


Deep Learning for Predicting Enhancing Lesions in Multiple Sclerosis from Noncontrast MRI

Ponnada A. Narayana, PhD • Ivan Coronado, MS • Sheeba J. Sujit, PhD • Jerry S. Wolinsky, MD • Fred D. Lublin, MD • Refaat E. Gabr, PhD

From the Departments of Diagnostic and Interventional Imaging (P.A.N., I.C., S.J.S., R.E.G.) and Neurology (J.S.W.), McGovern Medical School, University of Texas Health Science Center, 6431 Fannin St, Houston, TX 77030; and Department of Neurology, Icahn School of Medicine at Mount Sinai, New York, NY 10029-6574 (F.D.L.). Received May 9, 2019; revision requested July 2; revision received October 10; accepted October 25. **Address correspondence to P.A.N.** (e-mail: ponnada.a.narayana@uth.tmc.edu).

Funded by the National Institute of Neurological Disorders and Stroke/National Institute of Health (grant 1R56NS105857-01), a Chair in Biomedical Engineering endowment, and the John S. Dunn Foundation.

Conflicts of interest are listed at the end of this article.

Radiology 2020; 294:398–404 • <https://doi.org/10.1148/radiol.2019191061> • Content codes:  

Background: Enhancing lesions on MRI scans obtained after contrast material administration are commonly thought to represent disease activity in multiple sclerosis (MS); it is desirable to develop methods that can predict enhancing lesions without the use of contrast material.

Purpose: To evaluate whether deep learning can predict enhancing lesions on MRI scans obtained without the use of contrast material.

Materials and Methods: This study involved prospective analysis of existing MRI data. A convolutional neural network was used for classification of enhancing lesions on unenhanced MRI scans. This classification was performed for each slice, and the slice scores were combined by using a fully connected network to produce participant-wise predictions. The network input consisted of 1970 multiparametric MRI scans from 1008 patients recruited from 2005 to 2009. Enhanced lesions on postcontrast T1-weighted images served as the ground truth. The network performance was assessed by using fivefold cross-validation. Statistical analysis of the network performance included calculation of lesion detection rates and areas under the receiver operating characteristic curve (AUCs).

Results: MRI scans from 1008 participants (mean age, 37.7 years \pm 9.7; 730 women) were analyzed. At least one enhancing lesion was observed in 519 participants. The sensitivity and specificity averaged across the five test sets were 78% \pm 4.3 and 73% \pm 2.7, respectively, for slice-wise prediction. The corresponding participant-wise values were 72% \pm 9.0 and 70% \pm 6.3. The diagnostic performances (AUCs) were 0.82 \pm 0.02 and 0.75 \pm 0.03 for slice-wise and participant-wise enhancement prediction, respectively.

Conclusion: Deep learning used with conventional MRI identified enhanced lesions in multiple sclerosis from images from unenhanced multiparametric MRI with moderate to high accuracy.

©RSNA, 2019

Multiple sclerosis (MS) is the most common nontraumatic demyelinating neurologic disorder in young adults, affecting at least 2.5 million people worldwide (1). MRI is routinely used for both diagnosis and management of MS (2). A hallmark of MS is the presence of hyperintense lesions on images obtained with T2-weighted, proton density-weighted, and fluid-attenuated inversion recovery (FLAIR) MRI. Not all lesions seen on these images are active. Identification of active lesions is crucial for effective patient treatment (3). It is generally thought that active lesions show enhancement on T1-weighted MRI scans after the administration of gadolinium-based contrast agents (GBCAs). However, there are concerns about GBCA administration, including nephrogenic systemic fibrosis in patients with renal compromise (4) and long-term gadolinium deposition in various tissues (5–8). This is particularly a concern in patients with MS, who undergo frequent imaging with GBCA administration for regular clinical follow-up, which may result in higher cumulative gadolinium deposition in tissues. While acknowledging lack of documented evidence about long-term physiologic

effects of deposited gadolinium in tissues, the U.S. Food and Drug Administration states that “clinicians should limit use of GBCAs to situations where additional information provided by the contrast agent is needed and to assess the necessity of repeat MRIs with GBCAs” (<https://www.fda.gov/drugs/drug-safety-and-availability/fda-drug-safety-communication-fda-warns-gadolinium-based-contrast-agents-gbcas-are-retained-body>). Similar cautionary recommendations were also issued by a number of scientific organizations, such as the Consortium of Multiple Sclerosis Centers (www.mscares.org/mri) (9) and the International Society for Magnetic Resonance in Medicine (10). Alternative methods based on texture analysis, logistic regression, and chemical exchange saturation transfer techniques have been proposed to identify enhancing lesions without the administration of GBCA in patients with MS (11–13). These published studies are from a single center and/or based on a small sample size and require manual identification of the image features.

Deep learning (DL) is a subfield of machine learning that uses multiple nonlinear processing layers for a

Abbreviations

AUC = area under the receiver operating characteristic curve, CombiRx = Combination Therapy in Patients with Relapsing-Remitting Multiple Sclerosis, DL = deep learning, FLAIR = fluid-attenuated inversion recovery, GBCA = gadolinium-based contrast agent, MS = multiple sclerosis

Summary

Deep learning may be a viable alternative to gadolinium-based contrast agents for identifying enhancing lesions in multiple sclerosis on MRI scans.

