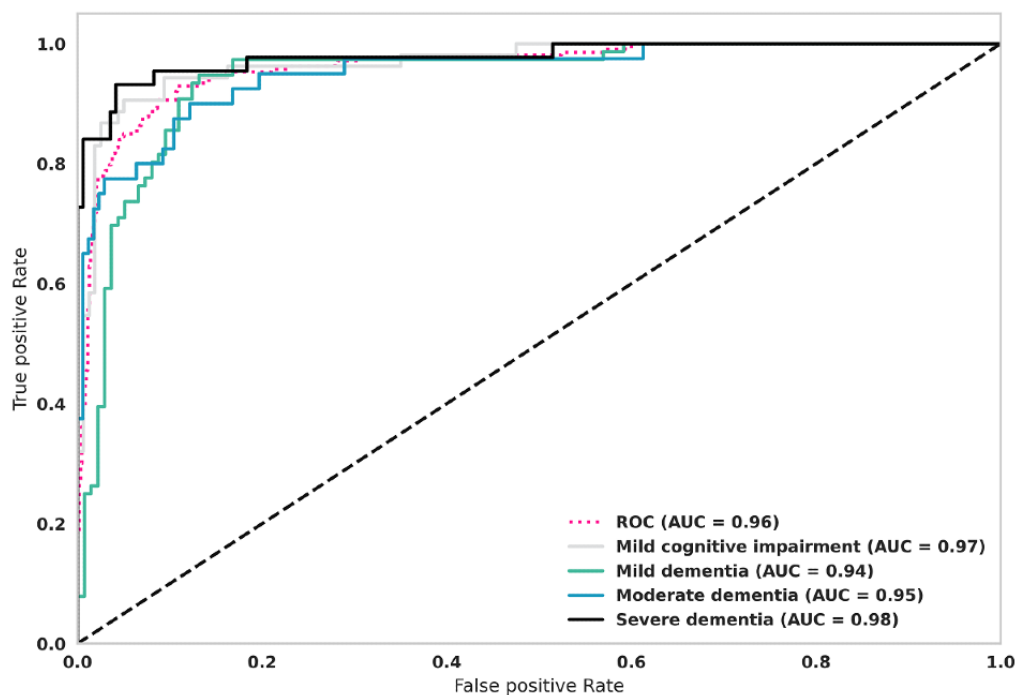


Figure 3. Receiver operating characteristic curves for each dementia stage predicted by the model. AUC: area under the curve; ROC: receiver operating characteristics.



Another way to assess the relationship between false positives and false negatives is to use a precision-recall curve, where high precision indicates a low false positive rate, and high recall denotes a low false negative rate. Figure 4 shows precision-recall curves of the overall model output by dementia stage. The

overall model performed well (AUPRC 0.91). When AUPRC metrics were compared for individual dementia stages, the model performed best when classifying severe dementia (AUPRC 0.95) and mild cognitive impairment (AUPRC 0.93). It was somewhat less able to discriminate between mild and moderate

dementia (Figure 4). These observations are similar to those seen when these relationships were evaluated with receiver operating characteristic curves as shown in Figure 3.

To confirm that the model could accurately predict dementia stage, we performed a permutation test, where we used randomly

mislabeled data in several iterations, grouped about the level expected by chance (Figure 5). The random permutation scores had a balanced accuracy between 0.2 and 0.3. This was well short of the classification score for the actual data, which had a balanced accuracy of 0.85, and a probability of obtaining this by chance $<.005$.

Figure 4. Precision-recall curves for each dementia stage as predicted by the model. AUC: area under the curve; PRC: precision-recall curve.

