1 **Fully Automated Detection of Paramagnetic Rims in Multiple Sclerosis Lesions on 3T**

2 **Susceptibility-Based MR Imaging**

3

4 **Carolyn Lou MS[1], Pascal Sati PhD[2,3], Martina Absinta MD, PhD[2,4], Kelly Clark MS[1], Jordan D.**

5 **Dworkin PhD[5,6], Alessandra M. Valcarcel PhD[1], Matthew K. Schindler MD, PhD[7], Daniel S.**

6 **Reich MD, PhD[2,4], Elizabeth M. Sweeney PhD[8*†], Russell T. Shinohara PhD[1,9*†]**

7

8 [1]Penn Statistics in Imaging and Visualization Endeavor (PennSIVE) Center, Department of

9 Biostatistics, Epidemiology, and Informatics, University of Pennsylvania, Philadelphia, PA, USA.

10 [2]Translational Neuroradiology Section, National Institute of Neurological Disorders and Stroke

11 (NINDS), National Institutes of Health (NIH), Bethesda, Maryland, USA.

12 [3]Department of Neurology, Cedars-Sinai Medical Center, Los Angeles, California, USA.

13 [4]Department of Neurology, Johns Hopkins School of Medicine, Baltimore, Maryland, USA.

14 [5]Department of Psychiatry, Columbia University Medical Center, New York, New York, USA.

15 [6]New York State Psychiatric Institute, New York, New York, USA.

16 [7]Department of Neurology, University of Pennsylvania, Philadelphia, PA, USA.

17 [8]Department of Population Health Sciences, Weill Cornell Medicine, New York, New York, USA.

18 [9]Center for Biomedical Image Computing and Analytics, Department of Radiology, University of

19 Pennsylvania, Philadelphia, Pennsylvania, USA.

20

21 *Both authors contributed equally to this work.

22

23 [†]Corresponding Authors:

24 Email: ems4003@med.cornell.edu (EMS)

25 Email: rshi@pennmedicine.upenn.edu (RTS)

26

Abstract

Background and Purpose: The presence of a paramagnetic rim around a white matter lesion has recently been shown to be a hallmark of a particular pathological type of multiple sclerosis (MS) lesion. Increased prevalence of these paramagnetic rim lesions (PRLs) is associated with a more severe disease course in MS. The identification of these lesions is time-consuming to perform manually. We present a method to automatically detect PRLs on 3T T2*-phase images.

Methods: T1-weighted, T2-FLAIR, and T2*-phase MRI of the brain were collected at 3T for 19 subjects with MS. The images were then processed with lesion segmentation, lesion center detection, lesion labelling, and lesion-level radiomic feature extraction. A total of 877 lesions were identified, 118 (13%) of which contained a paramagnetic rim. We divided our data into a training set (15 patients, 673 lesions) and a testing set (4 patients, 204 lesions). We fit a random forest classification model on the training set and assessed our ability to classify lesions as PRL on the test set.

Results: The number of PRLs per subject identified via our automated lesion labelling method was highly correlated with the gold standard count of PRLs per subject, r = 0.91 (95% CI [0.79, 0.97]). The classification algorithm using radiomic features can classify a lesion as PRL or not with an area under the curve of 0.80 (95% CI [0.67, 0.86]).

Conclusion: This study develops a fully automated technique for the detection of paramagnetic rim lesions using standard T1 and FLAIR sequences and a T2*phase sequence obtained on 3T MR images.

Keywords: magnetic resonance imaging; multiple sclerosis; chronic active lesions; paramagnetic rim lesions

Highlights:

2

56   • A fully automated method for both the identification and classification of paramagnetic

57   rim lesions is proposed.

58   • Radiomic features in conjunction with machine learning algorithms can accurately

59   classify paramagnetic rim lesions.

60   • Challenges for classification are largely driven by heterogeneity between lesions,

61   including equivocal rim signatures and lesion location.

62 **Introduction**

63

64 Multiple sclerosis (MS) is a demyelinating and inflammatory disease of the central nervous

65 system whose hallmark is lesions in the brain and spinal cord (1). These lesions can be detected

66 *in vivo* with magnetic resonance imaging (MRI) and are often quantified as total lesion volume

67 and lesion count, both of which can be used as measures of disease burden and to track disease

68 progression (2). Imaging biomarkers such as these are commonly used in the clinic and as

69 surrogate endpoints in clinical trials (3,4). However, other known biological processes of MS are

70 left uncaptured.

71

72 Chronic active lesions, which are a subset of MS lesions that are more prevalent in patients with

73 more severe disease (5–7), have imaging and histopathology findings suggestive of ongoing

74 tissue damage (8–10) and have until recently only been detectable by histopathology. These

75 lesions have been variously termed chronic active, slowly expanding, or smoldering lesions. At

76 an estimated prevalence of 10-15% of all MS lesions, this type of lesion is sufficiently common

77 and deleterious to warrant considerable efforts for biomarker development (6,8,9,11). On T2*-

78 phase MRI contrast, they are identifiable by curvilinear hypointensity along the edge of the

79 lesion that corresponds with of iron laden phagocytic cells observed on histopathological

80 specimens (8,9,12). These lesions have been variously termed chronic active, slowly expanding,

81 or smoldering lesions. Here, we refer to these lesions as paramagnetic rim lesions (PRLs).

