A Clinical Decision Support Tool to Reduce the Need for Magnetic Resonance Imaging for Disease Monitoring in Multiple Sclerosis

Michael Robinette*, Karla Gray-Roncal*, Kathryn Fitzgerald, Kadija Ferryman, Casey Overby Taylor, John Scott, Elias Sotirchos, Peter Calabresi, Ellen M. Mowry, William Gray-Roncal

**Authors contributed equally*

Abstract

People with multiple sclerosis (MS), a chronic neurologic disease, typically have a brain magnetic resonance imaging (MRI) scan annually to make sure the medication they are using is working well. Although with age and given treatment advances, these scans often do not show new MS activity, it is difficult to know when to advise a given person to stop this monitoring algorithm. In this study, leveraging analyses we have already completed, we generated personalized predictions about the utility of brain MRI at a given time point for individuals with MS. The longer-term goals are to 1) create an easily digestible visualization of the results, 2) evaluate how well the predictions work over time, 3) characterize and limit unintentional bias in the algorithm's predictions or deployment, and 4) assess if having the personalized information available reduces unnecessary brain MRIs, thereby improving the value of health care for people with MS.

Introduction

Multiple sclerosis (MS) is a relatively common disorder and a major cause of neurologic disability in early adults, impacting the ability of many to remain employed or engaged in family or personal care.[1-3] The earliest "relapsing" phase of MS, for the vast majority (85-90%) of people with the diagnosis, is characterized by the development of autoimmune-induced focal areas of demyelination throughout the central nervous system (CNS) which can be visible as "lesions" on brain and spinal cord magnetic resonance imaging. When they develop in regions most relevant to daily functioning, they produce new neurological symptoms (e.g., numbness, inability to see out of one eye, or trouble walking), which are known as MS exacerbations, relapses, flare-ups, or attacks. At onset, patients have this form of MS (relapsing remitting MS [RRMS]), but later in their course they often develop slowly-worsening disability (secondary progressive MS [SPMS]), which on the average begins at the age of 45 ± 10 years.[1]

Early treatments to prevent ongoing accrual of damage to the central nervous system can mitigate the risk of MS-related inflammatory activity. There are now more than 20 disease-modifying therapies (DMTs) that prevent the autoimmune “attacks” of MS. Observational studies evaluating early MS activity after starting a first-line, moderate efficacy DMT suggest that the development of a clinical exacerbation or more than one or two new lesions is associated with greater intermediate-term disability and typically is an indication to switch medications[6] to one of higher efficacy.[2] However, the modern treatment landscape for MS includes DMTs with a wide range of efficacy, where some have greater average likelihoods of suppressing new inflammatory activity than others. Further, the trend in treating MS has included early personalized optimization of DMT, such that breakthrough disease activity is increasingly rare.

Monitoring the treatment response to MS DMT has not yet been optimized, particularly in the modern treatment era. Current standard of care recommends imaging patients at routine intervals, which results in costs and patient discomfort that may not be necessary. When a person with MS begins a DMT, neurologists clinically evaluate that person with some periodicity (e.g. every 6 months) to ensure no new clinical exacerbations have occurred; they also repeat MRI of the brain and/or spinal cord to evaluate for new lesions that may have formed in the absence of new clinical symptoms. The appropriate frequency of such MRIs for specific patients or patient subgroups (e.g. older patients, those on higher efficacy DMT) has not been determined. Expert panels have suggested recommendations, but these are broad and intended for all MS patients, without offering algorithms that identify specific patients by incorporating individual patient characteristics (e.g. stability of disease) or thresholds for interval changes.[3,4]

Recent breakthroughs in machine-learning methods and tools allow for a systematic approach to explore various classification methods to optimize performance while emphasizing approaches that are relevant for clinical practice (e.g., interpretability). These algorithms can utilize large amounts of patient data (e.g. prior imaging results, relapse history, medication information) to use for training of their prediction models, with outputs that could potentially be read by providers and used as decision-making aids for clinically-relevant questions.

Our goal was to develop an algorithm to predict the presence of new lesions in a subsequent MRI. We envision that this information can be provided to clinicians through a future clinical decision tool to be used in shared decision-making conversations with patients.

Methods

Longitudinal, Equitable Systemized Imaging Operations for Neuroimmunology (LESION)

The LESION system has three components: data fusion and quality control, patient classification, and a graphical-user interface (GUI) of the classifier output. The data fusion step combines all the data inputs that will serve as variables for the classification algorithm. This step is important to ensure that valid and complete information gets prepared to serve as inputs to the classifier. A logistic regression classifier was selected as the engine of LESION, whose function output is a prediction of new lesions on subsequent MRIs. This prediction, along with a confidence value, is passed on to the GUI portion of the system. The interface then projects the findings for the clinician to review. The sections below explain in more detail the data sources that are merged as well as the classification algorithm used for LESION.