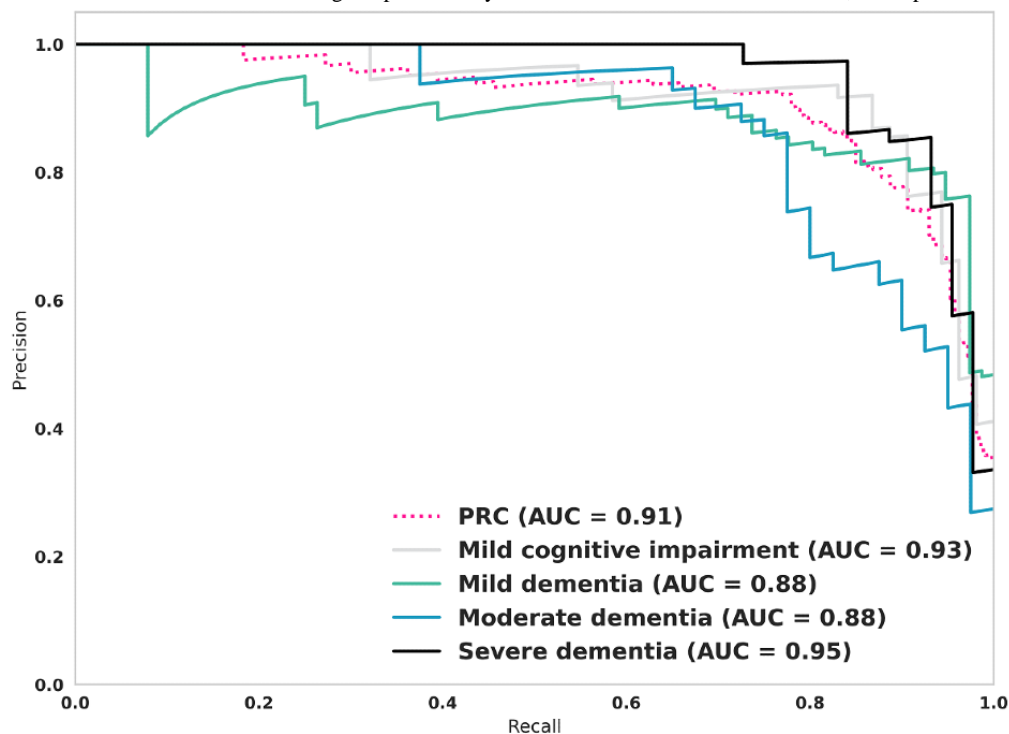
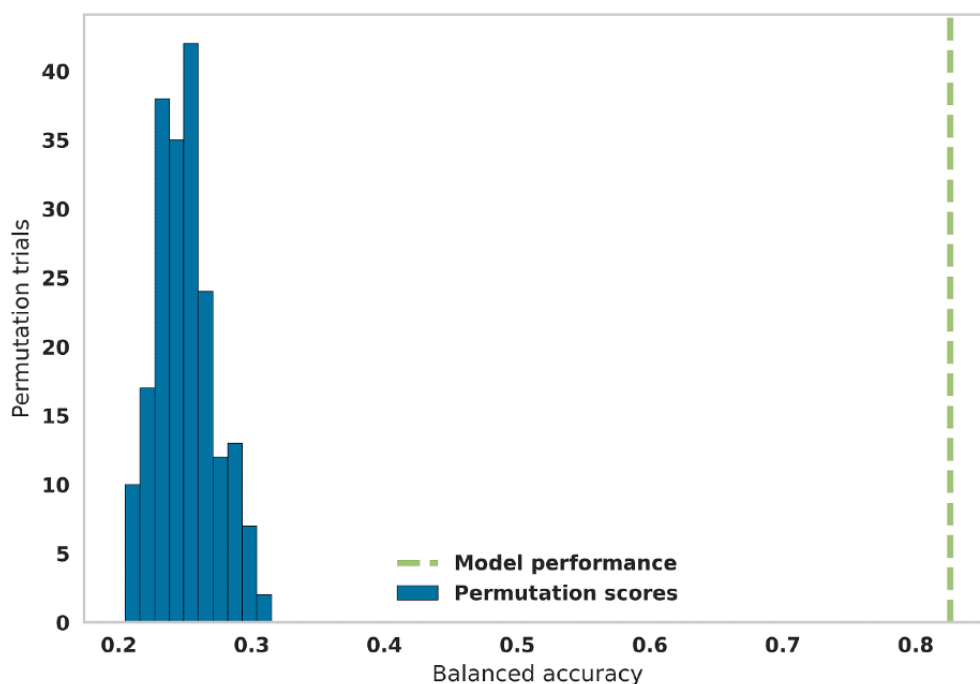


Figure 5. The classification scores obtained from models trained on permuted data were well short of scores obtained with the model trained on original data.



Discussion

This study aimed to stage dementia severity based on symptom profiles constructed with a standardized symptom menu from an online symptom tracking tool. We found that a support vector machine model consistently predicted each of the 4 dementia stages based on online symptom data reported by caregivers of persons with dementia. This approach to staging dementia severity will allow us to gain insights from online reported symptom data that can be collected by SymptomGuide Dementia and other similar platforms. In this way, symptom reporting can facilitate understanding dementia progression. For example, earlier work from this database suggests important qualitative differences in symptoms such as misplacing objects (eg, with dementia progression, less instances of simply forgetting where an item might be and more instances of placing items in an odd place [6]) or verbal repetition (eg, repetitive questioning, most often seen in mild dementia; it is characteristically dementia-defining when seen with early functional decline—difficulty operating familiar gadgets or appliances [5]). In this way, allowing the patients and carer voices to contribute to our understanding of dementia phenomenology can lead to recognizing patterns of both progression—as above—and of treatment [13]. The updated staging algorithm described here will further such inquiries.

We trained multiple machine learning algorithms and selected the best performing algorithm to use for our dementia stage classification task. A support vector machine model using a one-versus-rest approach demonstrated the best performance during model selection trials. The selected algorithm was then trained on the complete training data and validated using a test data set. The final model demonstrated excellent performance in discriminating dementia stages (balanced accuracy 0.85, AUROC 0.96). Receiver operating characteristic curves tend to present an optimistic picture of performance when the data set has a skewed distribution of the target variable [45]. For this reason, the performance of the model was also assessed with precision-recall curves. These too demonstrated that the model performed well, especially when classifying severe dementia. Since mild cognitive impairment and severe dementia can be considered bookends to the dementia spectrum, we can be reassured of both the model's precision and recall in classifying these extremes. For example, our model correctly classified a 75-year-old participant who reported 4 symptoms (social interaction/withdrawal, irritability and frustration, interest and initiative, aggression) as having mild cognitive impairment and a 76-year-old participant as having severe dementia based on a different set of 3 symptoms (wandering, delusions and paranoia, and aggression). The model was somewhat less accurate when classifying mild and moderate dementia. This is perhaps not surprising as symptom profiles in the middle of the dementia spectrum can exhibit a higher degree of overlap and can be difficult to distinguish clinically as well [46].