82

83 When first observed on MRI, the rim of a PRL was only visible on scans from ultra-high-field

84 strength (7T) magnets (13–16). Recently, PRLs have been shown to be identifiable on the more

85 commonly available high-field strength (3T) MRI scans as well, albeit with lower inter- and intra-

86 rater reliability (17). This development strengthens their viability as a target on clinical MRI

87 protocols, particularly because the sequences studied can be acquired with high spatial

88 resolution in less than 4 minutes (18). Previous studies of the PRLs have noted the geometric

89 nature of the rim and worked to identify the rim on the quantitative susceptibility mapping

90 (QSM) contrast as well (19–21).

91

92   Because visually inspecting every MS lesion for the presence of a paramagnetic rim is difficult,

93   time consuming, and prone to inter- and intra-rater variability, an automated method for

94   identifying PRLs would improve efficiency and facilitate translation of this imaging biomarker

95   into larger research studies and clinical practice. One way to identify PRLs is through

96   quantification of visual patterns that objectively characterize these data, which can be

97   accomplished through radiomics. Radiomics is an emerging field of research that encompasses

98   the extraction of quantitative features from biomedical images that may reflect underlying

99   pathophysiology (22). It has been shown to be a useful tool in the analysis of chest CT scans

100 (23,24) and MR images (25,26). Studies have shown that radiomic features are often useful

101 predictors of, or are associated with, known hallmarks of disease, although they have not been

102 used extensively in the MS literature. Here, we use radiomic features along with a random

103 forest classification model, which can flexibly model high dimensional data. Our method is fully

104 automated and uses a T2*-phase volume with isometric voxels and high spatial resolution that

105 is acquired in a clinically feasible acquisition time at 3T (18).

106

107

108 **Materials and Methods**

109

110 Study population:

111 We studied 19 subjects with MS who were scanned under an institutional review board–

112 approved natural history protocol at the National Institutes of Health (NIH). Subjects' age at the

113 time of scanning ranged from 20 to 66 years, with a mean age of 45 years (sd = 12) (Table 1).

114

Table 1: Demographics of Study Sample

Demographics	
N	19
Age (mean (SD))	45 (12)
Male (%)	8 (42)
Phenotype (%)	
Primary Progressive MS	3 (16)
Relapsing-Remitting MS	11 (58)
Secondary Progressive MS	5 (26)
Disease duration in years (mean (SD))	14.6 (9.1)
EDSS (median (range))	2.5 (1.0—7.0)
Treatment	
untreated	5 (26)
glatiramer acetate	1 (5)
interferon beta-1a	4 (21)
dimethyl fumarate	6 (32)
fingolimod	1 (5)
natalizumab	1 (5)
rituximab	1 (5)

115

116   Written informed consent was obtained from all participants. Data from this study can be

117   shared upon reasonable request and completion of a Data Transfer Agreement with the

118   National Institutes of Health.

119

120   MR Imaging acquisition:

121   All subjects were imaged on a Siemens Magnetom Skyra (Siemens, Erlangen, Germany) 3T

122   scanner, using a body transmit coil and a 32-channel receive array coil, at the National

123   Institutes of Health in Bethesda, Maryland. Imaging acquisition included the following

124   sequences:

125 • a whole-brain 3D T2-weighted fluid-attenuated inversion recovery (FLAIR) sequence

126 (repetition time, TR = 4800 ms; echo time, TE = 354 ms; inversion time, TI = 1800 ms; flip

127 angle, FA = 120°; acquisition time, TA = 6 minutes 30 seconds; 256 axial slices; 1mm

128 isometric voxel resolution);

129 • a whole-brain 3D T1-weighted magnetization-prepared rapid gradient echo (T1) sequence

130 (TR = 7.8 ms; TE = 3 ms; FA = 18°; TA = 3 minutes 35 seconds; 256 sagittal slices; 1mm

131 isometric voxel resolution), and

132 • a 3D segmented echo-planar imaging (EPI) sequence with whole-brain coverage providing

133 T2* magnitude and phase contrasts (TR = 64 ms; TE = 35 ms; flip angle, FA = 10°; TA = 5

134 minutes 46 seconds; 251 sagittal slices; 0.65mm isometric voxel resolution).

135 Additional standard MRI sequences, including a postcontrast 3D T1-weighted MPRAGE

136 sequence for the identification of gadolinium-enhancing lesions, were also acquired.

137

138 Manual paramagnetic rim lesion assessment:

139 Supratentorial non-gadolinium enhancing MS lesions were visually inspected for the presence

140 of a paramagnetic rim on T2* magnitude and unwrapped phase images by a neurologist with 14

141 years of experience in neuroimaging science (5,13,17). As previously described (27), a PRL is

142 identified when a hypointense signal on phase images is observed surrounding the periphery of

143 the lesion, while being either hyper- or isointense in its inner portion.

144

145 Image preprocessing:

146 Phase images were unwrapped and filtered as previously described (13). T1, FLAIR, and phase

147 images were then preprocessed using the *fslr* R package (28), an R wrapper for the FSL software

148 (29,30). Images were visualized with ITK-SNAP (31). The T2* magnitude contrast was not used

149 in this method.