Key Results

- A convolutional neural network was trained to identify enhancing lesions in multiple sclerosis based on images from unenhanced MRI.
- The convolutional neural network identified enhancing lesions in slices with sensitivity and specificity of 78% and 73%, respectively; the corresponding values for participant-wise prediction were 72% and 70%.
- The average areas under the receiver operating characteristic curve for slice-wise and participant-wise predictions of the deep learning model were 0.82 and 0.75, respectively.

hierarchical representation of the data (14). A unique feature that differentiates DL from classic machine learning methods is that DL can learn image features from the input data without manual identification (15). DL has the potential to identify enhancing lesions without GBCA administration. In fact, DL has been recently used to detect enhancing lesions in MS at a reduced GBCA dose (16). DL, based on a convolutional neural network, is particularly well suited for medical image processing, segmentation, classification, and prediction (17). DL has also been shown to be robust against data heterogeneity and image artifacts (18–20) that are typically encountered in multicenter studies.

The main objective of our study was to investigate the potential of DL in predicting enhancing lesions without GBCA administration. This is a classification, and not a segmentation, issue because our interest was mainly in identifying potentially enhancing lesions—not their volumes.

Materials and Methods

This study was a prospective analysis of MRI data acquired in patients with relapsing-remitting MS who participated in a phase III clinical trial. All data were anonymized. All participating sites received institutional review board approval for imaging of patients. Written informed consent was obtained from all patients. Our institutional review board approved the analysis of the MRI data. This study was fully compliant with the Health Insurance Portability and Accountability Act.

MRI scans were acquired with multiple platforms at 1.5-T (85%) and 3-T (15%) field strengths (GE Medical Systems, Milwaukee, Wis; Philips Healthcare, Best, the Netherlands; Siemens, Erlangen, Germany). The MRI protocol included acquisition of two-dimensional FLAIR (repetition time msec/echo time msec/inversion time msec, 10000/90–100/2400–2600) and two-dimensional dual-echo turbo spin-echo (repetition time msec/first echo time msec/second echo time msec,

6800/12–20/80–100) images. In addition, T1-weighted images (700–800/12–20) were obtained before and after contrast material administration with geometry identical to that with FLAIR and dual-echo turbo spin-echo images. The voxel dimensions for all images were 0.94 mm × 0.94 mm × 3 mm.

Study Participants

MRI scans were acquired in patients with relapsing-remitting MS who participated in a randomized, double-blind, multicenter phase III clinical trial (Combination Therapy in Patients with Relapsing-Remitting Multiple Sclerosis [CombiRx]; clinical trial NCT00211887). Patients were recruited consecutively between 2005 and 2009. Patients aged 18–60 years with an Expanded Disability Status Scale score of 0–5.5 were included. Sixty-eight centers participated in this study (21,22). In this feasibility study, we included images from the baseline examination and the follow-up examination at 6 months. MRI scans in this cohort were also analyzed to determine the regional atrophy (23), effect of inpainting (24), effect of intrinsic and extrinsic factors on cortical thickness (25), and tissue segmentation (26). The current study was based on the same cohort but investigated the prediction of lesional enhancement in MRI without use of contrast material.

The quality of the MRI scans was evaluated by using a procedure described elsewhere (27). Of the original 1008 patients recruited at baseline, 40 were lost to follow-up. Of the 1976 MRI data sets at both time points, six were excluded because of poor signal-to-noise ratio or motion artifacts. Thus, a total of 1970 MRI data sets were included in this study. Among the 1970 postcontrast T1-weighted MRI data sets, experts identified 519 patients with one or more area of enhancement with an enhancing volume of at least 7 voxels, with a total of 1390 enhancing lesions.

Image Preprocessing

All images were preprocessed with an MRI automatic processing pipeline (28–30). The preprocessing steps included skull stripping, bias correction, intensity normalization, and anisotropic diffusion filtering, as described elsewhere (28–30).

Input Data

The input images to the network included precontrast T1-weighted, T2-weighted, and FLAIR images. The input images were masked by the T2-hyperintense lesion mask obtained by using the MRI automatic processing pipeline, dilated by 3 voxels in each direction. The data were augmented by including 90° rotations and reflections about XY and YZ planes to avoid overfitting during network training. Sampling was done such that the MRI data sets from the two sessions (baseline and 6-month follow-up) in the same participant were assigned to the same group (either training or testing).

Areas of enhancement on postcontrast T1-weighted images were identified by two experts (P.A.N., with >30 years of experience in MRI of MS and other neurologic disorders, and J.S.W., with >30 years of experience in MRI of MS), whose findings served as the ground truth. Any disagreement between the two experts was resolved by consensus. For