The very low probability value from the permutation tests ($<.005$) reassures us that the model learned a real relationship between the data and dementia stages. It demonstrates that the classification performance of the model with respect to the test set is unlikely to have occurred as a result of chance.

Our data must be interpreted with caution. For model stability, symptoms were eliminated based on a set threshold of occurrence. While this worked well here, it might not hold in a larger data set. In addition, we used 3 separate data sets that used variations of our standardized symptom menu, with differences in the composition and order of presentation of the symptoms. Since most of these patient symptom profiles were constructed with the supervision of a clinician or a rater, the model may be less generalizable to web-based symptom profiles constructed without clinician facilitation or guidance.

Several other recent studies have applied machine learning algorithms for dementia research [47–51]. Most have used neuroimaging or biomarker data to train these models. Most models trained with neuroimaging data focus on distinguishing individual patients from healthy controls, whereas our model distinguished between different stages of dementia severity [52]. Extraction of image characteristics from neuroimaging data can be susceptible to variations in the scanner hardware and image acquisition protocols. This can produce models that may not be generalizable when applied to data acquired from different imaging sources [52]. Additionally, scans such as amyloid positron emission tomography imaging, used for diagnostic certainty regarding Alzheimer disease, can cost upward of US \$4000. Machine learning models that do not rely on neuroimaging data to stage or diagnose dementia, if used clinically, can potentially reduce the number of participants that require expensive neuroimaging tests [53].

More recent studies have also used data extracted from electronic health records which may include structured and unstructured data such as clinical notes, drug prescriptions, and diagnosis codes to develop predictive models [54–60]. These models have been trained to predict future onset of dementia [53–56] or diagnose undetected dementia [57,58,60] with varying levels of accuracy and can potentially serve as case-finding algorithms to target high-risk patients with further clinical assessments to confirm dementia diagnosis [58]. However, these models are contingent on the availability of consolidated electronic health records, sufficient health care interactions by the patient, and correctly transcribed notes and diagnosis codes [55,57,61]. In contrast, the model developed here does not use data extracted from electronic health records, rather it predicts dementia severity based on self-reported caregiver data and can be used to potentially unlock insights from online self-reported symptoms.

Few studies have used machine learning models to stage the severity of dementia or differentiate types of dementia [62]. One such study uses a combination of cognitive function tests and clinicians' assessments of patients to assess dementia severity on the Clinical Dementia Rating Scale [63]. On the other hand, a combination of neuropsychiatric assessment, mental status examination, and laboratory investigations have also been used to classify dementia severity with a high degree of accuracy [64]. Such approaches require trained interviewers and clinician assessment to obtain input data for the predictive models. This is in contrast to the model developed here, which is designed to stage dementia severity based on self-reported data thereby potentially offering a more economically viable screening tool for dementia severity.

Even though our sample size ($n=717$) is relatively small, it is larger than that of other machine learning studies in dementia, except for a 2019 report that used administrative data to diagnose incident dementia [47]. The advantage of utilizing patient reported outcomes such as SymptomGuide is that it reflects the lived experience of the patient or caregiver and focuses on what is meaningful to them. It is easier to source and computationally less expensive to train models when compared to imaging data or complex biomarkers [48,49]. Interestingly, Chiu et al [50] reported that a machine learning algorithm could be used to derive a screening instrument to distinguish normal cognition, mild cognitive impairment, and dementia. This further emphasizes that dementia symptoms can be used with machine learning to characterize various stages of dementia. On the other hand, our approach used a patient-derived library of symptoms to train a machine learning model, whereas Chui et al used machine learning to reduce the dimensionality of their screening instrument [50]. It is likely that, given the high dimensionality of late-life dementia, different machine learning approaches may be useful in dementia research. In our earlier work, we developed a model based on a neural network trained on 320 symptom profiles

reported by caregivers of persons with dementia [24]. This study expands on our previous work by increasing the sample size and diversity of the training data. We also examined the performance of multiple machine learning algorithms on the available data to maximize our interpretation. The support vector machine outperformed the neural network approach, highlighting the advantage of the current approach.

Future studies could include integrating the model developed here with an electronic interface by which end users could build a symptom profile and obtain the dementia stage. This instrument also has the potential to facilitate physician-patient discussions or to aid screening patients before their in-person memory clinic visit. This model can potentially be applied on other web-based data sets that contain symptom profiles of persons affected with dementia.

The model presented here can classify dementia stages from individualized symptom data. This real-world evidence will enable us to better understand the symptoms that matter most to people affected by dementia at each dementia stage. That information can greatly expand access to understanding the lived experience of dementia.

Conflicts of Interest

These analyses were conducted by DGI Clinical Inc. KR is president, chief science officer, and a shareholder in DGI Clinical Inc (formerly DementiaGuide Inc). In the last 3 years, KR has also sat on an advisory board for Roche/Genentech and Nutricia and has given 2 talks sponsored by Nutricia. AS, JS, SEH and TD are employees of DGI Clinical Inc.