Data Sources

Our tool development includes readily available data from two sources: the MS Performance Test (MSPT) and the MS Smartform. The MSPT is an iPad-based assessment that patients complete when they enter the clinic.[12] This assessment provides insight into a patient's MS status by monitoring physical (e.g., walking speed, dexterity scores) and cognitive (e.g., anxiety, sleep) metrics known to be overrepresented in, and indicative of disease burden among, people with MS. We focused on just three fields: anxiety, depression, and fatigue. These measures are readily available (as subscales of NeuroQoL[5]) and patient-reported, so their collection is not only standard of care at Johns Hopkins but can easily be incorporated by clinicians who may wish to use the tool at other centers where MSPT itself is not standard of care.

The MS Smartform is a tool for MS providers that is collected in a standardized data format and is also available in the EHR. Data used from this repository included relapses, number of lesions on MRI, and DMT category (Figure 1). The MS Smartform has been widely and freely shared through the Epic Foundation and Epic Library and has been adapted to the Cerner EHR;[13] it is in use at several major MS centers in the US and will remain accessible to external clinicians who wish to use LESION.

During the data merging process, we excluded patients that had only had one visit, in order to provide enough historical information for the prediction algorithm. Additionally, we excluded patients who had recently started a higher-efficacy DMT to avoid confounding bias as an

indication to switch to such medication means the patient is at higher risk of developing lesions in the near future.

Classification Approach

The output of the classifier portion of LESION is responsible for predicting if a new lesion will occur (or not) during the next scan. For classification, we use logistic regression to predict the target variable (i.e., new lesion present). Logistic regression was chosen for its ease of use and interpretability, both of which are critical for deployment of tools for clinical use. In our development process, we also explored more complex machine learning methods such as deep-learning-based classifiers, but found the results to be similar with higher computational requirements, less explainability, and less robustness. We used a "5-fold stratified cross-validation" technique, which trains and tests the data by iteratively splitting the set into 80%/20% training/test groups, resulting in each patient encounter being in the test set exactly once. In this classification paradigm, we leverage all of the data in each experiment, but ensure that no data sample appears in both the train and test set for a given experiment. This type of validation allows for training of an ensemble of models which can be used for prediction, and provides a full set of prediction for the patient cohort in aggregate, which enables operating point exploration and performance assessment.

After obtaining the results, we analyzed different decision thresholds for our patient cohort. By default, probabilities from logistic regression predictions are used with a 50% decision threshold. However, the decision threshold can be adjusted, especially in cases where the importance of false negatives or false positives may vary for a particular application or pose patient risk. During our analysis of the data, we identified a threshold that was conservative in selecting patients who are at high risk of developing new lesions (true positives) and should obtain a follow up scan, thus increasing its sensitivity. Because our LESION application is a clinical support tool, we wanted to be confident that patients who would benefit from a scan would be selected while eliminating unnecessary scans, We chose 8% as a threshold based on the distribution of the classification results and expert clinician input (Figure 2). For example, if the classifier predicts a new lesion will occur, LESION will flag a 'high risk" and display the data sources that support this result.

Results

Patient Characteristics

A total of 1045 patients' data were used for development and testing of the classifier in LESION. Their average age was 48 years, and there was a 3:1 female-to-male ratio; 807 (77%) self-identified as White; 191 (18) as Black; and 9 (0.9%) as Asian. Within this cohort, average time since last relapse was 8.7 (SD) years, 468 were on a high efficacy DMT, and 445 were on a moderate one (Table 1). In aggregate, 95 (9%) had a new lesion on follow up MRI (which served as the outcome for the model).

Algorithm Results

With the 8% threshold chosen, we obtained 67 true positives: new lesions predicted on next MRI, correctly chosen; 252 false positives, corresponding to new lesions being predicted incorrectly; 28 false negatives, corresponding to incorrectly missed lesions; and 698 true negatives, indicating an algorithm decision that correctly predicts no new lesion. This corresponds to a precision (i.e., positive predictive value) of 0.21 and a sensitivity (i.e., recall) of 0.705. The algorithm specificity is 0.735. This operating point is chosen to minimize missed lesions (false positives), as this could pose clinical risk. However, different thresholds can be chosen depending on the downstream clinical decision support tool paradigm.