150

151 We first applied the N4 inhomogeneity correction algorithm (32). We then rigidly registered

152 both the T1 and the FLAIR images to the T2*-phase image space, resampling to 0.65 mm

153 isometric resolution and using a mutual information cost function and nearest neighbor

7

154   interpolation. We used multi-atlas skull stripping (MASS) to identify cerebral tissue in the

155   images in T1 space (33). In two cases, MASS yielded poorly skull-stripped images based on

156   visual inspection. For those two cases, we instead used the FSL brain extraction tool for skull-

157   stripping (29). As a final step, we performed WhiteStripe intensity normalization on the

158   otherwise preprocessed T1, FLAIR, and phase images (34).

159

160   Lesion labelling:

161   Our lesion labelling method relies on access to maps that represent voxel-wise probabilities of

162   being a lesion, so we chose the automatic lesion segmentation method MIMoSA both for its

163   ability to integrate multimodal information and for its ability to provide voxel-level probability

164   maps (35). Manual lesion segmentation was conducted by a research assistant with 1 year of

165   experience, who was trained by a board-certified neurologist with extensive expertise in

166   neuroimmunology and MRI.

167

168   We trained the MIMoSA algorithm (36) with the manual segmentations as a gold standard and

169   the T1 and FLAIR images as input. We implemented a leave-one-out cross-validation approach,

170   where data from all but one subject was used to train a MIMoSA model, and that model was

171   subsequently applied to the remaining subject. This was done for every subject in our cohort.

172   When parallelized across 8 cores of a CPU of an Intel(R) Xeon(R) E5-2699 v4 @ 2.20GHz

173   processor in a high-performance computing environment, training a single model on 19

174   subjects took approximately 6 hours.

175

176   From each k-fold model, we extracted probability maps that contain voxel-wise probabilities of

177   being a white matter lesion. We then binarized these probability maps into lesion segmentation

178   maps via a subject-specific estimated optimal threshold that was identified out of a user-

179   provided range of possible thresholds and then chosen based on amount of overlap with a gold-

180   standard lesion segmentation as measured by a Sørensen-Dice coefficient (37). Because our

181   lesion segmentation masks did not always cover the entire area of a lesion, we then dilated the

8

182   masks by one voxel in each direction to increase the likelihood of detecting the paramagnetic

183   rim signal, which occurs on the boundary of lesions.

184

185   After lesion segmentation masks were obtained, we used the lesion probability maps as input

186   to a center detection method (38) to identify distinct lesions based on the texture of the lesion

187   tissue. We then used a nearest-neighbor approach to classify the remainder of the lesion

188   segmentation map into those identified lesions (Figure 1). At this point, we assigned PRL status

189   to the identified lesions based on the presence of any overlap with the manual PRL labels

190   described previously.