References

1. Kirova A, Bays RB, Lagalwar S. Working memory and executive function decline across normal aging, mild cognitive impairment, and Alzheimer's disease. *Biomed Res Int* 2015;2015:748212 [FREE Full text] [doi: [10.1155/2015/748212](https://doi.org/10.1155/2015/748212)] [Medline: [26550575](https://pubmed.ncbi.nlm.nih.gov/26550575/)]
2. Rockwood K, Sanon Aigbogun M, Stanley J, Wong H, Dunn T, Chapman CAT, et al. The symptoms targeted for monitoring in a web-based tracking tool by caregivers of people with dementia and agitation: cross-sectional study. *J Med Internet Res* 2019 Jun 28;21(6):e13360 [FREE Full text] [doi: [10.2196/13360](https://doi.org/10.2196/13360)] [Medline: [31254339](https://pubmed.ncbi.nlm.nih.gov/31254339/)]
3. Auer SR, Höfler M, Linsmayer E, Beránková A, Prieschl D, Ratajczak P, et al. Cross-sectional study of prevalence of dementia, behavioural symptoms, mobility, pain and other health parameters in nursing homes in Austria and the Czech Republic: results from the DEMDATA project. *BMC Geriatr* 2018 Aug 13;18(1):178 [FREE Full text] [doi: [10.1186/s12877-018-0870-8](https://doi.org/10.1186/s12877-018-0870-8)] [Medline: [30103672](https://pubmed.ncbi.nlm.nih.gov/30103672/)]
4. Rockwood K, Mitnitski A, Richard M, Kurth M, Kesslak P, Abushakra S. Neuropsychiatric symptom clusters targeted for treatment at earlier versus later stages of dementia. *Int J Geriatr Psychiatry* 2015 Apr;30(4):357-367. [doi: [10.1002/gps.4136](https://doi.org/10.1002/gps.4136)] [Medline: [24798635](https://pubmed.ncbi.nlm.nih.gov/24798635/)]
5. Reeve E, Molin P, Hui A, Rockwood K. Exploration of verbal repetition in people with dementia using an online symptom-tracking tool. *Int Psychogeriatr* 2017 Dec;29(6):959-966 [FREE Full text] [doi: [10.1017/S1041610216002180](https://doi.org/10.1017/S1041610216002180)] [Medline: [28274302](https://pubmed.ncbi.nlm.nih.gov/28274302/)]
6. McGarrigle L, Howlett SE, Wong H, Stanley J, Rockwood K. Characterizing the symptom of misplacing objects in people with dementia: findings from an online tracking tool. *Int Psychogeriatr* 2019 Nov;31(11):1635-1641. [doi: [10.1017/S104161021800220X](https://doi.org/10.1017/S104161021800220X)] [Medline: [30698122](https://pubmed.ncbi.nlm.nih.gov/30698122/)]
7. El Haj M, Roche J, Jardri R, Kapogiannis D, Gallouj K, Antoine P. Clinical and neurocognitive aspects of hallucinations in Alzheimer's disease. *Neurosci Biobehav Rev* 2017 Dec;83:713-720 [FREE Full text] [doi: [10.1016/j.neubiorev.2017.02.021](https://doi.org/10.1016/j.neubiorev.2017.02.021)] [Medline: [28235545](https://pubmed.ncbi.nlm.nih.gov/28235545/)]
8. Foguem C, Manckoundia P. Lewy body disease: clinical and pathological. *Curr Neurol Neurosci Rep* 2018 Apr 08;18(5):24. [doi: [10.1007/s11910-018-0835-5](https://doi.org/10.1007/s11910-018-0835-5)] [Medline: [29629495](https://pubmed.ncbi.nlm.nih.gov/29629495/)]
9. Woods B, Arosio F, Diaz A, Gove D, Holmerová I, Kinnaird L, et al. Timely diagnosis of dementia? family carers' experiences in 5 European countries. *Int J Geriatr Psychiatry* 2019 Jan;34(1):114-121 [FREE Full text] [doi: [10.1002/gps.4997](https://doi.org/10.1002/gps.4997)] [Medline: [30246266](https://pubmed.ncbi.nlm.nih.gov/30246266/)]
10. Knopman DS, Petersen RC. Mild cognitive impairment and mild dementia: a clinical perspective. *Mayo Clin Proc* 2014 Oct;89(10):1452-1459 [FREE Full text] [doi: [10.1016/j.mayocp.2014.06.019](https://doi.org/10.1016/j.mayocp.2014.06.019)] [Medline: [25282431](https://pubmed.ncbi.nlm.nih.gov/25282431/)]