Because all features are normalized, variables with higher-magnitude coefficients can be interpreted as more strongly driving prediction (Figure 3). For variables with high negative coefficients, they are strongly repressive of new lesions (e.g. aggressive DMT followed by higher age). Likewise, variables with high positive coefficients are strong potential drivers of new lesions (e.g. higher number of lesions at last MRI followed by higher number of relapses in last 2 years).

Discussion

Algorithm

Results from our predictive model and prototype visualization show promising results in identifying patients that may be able to take a longer interval between MRI scans. Because of the interpretability of the classifier and the relatively low proportion of missed lesions, this can be straightforwardly integrated into a clinical support tool. We have created an initial mock-up demonstrating how this might be integrated into an electronic health record system, recognizing that significant future development will likely be needed to aid a clinician-patient decision making conversation.

In this initial prototype, we identify patients that should obtain follow up scans as they have a high risk of demonstrating new lesions on the next MRI. But, also as important, we identified a large number of patients that could skip their next MRI with a low-risk of missing a new lesion. If the clinician has already been considering extending the interval between MRI scans, this result may provide additional reassurance or a “second opinion” to support the clinician’s assessment.

While not every patient will decide to skip the MRI, the patient burden and health system cost may be substantial at an individual and health system level (e.g., a typical brain MRI might cost $3500 per scan). In addition to the cost savings, this can target resources for those who need more frequent monitoring. Access to MRI scans can be limited, especially in certain geographic areas, both in the United States and in other countries; furthermore, even when MRIs are available, patients often must wait months to obtain an appointment. It is also important to emphasize the benefit to the patients who can forego frequent scans, especially those with claustrophobia, limited mobility, or chronic back pain, for whom an MRI scan can be a difficult ordeal, or for whom copay burden posits substantial resource strain.

Developing LESION allowed us to study strong predictors of future MRI lesions. Our significant features were congruent with findings from previous studies - patients using higher efficacy DMTs were less likely to have new disease activity on follow up MRI as were patients with stable disease (no recent relapses or lesions) and older age. Other features included in the model did not show significance and may warrant further study, especially as sample sizes increase. More complex models that consider these predictions could be used in addition to or instead of simple logistic regression, as was the primary engine in LESION. During our design process, we explored the use of more complex traditional models (e.g., Random Forest, support vector machines), as well as deep learning and emerging methods but found that they were either inappropriate (e.g., due to volume of data), or achieved similar performance with a loss of interpretability, which can be a major concern for gaining clinician and patient trust. Although we are very concerned with equity and appropriate care for patients from minoritized backgrounds, we chose to not explicitly encode race as a variable for the classifier to promote robustness, but plan to monitor the effect of the classifier on subgroups in implementation and through an audit report.

Implementation Considerations

Our implementation design needs to consider real-life implementation into the EHR and the values that are reported to the clinician. User-center design concepts such as transparency of

information need to be considered. For example, a clinician may need access not only to the recommendation but also the reasons behind the recommendation from the model perspective. This would allow the provider to assess if those aspects are consistent with the patient history and evidence (Figure 4).

Our initial design needs further validation with larger data sets and prospective studies. The threshold that was chosen was based on data from this study and could be different once it is studied prospectively. One component of our approach is to consider both algorithmic and implementation bias as tool implementation is studied. Sociodemographic information is available in the EHR to support these analyses, including patient sex at birth and gender, race, insurance data, and geocode-based determinants of health. A limitation worth considering is that the system relies on data availability from sources that need to manually entered by a clinician. If the MS Smartform is incomplete, algorithm training is affected and results may not be as meaningful, or a result may not be available for a particular patient.

In conclusion, LESION is a pilot decision-making support model to help clinicians triage patients that would benefit from a subsequent surveillance MRI versus those could be good candidates for decreased frequency of imaging. Once refined into a CDS tool, this approach may help support clinician-patient conversations regarding the value of image monitoring based on their individual patient characteristics.

Table 1: Baseline characteristics of patient data used for in LESION development.

Characteristic	
Total Patients, N	1045
Age yrs, mean (SD)	48.1 (9.6)
Race, N (%)	
White	807 (77)
Black	191 (18)
Asian	9 (0.9)
Female:Male ratio	3:1
DMT, N (%)	
Higher Efficacy	468 (44.8)
Moderate Efficacy	445 (42.6)
Patients w/ New Lesions, N (%)	95 (9)
Time since relapse yrs, mean (SD)	8.7 (9.5)

Traditional (moderate efficacy)	Aggressive (high efficacy)
• Glatiramer acetate (Copaxone, Glatopa) • Interferons (Avonex, Betaseron, Rebif, Plegridy) • Teriflunomide • Fumarates (Tecfidera, Vumerity, Bafiertam) • Sphingo-1 phosphate modulators (fingolimod, siponimod, ozanimod, ponesimod)	• Alemtuzumab (Lemtrada) • Natalizumab (Tysabri) • Ofatumumab (Kesimpta) • Ocrelizumab (Ocrevus) • Rituximab • Cladribine (Mavenclad)

Figure 1: Algorithm bins the DMTs into the above traditional (moderate efficacy) and aggressive (high efficacy) groups.

