DEVELOPMENT OF A MACHINE LEARNING ALGORITHM TO CLASSIFY DEMENTIA STAGE BASED ON REPORTED DEMENTIA SYMPTOMS

Rockwood K[1,2], Shehzad A[2], Stanley J[2], Dunn T[2], Howlett SE[1,2], Mitnitski AB[1,2], Chapman CAT[2]
(1) Dalhousie University, (2) DGI Clinical Inc.

BACKGROUND

- Dementia symptom menus provide a library from which the general public can identify and track symptoms that are important to them.
- The value of these data from a research standpoint would be enhanced if dementia stage could be identified.

OBJECTIVE

To apply supervised machine learning methods that can identify dementia stage by classifying symptoms reported by users.

METHODS

Sample

- We obtained data on 717 people in whom dementia had been staged in: a memory clinic in Halifax, NS; A long term care (LTC) study[1]; and the VASPECT[2] clinical trial.
- Symptom information was captured with either SymptomGuide® (memory clinic, VASPECT) or with Goal Attainment Scaling.
- Clinical stage was classified using either the Functional Assessment Staging Test or Global Deterioration Scale.

Machine Learning Algorithm

- We required prediction of four levels of cognitive impairment: Mild Cognitive Impairment (MCI), or mild, moderate, or severe dementia.
- Data were split into training (80%) and testing (20%) datasets.
- The algorithm was optimized for balanced accuracy taking into account varying proportions of stages in the data.
- Accuracy was adjudicated and iterated using measures of precision (Cohen's Kappa), sensitivity (recall) and Positive Predicted Value and tested using Area Under the Receiver Operating Characteristic Curve (AUC-ROC).

RESULTS

- Our sample was mostly female (59%) older adults (77.3±10.6 years, range 40-100) with mild-moderate dementia (**Table 1**).
- Patient age and 36 of 55 unique dementia symptoms most accurately distinguished stage.
- A Support Vector Machine (SVM) showed the best performance. The Algorithm successfully identified the correct dementia stage with 81% accuracy (range 70-90%, **Figure 3**).
- Best performance was seen when classifying severe dementia (**Figure 4**).
- Most misclassifications were within one stage of the clinical stage (**Figure 5**).

Table 1. Baseline subject characteristics

Characteristic	Clinic	LTC study[1]	VASPECT[2]	Total
Sample size	420	169	128	717
Age (Mean, SD)	74.6 (12.5)	81.0 (19.1)	75.4 (9.2)	77.3 (10.6)
Sex (% Women)	54.3	76.3	52.3	59.1
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)
Mean, SD Symptoms	4.7 (2.1)	4.6 (2.1)	7.6 (5.0)	5.2 (3.1)

Abbreviations: FAST, Functional Assessment Staging Test; GDS, Global Deterioration Scale; SD, Standard Deviation.