11. Cook C, Fay S, Rockwood K. Verbal repetition in people with mild-to-moderate Alzheimer Disease: a descriptive analysis from the VISTA clinical trial. *Alzheimer Dis Assoc Disord* 2009;23(2):146-151. [doi: [10.1097/WAD.0b013e318193cbef](https://doi.org/10.1097/WAD.0b013e318193cbef)] [Medline: [19474572](https://pubmed.ncbi.nlm.nih.gov/19474572/)]
12. Hamilton L, Fay S, Rockwood K. Misplacing objects in mild to moderate Alzheimer's disease: a descriptive analysis from the VISTA clinical trial. *J Neurol Neurosurg Psychiatry* 2009 Sep;80(9):960-965. [doi: [10.1136/jnnp.2008.166801](https://doi.org/10.1136/jnnp.2008.166801)] [Medline: [19293172](https://pubmed.ncbi.nlm.nih.gov/19293172/)]
13. Rockwood K, Fay S, Jarrett P, Asp E. Effect of galantamine on verbal repetition in AD: a secondary analysis of the VISTA trial. *Neurology* 2007 Apr 3;68(14):1116-1121. [doi: [10.1212/01.wnl.0000258661.61577.b7](https://doi.org/10.1212/01.wnl.0000258661.61577.b7)] [Medline: [17404193](https://pubmed.ncbi.nlm.nih.gov/17404193/)]
14. Lopez OL, Becker JT, Sweet RA, Klunk W, Kaufer DI, Saxton J, et al. Psychiatric symptoms vary with the severity of dementia in probable Alzheimer's disease. *JNP* 2003 Aug;15(3):346-353. [doi: [10.1176/jnp.15.3.346](https://doi.org/10.1176/jnp.15.3.346)]
15. Reisberg B, Ferris S, de Leon MJ, Crook T. The Global Deterioration Scale for assessment of primary degenerative dementia. *Am J Psychiatry* 1982 Sep;139(9):1136-1139. [doi: [10.1176/ajp.139.9.1136](https://doi.org/10.1176/ajp.139.9.1136)] [Medline: [7114305](https://pubmed.ncbi.nlm.nih.gov/7114305/)]
16. Sclan SG, Reisberg B. Functional Assessment Staging (FAST) in Alzheimer's disease: reliability, validity, and ordinality. *Int. Psychogeriatr* 2005 Jan 07;4(3):55-69. [doi: [10.1017/s1041610292001157](https://doi.org/10.1017/s1041610292001157)]
17. Stern Y, Albert SM, Sano M, Richards M, Miller L, Folstein M, et al. Assessing patient dependence in Alzheimer's disease. *Journal of Gerontology* 1994 Sep 01;49(5):M216-M222. [doi: [10.1093/geronj/49.5.m216](https://doi.org/10.1093/geronj/49.5.m216)]
18. Hughes CP, Berg L, Danziger W, Coben LA, Martin RL. A new clinical scale for the staging of dementia. *Br J Psychiatry* 2018 Jan 29;140(6):566-572. [doi: [10.1192/bjp.140.6.566](https://doi.org/10.1192/bjp.140.6.566)]
19. Wang H, Chen D, Yu H, Chen Y. Forecasting the incidence of dementia and dementia-related outpatient visits With Google Trends: evidence from Taiwan. *J Med Internet Res* 2015 Nov 19;17(11):e264. [doi: [10.2196/jmir.4516](https://doi.org/10.2196/jmir.4516)]
20. Snyder PJ, Kahle-Wroblewski K, Brannan S, Miller DS, Schindler RJ, DeSanti S, et al. Assessing cognition and function in Alzheimer's disease clinical trials: do we have the right tools? *Alzheimer's & Dementia* 2014 Nov 01;10(6):853-860. [doi: [10.1016/j.jalz.2014.07.158](https://doi.org/10.1016/j.jalz.2014.07.158)]
21. Yuan Q, Tan TH, Wang P, Devi F, Ong HL, Abidin E, et al. Staging dementia based on caregiver reported patient symptoms: Implications from a latent class analysis. *PLoS ONE* 2020 Jan 15;15(1):e0227857. [doi: [10.1371/journal.pone.0227857](https://doi.org/10.1371/journal.pone.0227857)]
22. Livingston G, Sommerlad A, Orgeta V, Costafreda SG, Huntley J, Ames D, et al. Dementia prevention, intervention, and care. *Lancet* 2017 Dec 16;390(10113):2673-2734. [doi: [10.1016/S0140-6736\(17\)31363-6](https://doi.org/10.1016/S0140-6736(17)31363-6)] [Medline: [28735855](https://pubmed.ncbi.nlm.nih.gov/28735855/)]
23. Rockwood K. An individualized approach to tracking and treating Alzheimer's disease. *Clin Pharmacol Ther* 2010 Oct;88(4):446-449. [doi: [10.1038/clpt.2010.68](https://doi.org/10.1038/clpt.2010.68)]