191

192   **Figure 1**
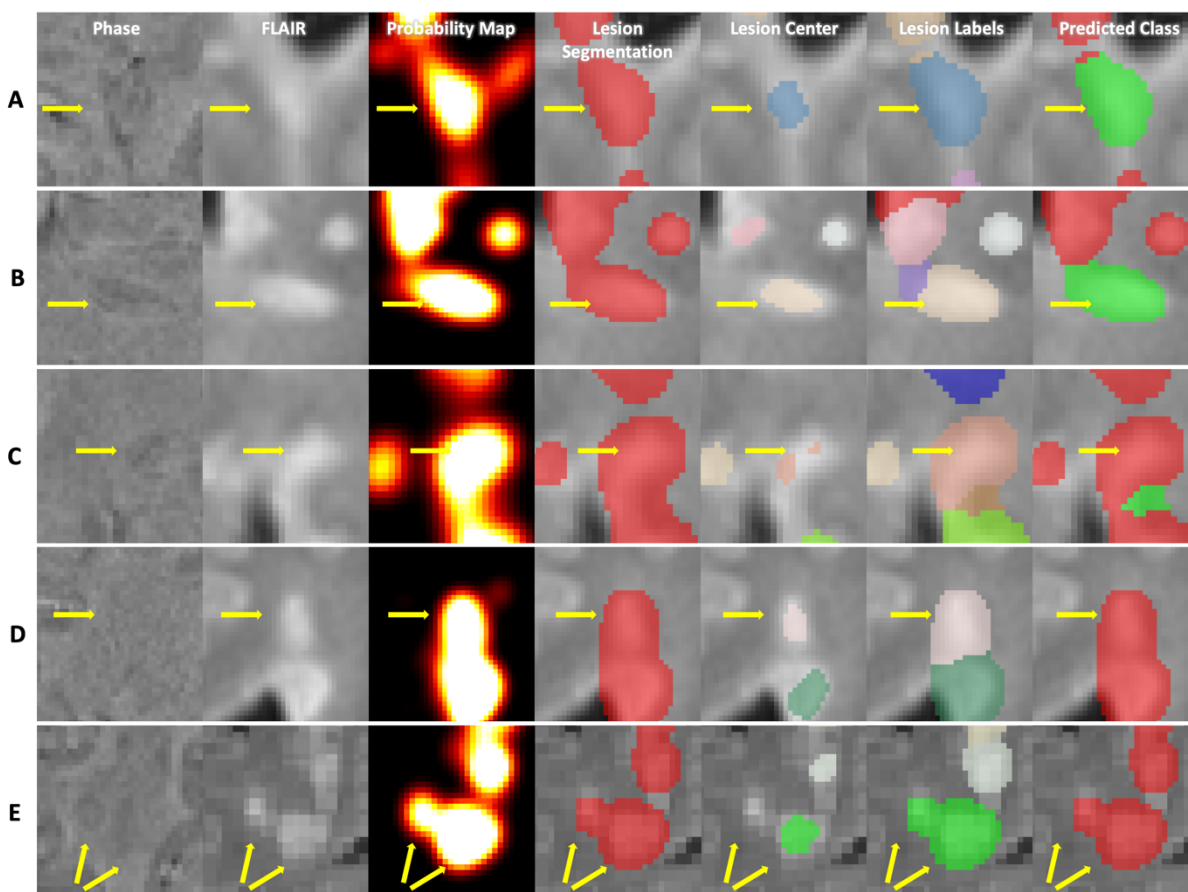
193

194   A visualization of the steps of the method for five different lesions. Each column corresponds to

195   a different part of the method, and each row corresponds to a different lesion of interest. In the

9

196   last column, lesions classified as PRLs are visualized as green, and lesions classified as not PRLs

197   are visualized as red. Subfigure A shows a lesion that was both identified as a PRL and classified

198   as a PRL, i.e. a true positive. Subfigure B shows a lesion that was identified as not a PRL but

199   classified as a PRL, i.e. a false positive. Correspondingly, subfigure C shows a false negative

200   lesion, and subfigure D shows a true negative lesion. Subfigure E shows a lesion that was

201   automatically labelled as a single lesion but is actually a confluence of lesions.

202

203

204   Due to failures in the lesion labelling process, a subset of abnormalities automatically identified

205   by our method might, to a manual rater, be considered clusters of confluent lesions. Because

206   we did not have access to manual segmentations of distinct lesions, we instead relied on a

207   combination of our lesion labelling method and connected components analysis to label lesions

208   as confluent. Specifically, if connected components only identified one cluster where our lesion

209   labelling method identified more than one lesion, we labelled the constituent lesions as

210   confluent.

211

212   Radiomics image analysis:

213   For lesions that were identified with our automatic pipeline, we conducted a radiomics analysis

214   to characterize each lesion with intensity-based statistics only on the phase contrast (39,40).

215   These include 44 features that summarize the intensities in an individual lesion with measures

216   that can be described in 3 general ways: statistics that describe the average and spread of the

217   intensities, statistics that describe the shape of the distribution of intensities, and statistics that

218   describe the diversity of intensities (40). For example, features like the mean, defined as

219    $\frac{1}{n}\sum_{i=1}^{n}x_i$, and interquartile range, defined as $abs(x_{75\%} - x_{25\%})$, are included in the first group,

220   where x_i represents intensity value at voxel i. Features like variance, defined as

221    $\frac{1}{n}\sum_{i=1}^{n}(x_i - mean(x))^2$, and skew, defined as $\frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - mean(x))^3}{sd(x)^3}$, are included in the second

222   group, and features like energy, defined as $\sum_{i=1}^{n}x_i^2$, uniformity, defined as $\sum_{i=1}^{n}p(x_i)^2$, and

223   entropy, defined as $\sum_{i=1}^{n}-p(x_i)\log_2 p(x_i)$, are included in the third group. A full list and

10

224   detailed equations for each of the first-order radiomic features can be found in the

225   supplemental material of (40).

226

227   Prediction model:

228   The radiomic features were used as candidate predictors in our subsequent prediction

229   modelling for classification of lesions as either being PRL or not. Class labels for each lesion

230   were previously assigned during the lesion labelling step. We split our dataset into a training set

231   and test set by subject, randomly assigning lesions from 15 subjects into the training set and

232   assigning lesions from the remaining 4 subjects into the test set, approximating an 80/20 split.

233   Both sets were examined to ensure that at least 100 lesions were present in each group.

234

235   Because PRLs were of a minority class (approximately 13% of the lesions were classified as

236   being a PRL), we used Synthetic Minority Oversampling TEchnique (SMOTE) to synthetically

237   balance our data (41). With SMOTE, we oversampled the minority class, the PRLs, by the

238   reciprocal of the percentage of PRLs present in the dataset, and we did not undersample the

239   majority class. We then trained a random forest classifier, chosen for its ability to flexibly model

240   a large number of features, with 10-fold cross-validation using the R package *caret* (42,43). We

241   summarized performance results using 0.5 as a threshold where applicable. We also derived

242   empirical confidence intervals for those measurements by randomly reassigning the training

243   and test set and repeating the above process 1000 times. We assessed variable importance in

244   the random forest as the percent increase in mean-squared error for a model with the variable

245   over a model with a permuted version of that variable. We then scaled that measure for

246   comparability across variables.