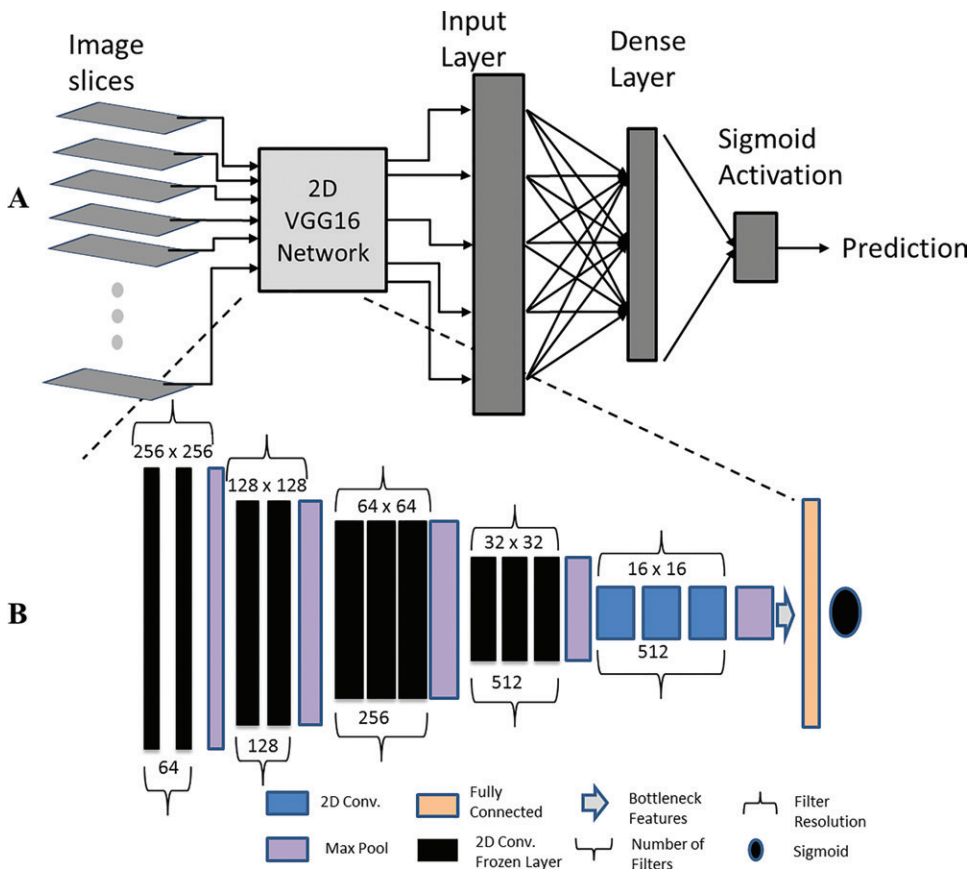


Figure 1: A, Illustration shows architecture of cascaded network for predicting enhancement. B, Illustration shows VGG16 network prediction of enhancement in each section. Scores from slices were combined by using a fully connected network to produce volume (participant-level) prediction. Numbers above VGG16 layers indicate image resolution; numbers underneath layers indicate number of filters (kernel size, 3 × 3). Conv. = convolution, 2D = two-dimensional.

training the DL classification model, image labels were generated for each slice and for the participant based on the presence or absence of contrast-enhancing lesions in each slice or volume, respectively.

Network Description

The DL model consisted of a cascade of two networks. The first network was a convolutional neural network that evaluated all two-dimensional slices from the MRI data for possible contrast enhancement in each slice (slice-wise enhancement). Another important evaluation from a clinical perspective was whether there were any enhancing lesions (ie, zero vs at least one) in a particular participant (participant-wise enhancement). Thus, a second fully connected network was used to combine the slice scores and produce a participant-wise prediction of enhancement.

The convolutional neural network model we used was based on transfer learning, a technique that uses knowledge from another model that has been trained on a different task (31). Transfer learning is shown to be effective in image analysis (32). It reduces computational time and memory requirements. VGG16, a two-dimensional network, was fine-tuned for identifying enhancing lesions. VGG16 achieved state-of-the-art performance on the ImageNet Large Scale Visual Recognition Challenge (<http://image-net.org/challenges/LSVRC/>).

The DL model used in this study consisted of a combination of two networks (Fig 1, A): VGG16 for predicting slice-wise enhancements and a fully connected network for participant-wise prediction. The VGG16 network is shown in Figure 1, B. It consists of five convolutional blocks and one block of fully connected layers followed by a softmax activation function at the output. Each convolutional block is composed of multiple convolutional layers, each with rectified linear unit activation, followed by a max pooling layer. Convolutional layers on each block are assigned a specific number of filters, beginning with 64 for convolutional layers in the first convolutional block. Subsequent blocks have double the number of filters from the previous block. The number of filters of the last convolutional block is not doubled but kept at 512. In our implementation, the receptive field was set at 3 × 3 with a stride size of 1 for all convolutional layers. The second fully connected network,

which combines the slice scores for participant-wise prediction, consists of an input layer (corresponding to the slice scores), a hidden layer of four nodes, and an output layer with a sigmoid activation function. The scripts and other information about this model can be found at <https://github.com/uthmri>.

Network Training

The training procedure consisted of two stages. First, the dense layers of the VGG16 network were removed; only the convolutional and max pooling layers were kept. The training set was then fed through the resulting architecture to generate bottleneck features (Fig 1, B). These features were used to train a new dense layer for identifying possible enhancing lesions. In this training stage, Adam, an algorithm for first-order gradient-based optimization of stochastic objective functions, was used as an optimizer (33) owing to its fast convergence and weight-dependent learning rate. Binary cross-entropy was used as the model loss function along with rectified linear unit activation for all but the last layer, which had sigmoid activation. In the second stage, the last convolutional block and the dense layers were trained on the image features. During this training stage, all other convolutional blocks in the network were frozen so that the weights of layers in those blocks were not updated. In this training stage, the stochastic gradient descent (33) optimizer was used, along with a small learning rate and high

momentum. This configuration restricted major changes to weights to retain previously learned features. Loss and activation functions remained the same as in the first stage. In our implementation, the network was trained for approximately 200 epochs (one epoch = one full iteration of training set); at each epoch, network weights were updated by means of back propagating the error between network output and ground truth. An initial learning rate of 10^{-4} was used. The fully connected network was trained for 100 epochs by using Adam as the optimizer, with an initial learning rate of 10^{-3} . Dropout was used to avoid overfitting. The data splitting for the fully connected network was the same as that used for the VGG16 network.