24. Rockwood K, Richard M, Leibman C, Mucha L, Mitnitski A. Staging dementia from symptom profiles on a care partner website. *J Med Internet Res* 2013 Aug 07;15(8):e145. [doi: [10.2196/jmir.2461](https://doi.org/10.2196/jmir.2461)]
25. Rockwood JK, Richard M, Garden K, Hominick K, Mitnitski A, Rockwood K. Precipitating and predisposing events and symptoms for admission to assisted living or nursing home care. *Can Geriatr J* 2014 Mar;17(1):16-21 [FREE Full text] [doi: [10.5770/cgj.17.93](https://doi.org/10.5770/cgj.17.93)] [Medline: [24596590](https://pubmed.ncbi.nlm.nih.gov/24596590/)]
26. Rockwood K, Mitnitski A, Black SE, Richard M, Defoy I, VASPECT study investigators. Cognitive change in donepezil treated patients with vascular or mixed dementia. *Can J Neurol Sci* 2013 Jul;40(4):564-571. [doi: [10.1017/s0317167100014670](https://doi.org/10.1017/s0317167100014670)] [Medline: [23786741](https://pubmed.ncbi.nlm.nih.gov/23786741/)]
27. Foster JR, Sclan S, Welkowitz J, Boksay I, Seeland I. Psychiatric assessment in medical long-term care facilities: Reliability of commonly used rating scales. *Int J Geriatr Psychiatry* 1988 Jul;3(3):229-233. [doi: [10.1002/gps.930030310](https://doi.org/10.1002/gps.930030310)]
28. Auer S, Reisberg B. The GDS/FAST staging system. *Int Psychogeriatr* 1997;9 Suppl 1:167-171. [doi: [10.1017/s1041610297004869](https://doi.org/10.1017/s1041610297004869)] [Medline: [9447440](https://pubmed.ncbi.nlm.nih.gov/9447440/)]
29. Richard M, Rockwood K, Mitnitski A. O2-13-06: Symptom profiles in relation to dementia staging. *Alzheimer's & Dementia* 2012 Jul 01;8(4S_Part_7):P263-P263. [doi: [10.1016/j.jalz.2012.05.701](https://doi.org/10.1016/j.jalz.2012.05.701)]
30. Han H, Wang WY, Mao BH. Borderline-SMOTE: A New Over-Sampling Method in Imbalanced Data Sets Learning. In: *Proceedings of the 2005 International Conference on Advances in Intelligent Computing - Volume Part I*. Berlin, Heidelberg: Springer-Verlag; 2005 Presented at: ICIC'05; August 23-26, 2005; Hefei, China p. 878-887. [doi: [10.1007/11538059_91](https://doi.org/10.1007/11538059_91)]
31. Laurikkala J. Improving Identification of Difficult Small Classes by Balancing Class Distribution. In: *Proceedings of the 8th Conference on AI in Medicine in Europe: Artificial Intelligence Medicine.*: Springer; 2001 Jun Presented at: Artificial Intelligence in Medicine in Europe. AIME; July 1-4, 2001; Cascais, Portugal p. 63-66. [doi: [10.1007/3-540-48229-6_9](https://doi.org/10.1007/3-540-48229-6_9)]
32. Mosley L. A balanced approach to the multi-class imbalance problem. Graduate Theses and Dissertations Iowa State University. Ames, Iowa; 2013. URL: <https://lib.dr.iastate.edu/cgi/viewcontent.cgi?referer=&httpsredir=1&article=4544&context=etd> [accessed 2020-11-06]
33. Cawley GC, Talbot NL. On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation. *Journal of Machine Learning Research* 2010 Mar 01;11:2079-2107. [doi: [10.5555/1756006.1859921](https://doi.org/10.5555/1756006.1859921)]
34. Krstajic D, Buturovic LJ, Leahy DE, Thomas S. Cross-validation pitfalls when selecting and assessing regression and classification models. *J Cheminform* 2014 Mar 29;6(1):10 [FREE Full text] [doi: [10.1186/1758-2946-6-10](https://doi.org/10.1186/1758-2946-6-10)] [Medline: [24678909](https://pubmed.ncbi.nlm.nih.gov/24678909/)]
35. Varma S, Simon R. Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics* 2006;7:91 [FREE Full text] [doi: [10.1186/1471-2105-7-91](https://doi.org/10.1186/1471-2105-7-91)] [Medline: [16504092](https://pubmed.ncbi.nlm.nih.gov/16504092/)]