Figure 1. Sample distribution by Clinical Stage

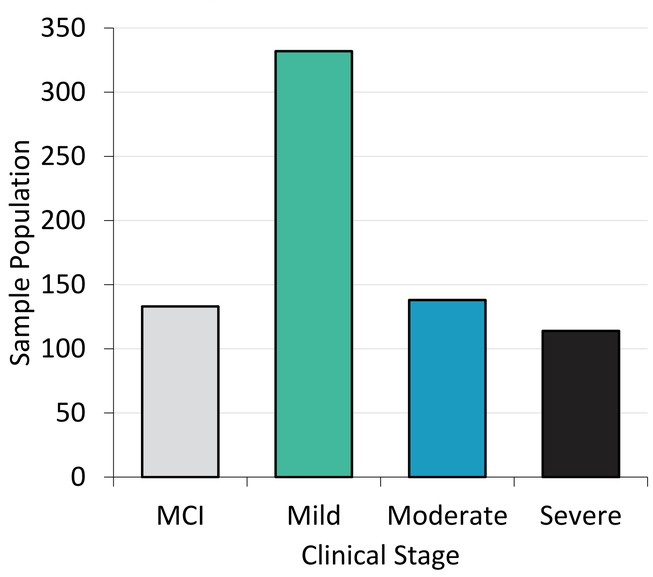


Figure 2. Visualization of two-component Principal Component Analysis

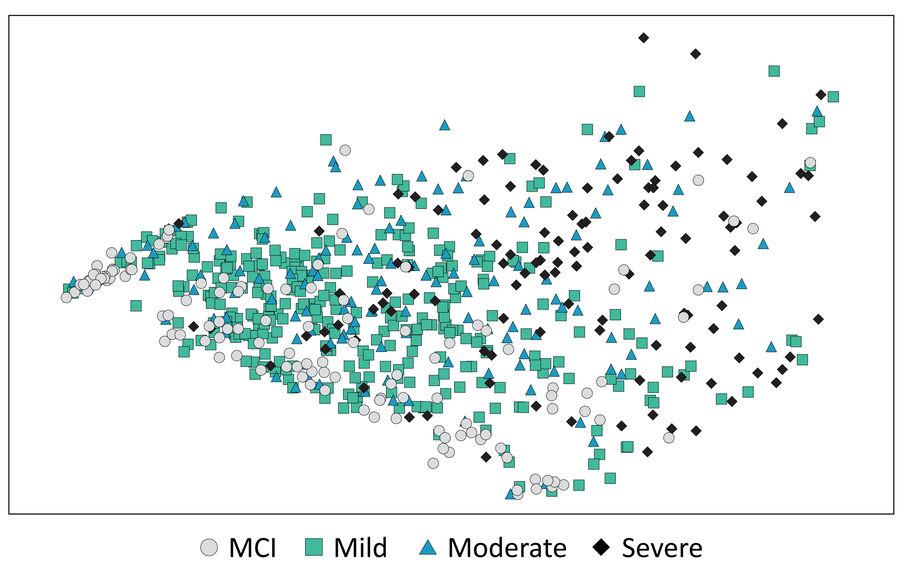


Figure 3. Balanced accuracy over 1000 iterations

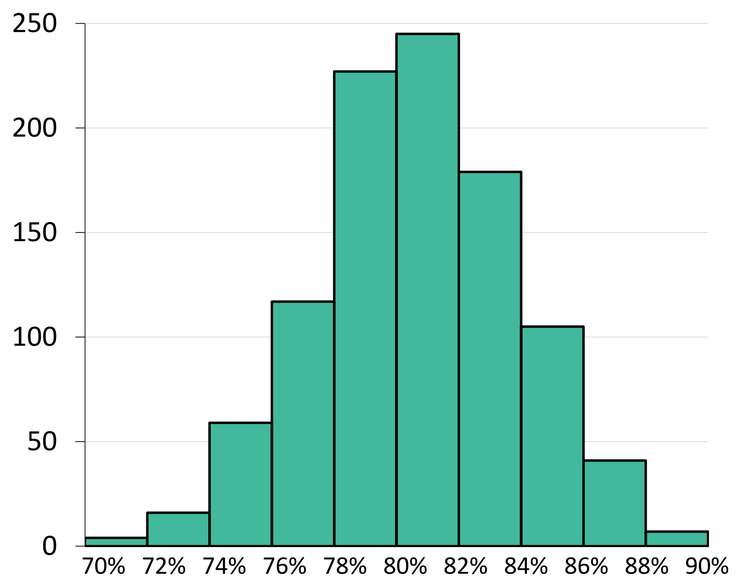


Figure 4. Algorithm performance (AUC-ROC) by classification

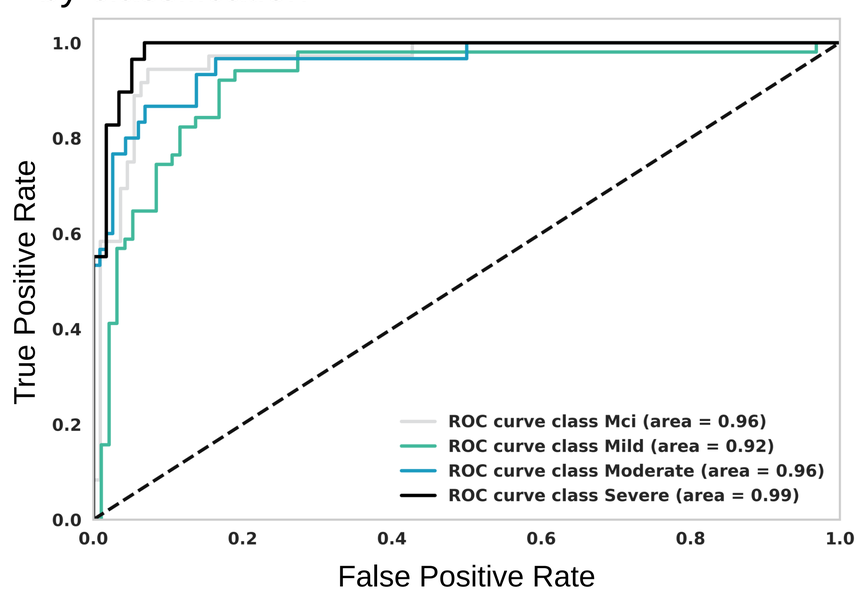
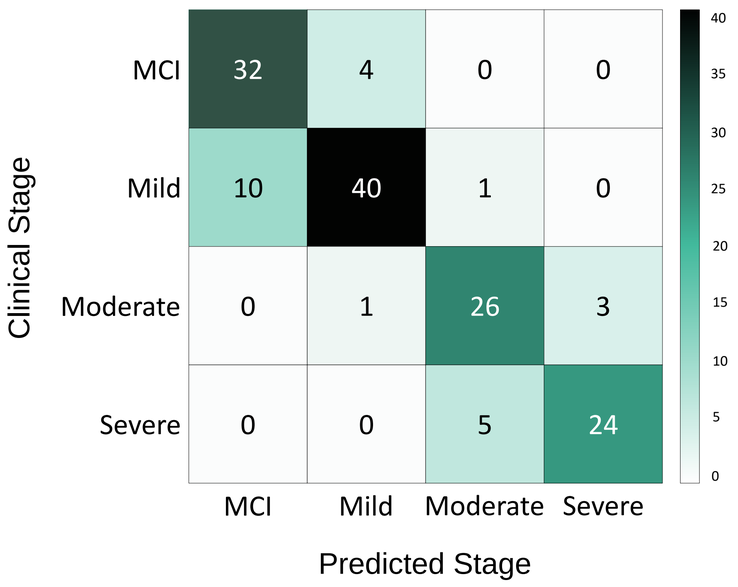


Figure 5. Clinical stage versus predicted stage (testing dataset)



All misclassifications in this example were classified within one stage of the clinical stage. Overall, 97% of misclassifications were within one stage.

CONCLUSIONS

- A supervised machine learning algorithm exhibited excellent performance in identifying dementia stage based on reported dementia symptoms.
- This novel dementia staging algorithm can be used in SymptomGuide® or other similar databases to identify dementia stage based on users' symptom profiles.

References:

1. Rockwood JKH et al. (2013) Precipitating and predisposing events and symptoms for admission to assisted living or nursing home care. *Can Geriatr J.* 17 (1): 16-21.
2. Rockwood K et al. (2013) Cognitive change in donepezil treated patients with vascular or mixed dementia. *Can J Neurol Sci.* 40 (4): 564-71