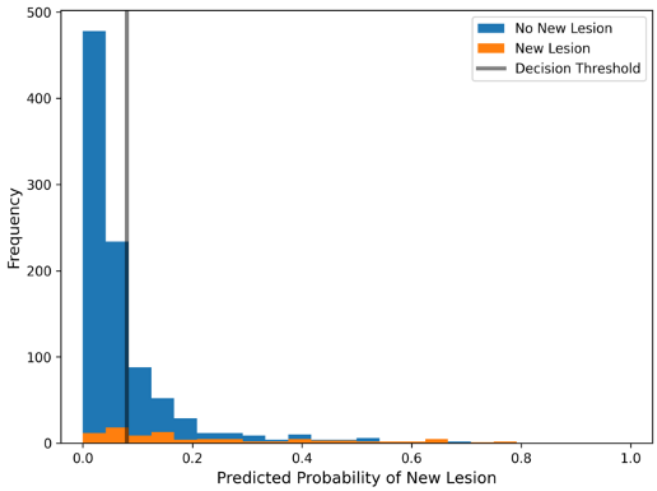

Figure 2: Histogram of lesions and decision thresholds. The patients with and without lesions on their MRI scans are shown. We compare these distributions with the decision threshold chosen. By shifting the decision threshold to the right (e.g., 0.2), we could further reduce the number of scans needed, but would miss a significant number of patients with lesions.

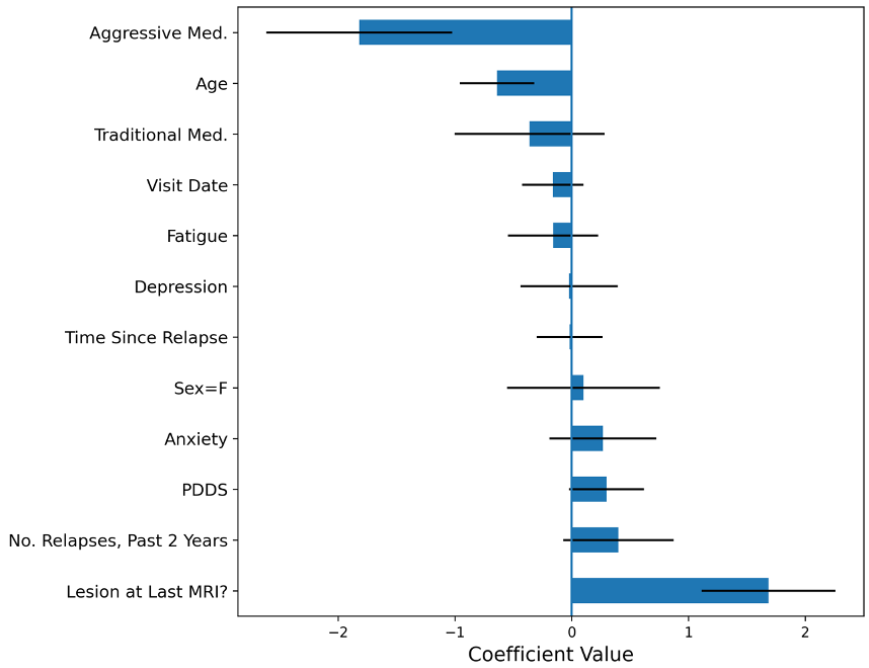

Figure 3: Logistic regression coefficients: We can assess the importance of each feature in predicting the presence of a new lesion.