36. Ojala M, Garriga G. Permutation tests for studying classifier performance. In: *Journal of Machine Learning Research*. USA: IEEE Computer Society; 2009 Dec Presented at: ICDM IEEE International Conference on Data Mining; Dec. 6 - Dec. 9 2009; Miami, Florida p. 908-913. [doi: [10.1109/ICDM.2009.108](https://doi.org/10.1109/ICDM.2009.108)]
37. Golland P, Fischl B. Permutation tests for classification: towards statistical significance in image-based studies. *Inf Process Med Imaging* 2003 Jul;18:330-341. [doi: [10.1007/978-3-540-45087-0_28](https://doi.org/10.1007/978-3-540-45087-0_28)] [Medline: [15344469](https://pubmed.ncbi.nlm.nih.gov/15344469/)]
38. van der Walt S, Colbert SC, Varoquaux G. The NumPy Array: A Structure for Efficient Numerical Computation. *Comput Sci Eng* 2011 Mar;13(2):22-30. [doi: [10.1109/MCSE.2011.37](https://doi.org/10.1109/MCSE.2011.37)]
39. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* 2020 Feb 3;17(3):261-272. [doi: [10.1038/s41592-019-0686-2](https://doi.org/10.1038/s41592-019-0686-2)]
40. Hunter JD. Matplotlib: a 2D graphics environment. *Comput Sci Eng* 2007 May;9(3):90-95. [doi: [10.1109/mcse.2007.55](https://doi.org/10.1109/mcse.2007.55)]
41. McKinney W. Data Structures for Statistical Computing in Python. In: *Proceedings of the 9th Python in Science Conference*. 2010 Presented at: Python in Science Conference; June 28 - July 3; Austin, Texas p. 51-56. [doi: [10.25080/majora-92bf1922-00a](https://doi.org/10.25080/majora-92bf1922-00a)]
42. Van Rossum G, Drake FL. *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace; 2009.
43. Varoquaux G, Buitinck L, Louppe G, Grisel O, Pedregosa F, Mueller A. Scikit-learn. *GetMobile: Mobile Comp and Comm* 2015 Jun;19(1):29-33. [doi: [10.1145/2786984.2786995](https://doi.org/10.1145/2786984.2786995)]
44. Lemaître G, Nogueira F, Aridas CK. Imbalanced-Learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning. *J Mach. Learn. Res* 2017 Jan;18(1):559-563 [FREE Full text]
45. Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS One* 2015;10(3):e0118432 [FREE Full text] [doi: [10.1371/journal.pone.0118432](https://doi.org/10.1371/journal.pone.0118432)] [Medline: [25738806](https://pubmed.ncbi.nlm.nih.gov/25738806/)]
46. Gauthier S, Reisberg B, Zaudig M, Petersen RC, Ritchie K, Broich K, International Psychogeriatric Association Expert Conference on mild cognitive impairment. Mild cognitive impairment. *Lancet* 2006 Apr 15;367(9518):1262-1270. [doi: [10.1016/S0140-6736\(06\)68542-5](https://doi.org/10.1016/S0140-6736(06)68542-5)] [Medline: [16631882](https://pubmed.ncbi.nlm.nih.gov/16631882/)]
47. Nori VS, Hane CA, Crown WH, Au R, Burke WJ, Sanghavi DM, et al. Machine learning models to predict onset of dementia: A label learning approach. *Alzheimers Dement (N Y)* 2019;5:918-925 [FREE Full text] [doi: [10.1016/j.trci.2019.10.006](https://doi.org/10.1016/j.trci.2019.10.006)] [Medline: [31879701](https://pubmed.ncbi.nlm.nih.gov/31879701/)]
48. Zheng Y, Guo H, Zhang L, Wu J, Li Q, Lv F. Machine learning-based framework for differential diagnosis between vascular dementia and Alzheimer's disease using structural MRI features. *Front Neurol* 2019;10:1097 [FREE Full text] [doi: [10.3389/fneur.2019.01097](https://doi.org/10.3389/fneur.2019.01097)] [Medline: [31708854](https://pubmed.ncbi.nlm.nih.gov/31708854/)]
49. Shigemizu D, Akiyama S, Asanomi Y, Boroevich KA, Sharma A, Tsunoda T, et al. A comparison of machine learning classifiers for dementia with Lewy bodies using miRNA expression data. *BMC Med Genomics* 2019 Oct 30;12(1):150 [FREE Full text] [doi: [10.1186/s12920-019-0607-3](https://doi.org/10.1186/s12920-019-0607-3)] [Medline: [31666070](https://pubmed.ncbi.nlm.nih.gov/31666070/)]
50. Chiu P, Tang H, Wei C, Zhang C, Hung G, Zhou W. NMD-12: A new machine-learning derived screening instrument to detect mild cognitive impairment and dementia. *PLoS One* 2019;14(3):e0213430 [FREE Full text] [doi: [10.1371/journal.pone.0213430](https://doi.org/10.1371/journal.pone.0213430)] [Medline: [30849106](https://pubmed.ncbi.nlm.nih.gov/30849106/)]
51. Zhou T, Thung KH, Zhu X, Shen D. Feature learning and fusion of multimodality neuroimaging and genetic data for multi-status dementia diagnosis. *Mach Learn Med Imaging* 2017 Sep;10541:132-140 [FREE Full text] [doi: [10.1007/978-3-319-67389-9_16](https://doi.org/10.1007/978-3-319-67389-9_16)] [Medline: [29376149](https://pubmed.ncbi.nlm.nih.gov/29376149/)]
52. Davatzikos C. Machine learning in neuroimaging: progress and challenges. *NeuroImage* 2019 Aug;197:652-656. [doi: [10.1016/j.neuroimage.2018.10.003](https://doi.org/10.1016/j.neuroimage.2018.10.003)]
53. Hane CA, Nori VS, Crown WH, Sanghavi DM, Bleicher P. Predicting onset of dementia using clinical notes and machine learning: case-control study. *JMIR Med Inform* 2020 Jun 03;8(6):e17819 [FREE Full text] [doi: [10.2196/17819](https://doi.org/10.2196/17819)] [Medline: [32490841](https://pubmed.ncbi.nlm.nih.gov/32490841/)]
54. Nori VS, Hane CA, Martin DC, Kravetz AD, Sanghavi DM. Identifying incident dementia by applying machine learning to a very large administrative claims dataset. *PLoS ONE* 2019 Jul 5;14(7):e0203246. [doi: [10.1371/journal.pone.0203246](https://doi.org/10.1371/journal.pone.0203246)]
55. Ben Miled Z, Haas K, Black CM, Khandker RK, Chandrasekaran V, Lipton R, et al. Predicting dementia with routine care EMR data. *Artif Intell Med* 2020 Jan;102:101771. [doi: [10.1016/j.artmed.2019.101771](https://doi.org/10.1016/j.artmed.2019.101771)] [Medline: [31980108](https://pubmed.ncbi.nlm.nih.gov/31980108/)]
56. Park JH, Cho HE, Kim JH, Wall MM, Stern Y, Lim H, et al. Machine learning prediction of incidence of Alzheimer's disease using large-scale administrative health data. *NPJ Digit Med* 2020;3:46 [FREE Full text] [doi: [10.1038/s41746-020-0256-0](https://doi.org/10.1038/s41746-020-0256-0)] [Medline: [32258428](https://pubmed.ncbi.nlm.nih.gov/32258428/)]
57. Barnes DE, Zhou J, Walker RL, Larson EB, Lee SJ, Boscardin WJ, et al. Development and validation of eRADAR: a tool using EHR data to detect unrecognized dementia. *J Am Geriatr Soc* 2020 Jan;68(1):103-111. [doi: [10.1111/jgs.16182](https://doi.org/10.1111/jgs.16182)] [Medline: [31612463](https://pubmed.ncbi.nlm.nih.gov/31612463/)]
58. Shao Y, Zeng QT, Chen KK, Shutes-David A, Thielke SM, Tsuang DW. Detection of probable dementia cases in undiagnosed patients using structured and unstructured electronic health records. *BMC Med Inform Decis Mak* 2019 Jul 09;19(1):128 [FREE Full text] [doi: [10.1186/s12911-019-0846-4](https://doi.org/10.1186/s12911-019-0846-4)] [Medline: [31288818](https://pubmed.ncbi.nlm.nih.gov/31288818/)]