For equal representation of both classes (enhancing and non-enhancing lesions) during training, minibatches were sampled with a 1:1 ratio, drawn randomly from participants with enhancing lesions ($n = 519$) and without enhancing lesions ($n = 1451$). This choice, effectively oversampling the minority class, minimizes problems related to class imbalance, which can be detrimental for classifier performance (34).

Statistical Analysis

The area under the receiver operating characteristic curve (AUC), sensitivity, specificity, accuracy, positive predictive value, and negative predictive value were calculated to evaluate the network performance. This analysis was performed for both slice-wise and participant-wise predictions. A fivefold cross-validation procedure was implemented to assess the stability of the DL model.

In each iteration, the data were split into two major sets for model development: 80% of the data set was used for training (65%) and validation (15%) and 20% was used for testing. This process was repeated five times by changing the assignments for the partitioned data.

All processing was performed on the Maverick2 cluster at the Texas Advanced Computing Center with graphics processing unit cards (Tesla V100; Nvidia, Santa Clara, Calif). Implementation was carried out by using the Keras Python library (version 2.2.4) (35) and TensorFlow (version 1.12.0) (36).

Results

A summary of demographic characteristics and clinical data for the entire CombiRx cohort is given in Table 1. Figure 2 summarizes the total number of scans included and reasons for exclusion. The total number of enhancing lesions in all 519 participants was 1390.

Results of cross-validation for the test data for each run are summarized in Table 2. An average accuracy of 75% was obtained in the slice-wise enhancement prediction, and the average accuracy in predicting enhancements in participants with at least one enhancing lesion (participant-wise enhancement) was 70%. Figure 3 shows examples of selected slices demonstrating true-positive, false-positive, and false-negative classifications by the network. For comparison, the postcontrast T1-weighted images showing the enhancing lesions are also included. Figure 4 shows the receiver operating characteristic curves in which the true-positive rate (or sensitivity) is plotted against the false-positive rate (false positive rate = $1 - \text{specificity}$) at various threshold

Table 1: Summary of Demographic and Clinical Data for the CombiRx Cohort

Parameter	Value
Age (y)*	37.7 ± 9.7
F/M	730/278
Race	
White	883
African American	73
Other	52
Ethnicity	
Hispanic	63
Non-Hispanic	902
Other	43
Symptom duration (y)*	4.8 ± 5.6
Median EDSS at screening [†]	2 (0–6.5)

Note.—Unless otherwise specified, data are numbers of participants. Adapted from reference 22. CombiRx = Combination Therapy in Patients with Relapsing-Remitting Multiple Sclerosis, EDSS = Expanded Disability Status Scale.

* Means ± standard deviations.

[†] Numbers in parentheses are the range.

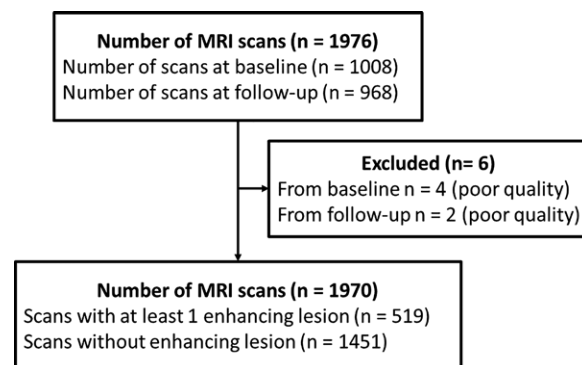


Figure 2: Study flowchart.

settings. The sensitivity and specificity in Table 2 were computed at the optimal thresholds that maximize the Youden index (sensitivity + specificity - 1). The average AUCs (\pm standard deviations) were 0.82 ± 0.02 and 0.75 ± 0.03 for predicting slice-wise and participant-wise enhancements, respectively.

Discussion

Administration of gadolinium-based contrast agents (GBCAs) is crucial for the management of multiple sclerosis (MS). Repeated administration of GBCAs raised concerns about patient safety. In this study, we investigated the feasibility of applying deep learning (DL) to identify enhancing lesions on unenhanced images. The ability to identify enhancing lesions without GBCA administration by using DL can potentially minimize the need for GBCA administration, improve patient safety, and reduce costs associated with clinical care. The annotated Combination Therapy in Patients with Relapsing-Remitting Multiple Sclerosis (CombiRx) MRI data that we used provided an excellent platform to investigate the potential of DL in the prediction of enhancing lesions on unenhanced images. Expert-identified

Table 2: Results for Each Iteration (Run) of Fivefold Cross-Validation of the Test Data for the Prediction of Enhancing MS Lesions

Prediction Type	Sensitivity (%)	Specificity (%)
Slice-wise prediction		
Run 1	76 (299/393)	74 (291/393)
Run 2	71 (278/391)	75 (293/391)
Run 3	77 (279/362)	76 (275/362)
Run 4	83 (350/422)	71 (300/422)
Run 5	83 (334/407)	69 (281/407)
Average*	78 ± 4.3 (74, 82)	73 ± 2.7 (71, 75)
Participant-wise prediction		
Run 1	81 (88/109)	62 (177/286)
Run 2	67 (68/102)	69 (206/298)
Run 3	71 (71/100)	72 (211/293)
Run 4	83 (86/104)	66 (189/287)
Run 5	58 (60/103)	80 (230/287)
Average*	72 ± 9.0 (64, 80)	70 ± 6.3 (64, 75)

Note.—Unless otherwise specified, numbers in parentheses are raw data. Sensitivity is reported as the number of enhancing lesions predicted/total number of cases with enhancing lesion. Specificity is reported as the number of cases predicted as non-enhancing/number of cases without enhancing lesions. MS = multiple sclerosis.