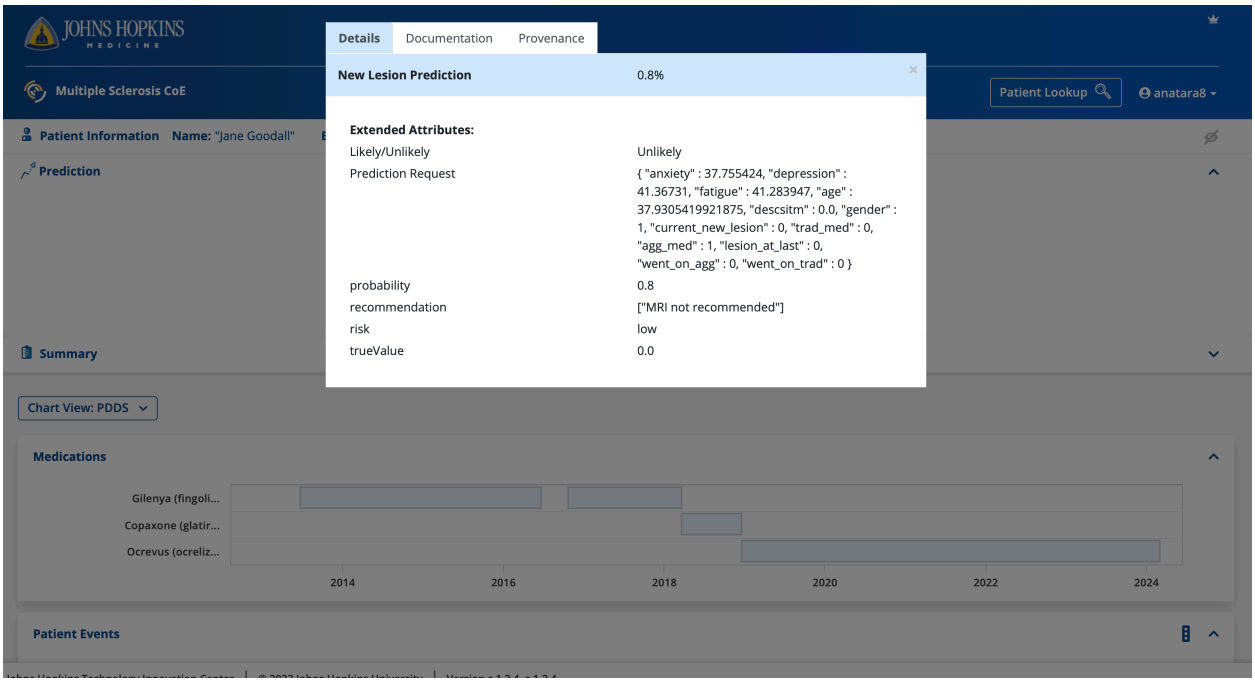

Figure 4: Decision-aid support tool prototype for an example patient, to illustrate how our model may be used in future clinician-patient conversations.

References

1. Zeydan B, Kantarci O. Progressive forms of multiple sclerosis. Distinct entity or age-dependent phenomenon. Neurol Clin 2018;36:163-171.
2. Sormani MP, Gasperini C, Romeo M, Rio J, Calabrese M, Cocco E, Enzingher C, Fazekas F, Filippi M, Gallo A, Kappos L, Marrosu MG, Martinelli V, Prosperini L, Rocca MA, Rovira A, Sprenger T, Stromillo ML, Tedeschi G, Tintorè M, Tortorella C, Trojano M, Montalban X, Pozzilli C, Comi G, De Stefano N; MAGNIMS study group. Assessing response to interferon-β in a multicenter dataset of patients with MS. Neurology 2016;87(2):134-40.
3. Wattjes MP, Ciccarelli O, Reich DS, Banwell B, de Stefano N, Enzinger C, Fazekas F, Filippi M, Frederiksen J, Gasperini C, Hacohen Y, Kappos L, Li DKB, Mankad K, Montalban X, Newsome SD, Oh J, Palace J, Rocca MA, Sastre-Garriga J, Tintoré M, Traboulsee A, Vrenken H, Yousry T, Barkhof F, Rovira À; Magnetic Resonance Imaging in Multiple Sclerosis study group; Consortium of Multiple Sclerosis Centres; North American Imaging in Multiple Sclerosis Cooperative MRI guidelines working group. 2021 MAGNIMS-CMSC-NAIMS consensus recommendations on the use of MRI in patients with multiple sclerosis. Lancet Neurol 2021;20:653-670.
4. Saslow L, Li DKB, Halper J, Banwell B, Barkhof F, Barlow L, Costello K, Damiri P, Dunn J, Giri S, Maes M, Morrow SA, Newsome SD, Oh J, Paul F, Quarterman P, Reich DS, Shewchuk JR, Shinohara RT, Van Hecke W, van de Ven K, Wallin MT, Wolinsky JS, Traboulsee A. An international standardized magnetic resonance imaging protocol for diagnosis and follow-up of patients with multiple sclerosis: advocacy, dissemination, and implementation strategies. Int J MS Care 2020;22:226-232.
5. Medina LD, Torres S, Alvarez E, Valdez B, Nair KV. Patient-reported outcomes in multiple sclerosis: Validation of the Quality of Life in Neurological Disorders (Neuro-QoL™) short forms. Mult Scler J Exp Transl Clin. 2019 Nov 26;5(4):2055217319885986.