59. Mar J, Gorostiza A, Ibarrondo O, Cernuda C, Arrospide A, Iruin Á, et al. Validation of random forest machine learning models to predict dementia-related neuropsychiatric symptoms in real-world data. *J Alzheimers Dis* 2020;77(2):855-864. [doi: [10.3233/JAD-200345](https://doi.org/10.3233/JAD-200345)] [Medline: [32741825](https://pubmed.ncbi.nlm.nih.gov/32741825/)]
60. Ford E, Rooney P, Oliver S, Hoile R, Hurley P, Banerjee S, et al. Identifying undetected dementia in UK primary care patients: a retrospective case-control study comparing machine-learning and standard epidemiological approaches. *BMC Med Inform Decis Mak* 2019 Dec 02;19(1):248 [FREE Full text] [doi: [10.1186/s12911-019-0991-9](https://doi.org/10.1186/s12911-019-0991-9)] [Medline: [31791325](https://pubmed.ncbi.nlm.nih.gov/31791325/)]
61. Zhu CW, Ornstein KA, Cosentino S, Gu Y, Andrews H, Stern Y. Misidentification of dementia in Medicare claims and related costs. *J Am Geriatr Soc* 2019 Feb;67(2):269-276 [FREE Full text] [doi: [10.1111/jgs.15638](https://doi.org/10.1111/jgs.15638)] [Medline: [30315744](https://pubmed.ncbi.nlm.nih.gov/30315744/)]
62. Pellegrini E, Ballerini L, Hernandez MDCV, Chappell FM, González-Castro V, Anblagan D, et al. Machine learning of neuroimaging for assisted diagnosis of cognitive impairment and dementia: A systematic review. *Alzheimers Dement (Amst)* 2018;10:519-535 [FREE Full text] [doi: [10.1016/j.dadm.2018.07.004](https://doi.org/10.1016/j.dadm.2018.07.004)] [Medline: [30364671](https://pubmed.ncbi.nlm.nih.gov/30364671/)]
63. Shankle WR, Mania S, Dick MB, Pazzani MJ. Simple models for estimating dementia severity using machine learning. *Stud Health Technol Inform* 1998;52 Pt 1:472-476. [Medline: [10384501](https://pubmed.ncbi.nlm.nih.gov/10384501/)]
64. Joshi S, GG VS, P DS, KR V, LM P. Classification and treatment of different stages of Alzheimer's disease using various machine learning methods. *Int J of Bioi Res* 2010 Jun 30;2(1):44-52. [doi: [10.9735/0975-3087.2.1.44-52](https://doi.org/10.9735/0975-3087.2.1.44-52)]

Abbreviations

AUROC: area under the receiver operating characteristic curve

AUPRC: area under the precision-recall curve

FAST: Functional Assessment Staging Test

GDS: Global Deterioration Scale

Edited by G Eysenbach; submitted 01.06.20; peer-reviewed by L McGarrigle, P Bleicher; comments to author 31.07.20; revised version received 17.08.20; accepted 24.10.20; published 11.11.20

Please cite as:

Shehzad A, Rockwood K, Stanley J, Dunn T, Howlett SE

Use of Patient-Reported Symptoms from an Online Symptom Tracking Tool for Dementia Severity Staging: Development and Validation of a Machine Learning Approach

J Med Internet Res 2020;22(11):e20840

URL: <http://www.jmir.org/2020/11/e20840/>

doi: [10.2196/20840](https://doi.org/10.2196/20840)

PMID: [33174853](https://pubmed.ncbi.nlm.nih.gov/33174853/)

©Aaqib Shehzad, Kenneth Rockwood, Justin Stanley, Taylor Dunn, Susan E Howlett. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 11.11.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.