* Numbers in parentheses are the 95% confidence interval.

enhancement on postcontrast T1-weighted images was considered as the ground truth. The underlying hypothesis was that multiple MRI sequences (T1-weighted, T2-weighted, and fluid-attenuated inversion recovery) provide complementary texture information that could be used to identify enhancement without the administration of contrast material. Our results indicate that enhancement can be predicted with an accuracy of 75% and 70% for slice-wise and participant-wise evaluation, respectively. These results are promising for a protocol that avoids contrast material administration. Reasonable consistency of the results was noted among the five cross-validation runs, suggesting that the network performance was stable.

The MRI data from the CombiRx cohort are quite heterogeneous in the sense that imaging was performed at different centers by using different MRI systems and field strengths. The results obtained with such heterogeneous data are perhaps more robust and generalizable than those from a single-center study. Owing to the high computational demands, the analysis was restricted to the first two time points in the CombiRx data.

The reduced accuracy for participant-wise prediction (70% compared with slice-wise accuracy of 75%) is not surprising; errors in slice-wise predictions are amplified and may completely alter the prediction of patients without enhancing lesions (resulting in a false-positive result) or patients with a single enhancing lesion (resulting in a false-negative result).

The prediction accuracy could be further improved by optimizing the network model and its hyperparameters (eg, learning rate, dropout rate). Another way to improve the performance would be to use an ensemble of networks, with each network trained with a different randomly chosen input. Such studies are in progress. We chose the VGG16 network for our study because

it allows implementation of transfer learning. We found this architecture to be adequate for our task. The VGG16 network with 16 layers may be considered rather shallow when compared with deeper architectures such as Inception (37) and ResNet (38). Whereas deeper architectures have features specialized for large tasks (eg, ImageNet for classification of natural scenes), a shallower architecture has more generalizable features that can be extended more easily to different domains, such as medical imaging. Transfer learning could have been more effective if we had used a network trained on medical images rather than ImageNet; however, we are currently unaware of such a resource.

The CombiRx trial recruited only patients with relapsing-remitting MS, the most common MS phenotype. Lesional enhancement is less common in other MS phenotypes (eg, primary progressive MS). It is also possible that differences in the prevalence and morphologic features exist between different MS phenotypes, which might affect the generalizability of our results.

Patients with MS who are being treated with natalizumab and who are human polyomavirus 2 seropositive, depending on time of exposure to natalizumab, are at risk for developing progressive multifocal leukoencephalopathy (39). However, progressive multifocal leukoencephalopathy lesions are unlikely to show enhancement with GBCA. The exception is with those patients with a high suspicion of progressive multifocal leukoencephalopathy whose disease-modifying therapy has been stopped and who develop immune reconstitution inflammatory syndrome reactions that often include extensive GBCA enhancement as human polyomavirus 2 cytotoxic T cells massively infiltrate the central nervous system. None of the participants in this study showed clinical or radiologic evidence of progressive multifocal leukoencephalopathy.

Our study had some limitations that must be considered when interpreting the data. Although our study was based on a relatively large cohort, the number of patients with enhancing lesions is much smaller than the total number of patients with T2-hyperintense lesions. Our plan to overcome the sample size limitation was to pool data from all centers, regardless of imager type or field strength. In addition, we analyzed two-dimensional images obtained with a slice thickness of 3 mm in the CombiRx trial. This could introduce partial averaging that could compromise the quality of classification. Higher-spatial-resolution images that are typically acquired in the three-dimensional mode would have been preferable (40). However, we were limited by the CombiRx MRI protocol.

Even though we used data acquired at different centers with different MRI system platforms, the MRI parameters were well controlled in this clinical trial. For generalizability of the model, it is essential that further testing be conducted on a heterogeneous data set. Currently, we are not aware of the availability of data sets that use the same protocol as CombiRx but different sequence parameters. We are currently working on generalizing our model so that it can be applied to more heterogeneous data sets.

Another limitation of this study was that it was based on conventional MRI sequences. Advanced MRI sequences such as diffusion-weighted imaging, magnetization transfer ratio imaging, and myelin fraction mapping were not