247

248

249   Post-hoc analysis:

250   An additional board-certified neurologist (MS) with extensive expertise in neuroimmunology

251   and MRI, who was not involved in the generation of the manual PRL labels, examined each

252   misclassified lesion. We rated lesions on a 5-point scale, where 1 indicated definitely not a PRL,

253   2 indicated probably not a PRL, 3 indicated uncertain, 4 indicated probably a PRL, and 5

254   indicated definitely a PRL. Some lesions were automatically labelled as one lesion but were

255   actually a confluence of lesions (Figure 1). We assigned manual ratings to these confluent

256   clusters based on the presence of at least one PRL. We also assessed the method's performance

257   only for lesions that were not part of a confluent cluster.

258

259

260   **Results**

261

262   The final dataset included a total of 877 lesions in 19 subjects identified by our automated

263   lesion labelling method, 118 (13%) of which we found to be PRLs by overlap with the manual

264   annotation. The average number of lesions per subject was 46.2 (sd = 19.8), and the average

265   number of PRLs per subject was 6.2 (sd = 4.0). Table 2 summarizes by subject the total number

266   of lesions identified from our lesion labelling method, the number of PRLs identified from our

267   lesion labelling method, and the number of PRLs identified by a manual rater. The number of

268   identified PRLs by our method was highly correlated with the gold standard count of PRLs, r =

269   0.91 (95% CI [0.79, 0.97]) (Figure 2).

270

271 **Figure 2**
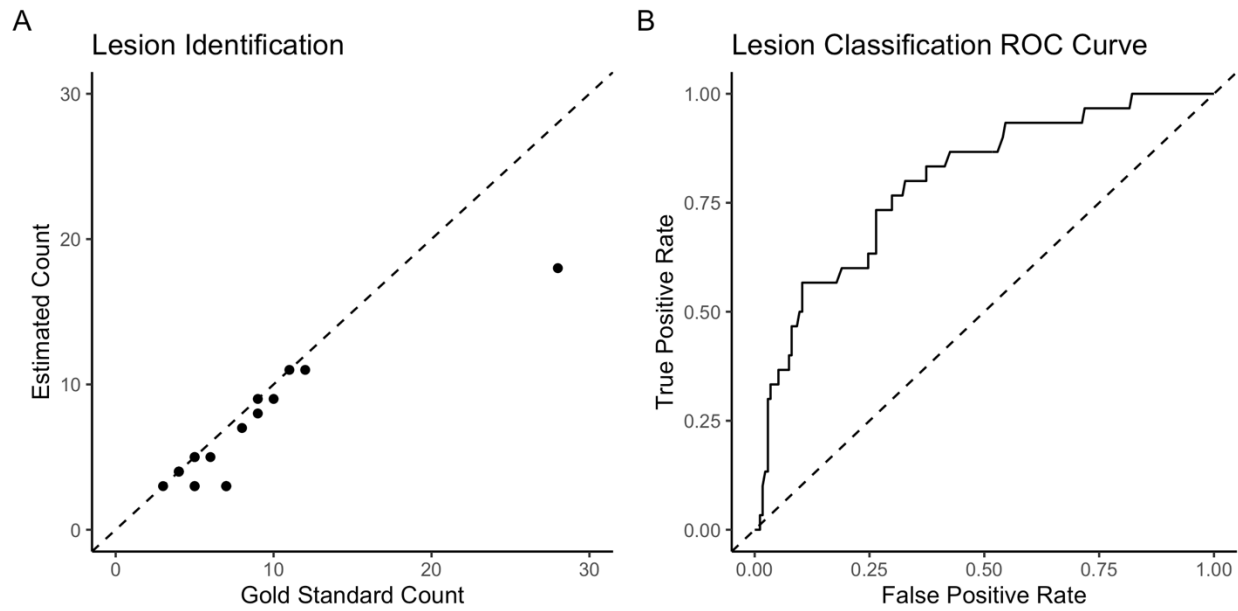
272

273   Subfigure A shows the gold standard count of PRLs against the number of PRLs identified via

274   our lesion identification method, r = 0.91 (0.79, 0.97). Subfigure B shows the ROC curve after

275   classification, AUC = 0.80 (0.67, 0.86).

276

277   **Table 2: Lesion Counts by Subject**

Subject ID	Automated Total Lesion Count	Automated PRL Count	Manual PRL Count
1	83	9	9
2	35	3	5
3	31	5	6
4	22	3	7
5	40	9	10
6	69	18	28
7	14	3	7
8	54	8	9
9	42	3	7
10	52	11	12
11	27	4	4
12	72	4	4
13	40	3	5
14	28	5	5
15	39	11	11
16	78	4	4
17	36	5	5
18	47	7	8
19	68	3	3

278   The table summarizes the total number of lesions identified per subject, the number of

279   identified paramagnetic rim lesions (PRL) by our lesion labelling method, and the number of

280   PRLs identified by a manual rater.

281

282

14

283   We trained a random forest classification model using PRL status from the lesion labelling

284   method as the label. In the iteration that we used to derive performance measures, there were

285   673 lesions in the training set, 88 of which were PRLs, and 204 lesions in the testing set, 30 of

286   which were PRLs. We were able to classify lesions as PRL or not with an AUC of 0.80 (95% CI

287   [0.67, 0.86]). Using 0.5 as a probability threshold, 150 lesions were accurately classified as not

288   PRL, 24 lesions were false positives, 13 were false negatives, and 17 were classified correctly as

289   PRL (Table 3). A breakdown of the classification results for the test set lesions by subject is also

290   provided in Table 3, from which we can see that variability in classification accuracy does not

291   seem to be driven by poor performance in a minority of subjects but rather by heterogeneity in

292   the lesions themselves.

