Multiple Sclerosis Management: Predicting Disease Trajectory of People with Multiple Sclerosis Utilizing Multi-Dimensional Data Including Digital Cognitive Assessments and Patient Reported Outcomes Using Machine Learning Techniques

Mark Gudesblatt MD[1,7,10], Jared Srinivasan[1], Daniel Golan MD[5,6], Myassar Zarif MD[1], Barbara Bumstead ANP[1], Marijean Buhse ANP PhD[1,2], Lori Fafard RN[1], Daniel Kantor MD[7,9,10],Timothy Fratto PhD[3,4], Jeffrey Wilken PhD[3,4], Cynthia Sullivan PhD[3], Johanna Kaplan PhD[13], Glen Doniger PhD[11,12], Amrit Baveja[3], Ben Epstein[3], Monte Zweben[3]

[1]South Shore Neurologic Associates, Islip, NY, USA. 2. Dept. of Nursing Stony Brook University, Stony Brook NY. 3. Washington Neuropsychology Research Group, Fairfax, VA. 4. Department of Neurology, Georgetown University, Washington DC 5. Department of Neurology, Carmel Medical Center, Haifa, Israel, 6. Faculty of Medicine, Technion-Israel Institute of Technology, Haifa, Israel [7]Precision Innovative Network (PIN), New York, NY, USA. [8]Splice Machine, San Francisco, CA, USA. [9]Schmidt College of Medicine, Florida Atlantic University, Boca Raton, FL, USA. [10]Medical Partnership 4 MS (MP4MS), Coconut Creek, FL, USA. 11.NeuroTrax Corporation, Modiin, Israel. 12. Joseph Sagol Neuroscience Center & Center of Advanced Technologies in Rehabilitation, Sheba Medical Center, Ramat Gan. 13. Washington Anxiety Center of Capitol Hill, Washington DC

Background

Multiple Sclerosis (MS) disease impact and progression is traditionally measured by: neuro-radiologist reported MRI changes, patient reported relapse rates and neurological examination determined (EDSS). EDSS does not reflect cognitive impairment (CI) and neurologist detection of CI is neither sensitive nor quantitative. CI, independent of EDSS, can impact patients' employability, ability to drive, fall risk, and Quality of Life (QoL). Patient reported outcomes (PROs) and digital functional measures (cognition, gait) are emerging as more meaningful, patient-centered measures.

Combining multidimensional PRO and quantified digital objective examiner independent disease impact information independent of EDSS might enhance clinical decision making.

Machine learning can help clinicians predict the trajectory of MS using ongoing streams of such population data. Real-time collection of de-identified data can continuously improve predictions and surface the relevant features that predict both treatment and expected PRO outcomes. Supervised machine learning models trained on such comprehensive information, demographic data and digital quantitative objective multi-dimensional batteries can serve to effectively predict and perhaps optimize meaningful and important outcomes as well as perhaps improve treatment selection or timing of such treatment change. As patient-tracking sources expand beyond what is traditionally captured in an office visit, clinicians need tools to effectively integrate and understand these varied streams of data. Machine learning has the potential to not only help identify early critical signals of meaningful change but improve clinicians prediction of patient outcomes from multi-dimensional and quantified data sources to enhance shared decision making to pre-empt development of visible and accumulative meaningful disability.

Objective

To demonstrate the feasibility of predicting clinical outcomes in people with Multiple Sclerosis using standard machine learning methods on multi-dimensional data including objective examiner independent digital cognitive assessments and Patient Reported Outcomes (PROs) that could serve to enhance the Shared Decision Making process.

Methods

- Machine learning models were trained on EHR data, cognitive domain scores and PRO data. A model was created to predict what the patient is likely to report. **70%** of the dataset was used in training, **20%** in testing and **10%** on independent validation sets with an ensemble learning method (random forest classifiers) used to construct a multitude of training decision trees, which then outputted the mean prediction of the individual trees. This would allow a purely independent set of data, never seen by the model to fully verify the accuracy metrics obtained without bias.
- A what-if tool was created that allows a clinician to alter a data field to anticipate a possible future value, and run the model to determine best predictions of PRO score and Disability Milestone. (enhanced decision making)

Results

- The training set cohort consisted of **258** people with Multiple Sclerosis over a three year period at a single MS Center.
- Gender: 72.5% female.
- Average Age: 46.2 ± 10.2 years.

Multiple Sclerosis Disease Impact: Machine Learning Prediction Model
- What will the patient report and experience?

Disability Milestone Patient Reported Outcome (PRO)	Predicted/Classified False negative/False positive F1	Weighted Precision	Weighted Recall	Accuracy
Driving (yes/no)	96% +/- 2%	96% +/- 2%	96% +/- 2%	96% +/- 2%
Modified Falls Efficacy Scale (MFES PRO score)	83% +/- 3%	84% +/-2%	83% +/-2%	83% +/-2%
Depression (Beck) (BDI PRO score)	83% +/-4%	83% +/-4%	83% +/-4%	83% +/-4%
Fatigue (MFIS PRO score)	72% +/-4%	77% +/-4%	72% +/-3%	72% +/-3%
Employment (yes/no)	91% +/-4%	91% +/-4%	91% +/-4%	91% +/-4%
Quality of Life (Likert Scale)	85% +/-3%	86 % +/-2%	85% +/-3%	85% +/-3%

Term	Definition	Dictionary
F1	**2[(Precision-Recall)/(Precision+Recall)]**	**TP** = True Positives **FP** = False Positives **FN** = False Negatives **TN** = True Negative Values **P** = Total Positive Values **N** = Total Negative Values
Precision	**TP/(TP+FP)**	
Recall	**TP/(TP+FN) or TP/P**	
Accuracy	**(TP+TN)/(P+N)**	

Definition of Terms

F1 can be interpreted as the harmonic mean of precision and recall. F1 combines the two measures equally.

Precision can be interpreted as the "positive predictive value" for a classifier. It is calculated as the proportion of all positive classifications that are in fact true positives. Weighted precision calculates the precision per class and then weights that value by the number of instances of each class.

Recall can be interpreted as the "True Positive Rate" or "sensitivity" for a classifier. It is calculated as the proportion of all positive values that are classified as such. Weighted recall calculates the recall per class and then weights that value by the number of instances of each class.

Accuracy can be interpreted as the proportion of correctly classified values. It is calculated as the number of correct classifications divided by the total number of predictions.

Conclusion

Machine learning combined with objective measures of disease impact and PRO can provide important information to predict economically important and disability relevant outcomes, potentially enhancing treatment decisions. These results show promising predictive accuracy to be used in a variety of advisory applications and potentially reduce disease related disability. The results of the other models demonstrate the feasibility of using machine learning in a broader network of clinical sites that will allow for greater accuracy, precision and recall. The eventual goal is that these models can be used as an aid to enhance the shared decision making process, and to reduce both inappropriate healthcare costs and economically impactful and important patient centric disability.