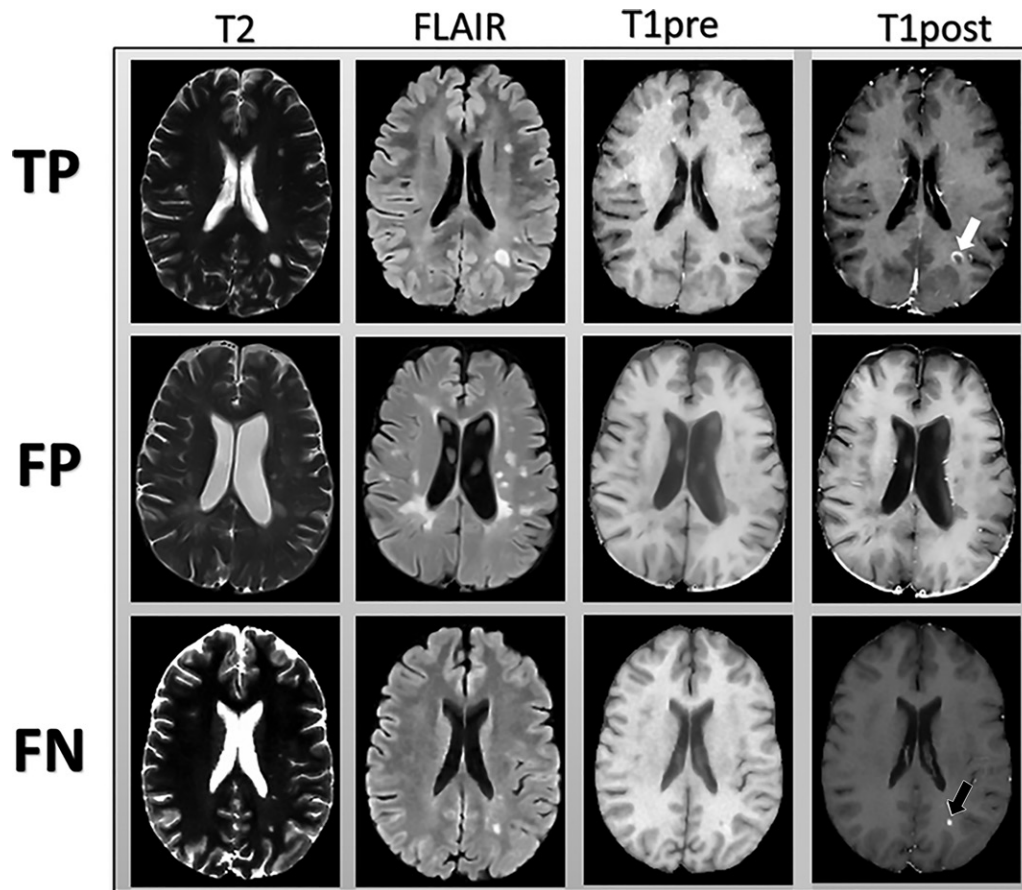


Figure 3: Examples of images input to the network (T2-weighted [T2], fluid-attenuated inversion recovery [FLAIR], and pre-contrast T1-weighted [T1pre] images). Postcontrast T1-weighted images (T1 post) demonstrating areas of true-positive (white arrow) and false-negative (black arrow) enhancement are shown for comparison. FN = false-negative classification of enhancement, FP = false-positive classification of enhancement, TP = true-positive classification of enhancement.

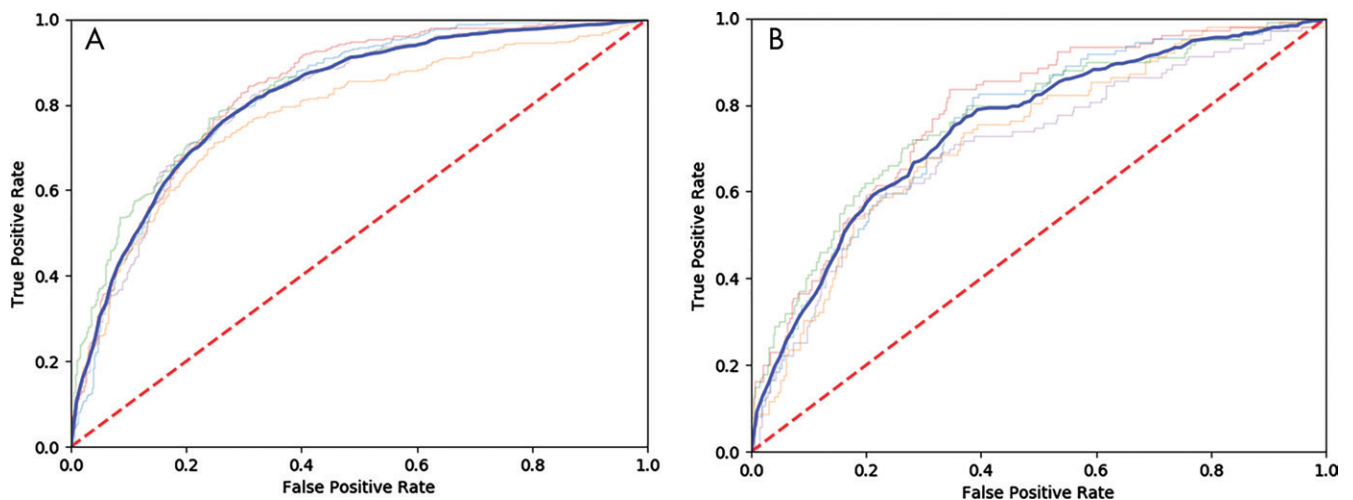


Figure 4: Graphs show receiver operating characteristic curves with true-positive rate plotted against false-positive rate for, A, slice-wise and, B, participant-wise prediction of enhancement. Dashed red line is the chance line. Thin lines show receiver operating characteristic curves for five runs, and solid blue line is average receiver operating characteristic curve. Mean areas under the receiver operating characteristic curve are 0.82 ± 0.02 and 0.75 ± 0.03 for slice-wise and participant-wise prediction, respectively.

included because the CombiRx protocol did not use these. It is possible that inclusion of these additional sequences would increase the accuracy of identification of enhancing lesions.

We used cross-validation to assess the generalization of prediction when applied to new data. Cross-validation showed good generalization, with a mean accuracy of $75\% \pm 1.3$ and $70\% \pm 2.6$ for slice-wise and participant-wise predictions, respectively.

However, further testing is needed to assess generalization to other MRI data sets acquired with a similar protocol.