293

15

294   **Table 3: Summary of Classification Performance Measures**

Contingency Table		
	Reference	
Prediction	Rim Negative	Rim Positive
Rim Negative	150	13
Rim Positive	24	17

Performance Measures	
Accuracy	0.82 (0.71, 0.86)
Positive Predictive Value	0.41 (0.16, 0.53)
Negative Predictive Value	0.92 (0.87, 0.97)
False Positive Rate	0.14 (0.08, 0.27)
False Negative Rate	0.43 (0.22, 0.72)
Sensitivity	0.57 (0.29, 0.74)
Specificity	0.86 (0.72, 0.92)

Testing Set Lesion Classification Count by Subject				
Subject	True Negative	False Negative	False Positive	True Positive
1	70	4	4	5
4	18	2	1	1
11	30	5	11	6
20	32	2	8	5

295   The table summarizes the performance measures we observed for the classification of lesions

296   as PRLs or not.

297

298

299   We also examined the results of the method for lesions that were not part of a confluent

300   cluster. A total of 62 lesions in the test set were not confluent, and we were able to classify

301   them with an AUC of 0.91. Using 0.5 as a probability threshold, 50 lesions were accurately

302   classified as not PRL, 4 were false positive, 2 were false negative, and 6 were accurately

16

303 classified as PRL (Table 4). We provide a summary of additional performance measures in Table

304 4.

305

306 **Table 4: Summary of Classification Performance Measures, Excluding Confluent Lesions**

Contingency Table		
	Reference	
Prediction	Rim Negative	Rim Positive
Rim Negative	50	4
Rim Positive	2	6
Performance Measures		
Accuracy	0.90	
Positive Predictive Value	0.75	
Negative Predictive Value	0.92	
False Positive Rate	0.04	
False Negative Rate	0.40	
Sensitivity	0.60	
Specificity	0.96	

307 The table summarizes the performance measures we observed for the classification of lesions

308 after exclusion of lesions in confluent clusters.

309

310

311 A visualization of lesions that were true positive, false positive, false negative, and true negative

312 respectively is provided in Figure 1. From subfigure B, where we see the method illustrated for

313 a lesion that was identified as a not a PRL but classified as a PRL, we can see that

314 hypointensities can manifest around a lesion even when they cannot be rated as a rim.

315 Conversely, from subfigure C, which shows a lesion that was identified as a PRL but classified as

316 not a PRL, we can see that despite the presence of hypointensities that are visible to the eye,

317 certain PRLs may not display a signal strong enough to be captured by radiomic features.

318

319   The random forest identified uniformity, entropy, and energy as the most important radiomic

320   features in classifying lesions, which are all radiomic features that aim to describe the diversity

321   of the data points. (Figure 3). Other radiomic features that were important were mode,

322   kurtosis, skew, geometric mean, and quantile features. Entropy and uniformity were both

323   higher in lesions that were not PRL, and energy was higher in lesions that were PRL.

324

325   **Figure 3**
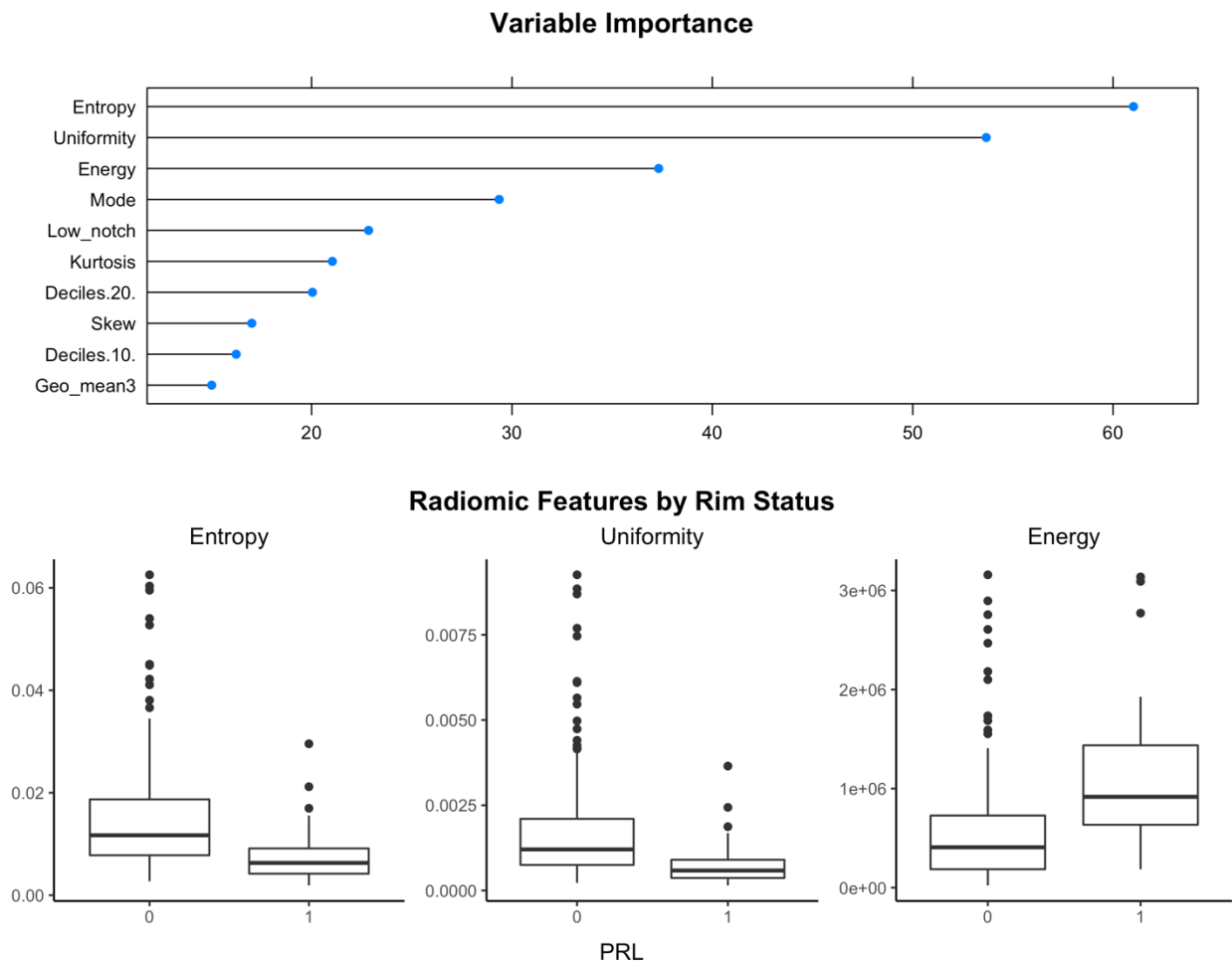
326

327   The variables identified as the most important by our model for determining the presence of

328   PRLs were entropy, uniformity, and energy. Here, we measure variable importance as the

329   percent increase in mean squared error for the model with the variable over the model with a

330   permuted version of that variable, scaled for comparability across variables. Boxplots of

331   entropy, uniformity, and energy on the lesions from the test set show that PRLs and non-PRLs

18

332   seem to differ on those measures, supporting the theory that they are important for

333   distinguishing the two kinds of lesions.

334

335

336   A second expert manually rated the 37 lesions that were misclassified by the model. The rater

337   deemed that 3 lesions included too much artifact to assess PRL status, and 18 lesions were

338   confluent lesions. Of the lesions not part of a confluent cluster, 10 were false positive lesions

339   and 6 were false negative lesions. Of those 10 false positive lesions, 2 of these lesions were

340   rated as definitely a PRL, 2 were rated as probably a PRL, 4 were rated as probably not a PRL,

341   and 2 were rated as definitely not a PRL. For the 6 false negative lesions that were not

342   confluent, 5 were rated as definitely a PRL and 1 was rated as definitely not a PRL.

343

344   As for confluent clusters, 11 were deemed false positives and 7 were false negatives. These

345   were rated according to the presence of at least one PRL in each confluent cluster. Of the 11

346   false positive lesions, 4 were rated as definitely not a PRL, 3 were rated as probably a PRL, and 4

347   were rated as definitely a PRL. Of the 7 false negative lesions, 2 were rated as probably a PRL

348   and 5 were rated as definitely a PRL. Note that the confluence defined here was as judged by

349   the manual rater. This differs from but complements the confluence definition employed for

350   the primary test set analysis, which was the definition based on the automated analysis used to

351   derive the performance measures reported in Table 4.

352

353

354   **Discussion**

355

356   Preliminary studies have shown that the existence of a paramagnetic rim around an MS lesion is

357   an important biomarker with possible clinical implications, shown to be indicative of chronic

358   inflammation, associated with heightened disability, and resistant to current disease-modifying

359   treatments (5). However, paramagnetic rims are time-consuming to identify manually, even by

360   highly trained experts (17). In this paper, we developed a fully automatic method for the

361   detection of the paramagnetic rim on a 3T MRI using a submillimeter isometric, clinically

362   feasible, segmented-EPI sequence (17,18). Automation of PRL identification that relies on

363   objective assessment will aid larger scaled studies assessing this promising imaging biomarker

364   in MS.

365

366   The proposed method relies on radiomics for automated PRL identification and classification.

367   Radiomic features have been used previously in other contexts, but none were used specifically

368   to classify PRLs. The radiomic features that were the most important in this context aimed to

369   measure the variability of intensity within a lesion (entropy and uniformity) or quantify the

370   magnitudes of the intensities themselves (energy).

371

372   Both entropy and uniformity are measures based on the probability of observing a particular

373   intensity within a lesion. Because we did not bin the voxel intensities, this probability of

374   observing a particular intensity is fairly low, which is reflected in the observed range of

375   uniformity in this study. Uniformity is a direct measure of homogeneity of the intensities within

376   a lesion. We expect uniformity to be lower for PRLs due to the presence of both intensities

377   representing normal appearing tissue and hypointensities from the paramagnetic rim. Lesions

378   that are not PRLs do not appear with any distinct signature on the phase image, leading to a

379   higher uniformity.