In conclusion, we investigated the feasibility of using deep learning (DL) for predicting enhancing lesions from MRI scans acquired without administration of gadolinium-based contrast agents (GBCAs). MRI data from a large number of patients with multiple sclerosis (MS) were used to train and test the network performance. Participants with enhancing lesions were classified with 70% accuracy. It is anticipated that inclusion of other MRI sequences could further improve DL performance. This will be necessary before DL is accepted as a viable alternative to GBCAs for identifying enhancing lesions in MS.

Acknowledgment: The Texas Advanced Computing Center, Austin, Tex, provided access to the Maverick2 computer cluster.

Author contributions: Guarantors of integrity of entire study, P.A.N., S.J.S., J.S.W.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, P.A.N., I.C., S.J.S., J.S.W., R.E.G.; clinical studies, S.J.S., J.S.W., F.D.L.; experimental studies, I.C., S.J.S., J.S.W., R.E.G.; statistical analysis, P.A.N., I.C., S.J.S., R.E.G.; and manuscript editing, P.A.N., I.C., S.J.S., J.S.W., R.E.G.

Disclosures of Conflicts of Interest: P.A.N. Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: received funding from SanBio and Sunovion Pharmaceutical. Other relationships: disclosed no relevant relationships. I.C. disclosed no relevant relationships. S.J.S. disclosed no relevant relationships. J.S.W. Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: received personal compensation for consulting, serving on a scientific advisory board, speaking, and other activities from Biogen, Sanofi Genzyme, Actelion; Alkermes, Bionest, Celgene, Clene Nanomedicine, EMD Serono, Forward Pharma A/S; Genzyme, MedDay Pharmaceuticals, Novartis, Otsuka, PTC Therapeutics, Roche/Genentech, and GW Pharma; receives personal compensation for consulting and serving on data safety monitoring board from Brainstorm Cell Therapeutics; receives personal compensation for consulting from AbbVie and Acorda Therapeutics. Other relationships: receives royalties from a patent licensed to UTHealth; has a patent issued to the Board of Regents of the University of Texas System. F.D.L. Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: receives personal fees for being on the advisory board of Bayer HealthCare, Biogen Idec, EMD Serono, Teva Neuroscience, Actelion, Sanofi/Genzyme, Acorda, Roche/Genentech, Celgene/Receptos, Medimmune, TG Therapeutics, Abbvie, MedDay, Atara Biotherapeutics, Innate Immunotherapeutics, Polpharma, Orion Biotech, Brainstorm Cell Therapeutics, Novartis, Jazz Pharmaceuticals, GW Pharma, Biogen, Apitope, and Forward Pharma; received personal fees for patent review from Acorda, Novartis, and Forward Pharma. Other relationships: disclosed no relevant relationships. R.E.G. disclosed no relevant relationships.

References

- Tullman MJ. Overview of the epidemiology, diagnosis, and disease progression associated with multiple sclerosis. *Am J Manag Care* 2013;19(2 Suppl):S15–S20.
- Thompson AJ, Banwell BL, Barkhof F, et al. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *Lancet Neurol* 2018;17(2):162–173.
- Lublin FD, Reingold SC, Cohen JA, et al. Defining the clinical course of multiple sclerosis: the 2013 revisions. *Neurology* 2014;83(3):278–286.
- Grobner T. Gadolinium—a specific trigger for the development of nephrogenic fibrosis and nephrogenic systemic fibrosis? *Nephrol Dial Transplant* 2006;21(4):1104–1108.
- Fraum TJ, Ludwig DR, Bashir MR, Fowler KJ. Gadolinium-based contrast agents: a comprehensive risk assessment. *J Magn Reson Imaging* 2017;46(2):338–353.
- Kanal E. Gadolinium-based contrast agents: the plot thickens. *Radiology* 2017;285(2):340–342.
- Kanda T, Ishii K, Kawaguchi H, Kitajima K, Takenaka D. High signal intensity in the dentate nucleus and globus pallidus on unenhanced T1-weighted MR images: relationship with increasing cumulative dose of a gadolinium-based contrast material. *Radiology* 2014;270(3):834–841.
- McDonald RJ, McDonald JS, Dai D, et al. Comparison of gadolinium concentrations within multiple rat organs after intravenous administration of linear versus macrocyclic gadolinium chelates. *Radiology* 2017;285(2):536–545.
- Trabousee A, Oh J, Barlow L, et al. Consensus statement on the use of gadolinium for magnetic resonance imaging (MRI) used in the diagnosis and follow-up of patients with multiple sclerosis (MS). *J Neurol Sci* 2017;381(Suppl):957.
- Gulani V, Calamante F, Shellock FG, Kanal E, Reeder SB; International Society for Magnetic Resonance in Medicine. Gadolinium deposition in the brain: summary of evidence and recommendations. *Lancet Neurol* 2017;16(7):564–570.
- Michoux N, Guillet A, Rommel D, Mazzamuto G, Sindic C, Duprez T. Texture Analysis of T2-Weighted MR Images to Assess Acute Inflammation in Brain MS Lesions. *PLoS One* 2015;10(12):e0145497.
- Shinohara RT, Goldsmith J, Mateen FJ, Craicaneu C, Reich DS. Predicting breakdown of the blood-brain barrier in multiple sclerosis without contrast agents. *AJNR Am J Neuroradiol* 2012;33(8):1586–1590.
- Lin Z, Li Y, Su P, et al. Non-contrast MR imaging of blood-brain barrier permeability to water. *Magn Reson Med* 2018;80(4):1507–1520.
- Goodfellow I, Bengio Y, Courville A, Bengio Y. *Deep learning*. Cambridge, Mass: MIT Press, 2016.
- LeCun Y, Bengio Y, Hinton G. *Deep learning*. *Nature* 2015;521(7553):436–444.
- Gong E, Pauly JM, Wintermark M, Zaharchuk G. Deep learning enables reduced gadolinium dose for contrast-enhanced brain MRI. *J Magn Reson Imaging* 2018;48(2):330–340.
- Ravi D, Wong C, Deligianni F, et al. *Deep Learning for Health Informatics*. *IEEE J Biomed Health Inform* 2017;21(1):4–21.
- Moeskops P, de Bresser J, Kuijff HJ, et al. Evaluation of a deep learning approach for the segmentation of brain tissues and white matter hyperintensities of presumed vascular origin in MRI. *Neuroimage Clin* 2017;17:251–262.
- Laukamp KR, Thiele F, Shakirin G, et al. Fully automated detection and segmentation of meningiomas using deep learning on routine multiparametric MRI. *Eur Radiol* 2019;29(1):124–132.
- Rachmadi ME, Valdés-Hernández MDC, Agan MLF, Di Perri C, Komura T; Alzheimer's Disease Neuroimaging Initiative. Segmentation of white matter hyperintensities using convolutional neural networks with global spatial information in routine clinical brain MRI with none or mild vascular pathology. *Comput Med Imaging Graph* 2018;66:28–43.
- Lindsey JW, Scott TF, Lynch SG, et al. The CombiRx trial of combined therapy with interferon and glatiramer acetate in relapsing remitting MS: design and baseline characteristics. *Mult Scler Relat Disord* 2012;1(2):81–86.
- Lublin FD, Cofield SS, Cutter GR, et al. Randomized study combining interferon and glatiramer acetate in multiple sclerosis. *Ann Neurol* 2013;73(3):327–340.
- Datta S, Straewen TD, Cofield SS, et al. Regional gray matter atrophy in relapsing remitting multiple sclerosis: baseline analysis of multi-center data. *Mult Scler Relat Disord* 2015;4(2):124–136.
- Govindarajan KA, Datta S, Hasan KM, et al. Effect of in-painting on cortical thickness measurements in multiple sclerosis: a large cohort study. *Hum Brain Mapp* 2015;36(10):3749–3760.
- Govindarajan KA, Freeman L, Cai C, Rahbar MH, Narayana PA. Effect of intrinsic and extrinsic factors on global and regional cortical thickness. *PLoS One* 2014;9(5):e96429.
- Gabr RE, Coronado I, Robinson M, et al. Brain and lesion segmentation in multiple sclerosis using fully convolutional neural networks: a large-scale study. *Mult Scler* doi: 10.1177/1352458519856843. Published online June 13, 2019. <https://doi.org/10.1177/1352458519856843>.
- Narayana PA, Govindarajan KA, Goel P, et al. Regional cortical thickness in relapsing remitting multiple sclerosis: a multi-center study. *Neuroimage Clin* 2012;2(1):120–131.
- Datta S, Sajja BR, He R, Wolinsky JS, Gupta RK, Narayana PA. Segmentation and quantification of black holes in multiple sclerosis. *Neuroimage* 2006;29(2):467–474.
- Datta S, Sajja BR, He R, Gupta RK, Wolinsky JS, Narayana PA. Segmentation of gadolinium-enhanced lesions on MRI in multiple sclerosis. *J Magn Reson Imaging* 2007;25(5):932–937.
- Sajja BR, Datta S, He R, et al. Unified approach for multiple sclerosis lesion segmentation on brain MRI. *Ann Biomed Eng* 2006;34(1):142–151.
- Torrey L, Shavlik J. Transfer learning. In: Soría E, Martín J, Magdalena R, Martínez M, Serrano A, eds. *Handbook of Research on Machine Learning Applications*. Hershey, Pa: IGI Global, 2010. <http://ftp.cs.wisc.edu/machine-learning/shavlik-group/torrey.handbook09.pdf>.
- Yang Y, Yan LF, Zhang X, et al. Glioma Grading on Conventional MR Images: A Deep Learning Study With Transfer Learning. *Front Neurosci* 2018;12:804.
- Kingma D, Ba J. Adam: A Method for Stochastic Optimization. *International Conference for Learning Representations*. <https://arxiv.org/abs/1412.6980>. Published December 22, 2014.
- Buda M, Maki A, Mazurowski MA. A systematic study of the class imbalance problem in convolutional neural networks. *Neural Netw* 2018;106:249–259.
- Chollet F. Keras: Deep learning library for Theano and TensorFlow. <https://keras.io/>. 2015;7(8).
- Abadi M, Barham P, Chen J, et al. Tensorflow: a system for large-scale machine learning. *OSDI*, 2016; 265–283.
- He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016; 770–778.
- Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015; 1–9.
- Clifford DB, De Luca A, Simpson DM, Arendt G, Giovannoni G, Nath A. Natalizumab-associated progressive multifocal leukoencephalopathy in patients with multiple sclerosis: lessons from 28 cases. *Lancet Neurol* 2010;9(4):438–446.
- Crombé A, Saranathan M, Ruet A, et al. MS lesions are better detected with 3D T1 gradient-echo than with 2D T1 spin-echo gadolinium-enhanced imaging at 3T. *AJNR Am J Neuroradiol* 2015;36(3):501–507.