380

381   Entropy takes the probability of observing a particular intensity within a lesion and transforms it

382   such that the measure reflects the amount of variation observed. Because of the

383   aforementioned lack of binning, entropy here more accurately reflects lesion size in that given

384   our more homogenous set of probabilities, a smaller probability of observing a given intensity

385   results in a smaller measure of entropy, and larger lesions yield a smaller probability of

386   observing a given intensity. In this dataset, PRLs tend to have smaller values of entropy,

387   possibly reflecting a larger size.

388

389 Energy is a measure of the magnitude of intensities within a lesion. Here, PRLs manifest with

390 higher energy because of the way the phase image was created and the subsequent range of

391 the intensities. Hypointensities on the phase image used in this study represent more extreme

392 negative values instead of values closer to 0, with more extreme hypointensities resulting in

393 more extreme energy values.

394

395 Many of the lesions that the model misclassified were confluent lesions that were labelled as a

396 single lesion. While the percentages of confluent lesions among correctly classified lesions was

397 66%, the percentage of confluent lesions among incorrectly classified lesions was 84%,

398 suggesting that confluence negatively influences the model's ability to classify lesions as PRL or

399 not. We provide an example of one of these confluent lesions in Figure 1, Subfigure E. In this

400 lesion, although one of the encompassed lesions contained a clear rim signal, the larger of the

401 two does not. Because the majority of the voxels included in the confluent lesion belong to the

402 encompassed one without a rim signal, the first-order radiomic features extracted from this

403 confluent lesion reflected that signal.

404

405 Artifact made it difficult for a manual rater to rate some of the lesions; our model typically

406 (perhaps incorrectly) rated these as PRL. Of the "false positive" lesions, as determined by the

407 initial PRL manual delineations, while half of those were separately rated as definitely or

408 probably not a PRL, half were rated as definitely or probably a PRL. We also note that for the

409 false positives, around half of the manual ratings were between 2 and 4 on a 5-point scale

410 indicating that even for an expert rater, a large portion of these lesions were difficult to classify.

411 Of the false negative lesions, almost all were rated as definitely a PRL.

412

413 We dilated our lesion segmentation map to increase the likelihood that a rim signal would be

414 included in a lesion label. Because of this artificial augmentation, periventricular lesions and

415 lesions closer to the cortex could be difficult to classify due to inclusion of non-lesional phase-

416 hypointensities in a lesion map, such as ventricles or cortical tissue.

417

418   These issues could be addressed by taking a more nuanced approach to modelling the

419   probability of having a rim. Here, we treated the identification of PRLs as a binary classification

420   problem, invoking a random forest to predict if a given lesion was a PRL or not. However, the

421   identification of PRLs can be difficult because of the myriad of factors that drive the clarity and

422   strength of a rim signature, some of which are technical and some of which reflect biological

423   processes. As noted in Figure 1, while some lesions exhibit a rim unequivocally, other lesions

424   exhibit a more equivocal signature. This renders the task of rating lesions as PRL or not difficult,

425   both for manual raters and automated classifiers. In fact, previous research has shown that

426   intra- and interrater reliability for paramagnetic rim evaluation are substantial but not perfect,

427   with a Cohen κ of 0.77 and 0.71 respectively (17). A future, more nuanced approach could treat

428   the presence of a rim as a continuous measure instead of a binary classification. This would

429   likely more accurately reflect underlying biological processes as well, as the amount of iron-

430   containing phagocytes at the edge of a lesion can vary across lesions (8).

431

432   Limitations:

433   A major limitation to current assessments of paramagnetic rims is that no international

434   consensus exists on criteria for determining this imaging signature. This limitation may hinder

435   the application of the proposed methodology to new studies in which differing definitions of

436   paramagnetic rims may be desired based on local practices. While signal-to-noise ratio is higher

437   on a 7T MR image, allowing for higher inter- and intra-rater reliability, they remain low across

438   contrast types on 3T (17). However, our study relies on techniques that perform well on 3T

439   images, so extensions to 7T would require additional validation.

440

441   In addition, PRLs are a less common type of lesion. In the current study, 13% of the lesions that

442   were identified had rims. Because they are a rare event, classical machine learning models may

443   need to be adjusted in order to classify them with appropriate consideration. In the current

444   study, we employed SMOTE to artificially balance our training data. Other machine learning

445   methods may benefit more from other solutions.

446

447 **Conclusion**

448

449 This study introduces a fully automated method for the identification and classification of

450 paramagnetic rim lesions relying solely on 3T MR images, which are commonly available in a

451 clinical setting. Automation of this process is important for the continued development of the

452 scientific community's knowledge around these lesions and their implications for disease

453 burden.

454

455

467

468 **Declarations of interest**

469 None

470

473

474 **References**

475 1. Sahraian MA, Radü E-W. MRI Atlas of MS Lesions. Springer Science & Business Media; 2007.
476   184 p.

477 2. Popescu V, Agosta F, Hulst HE, Sluimer IC, Knol DL, Sormani MP, et al. Brain atrophy and
478   lesion load predict long term disability in multiple sclerosis. J Neurol Neurosurg Psychiatry.
479   2013 Oct 1;84(10):1082–91.

480 3. Filippi M, Agosta F. Imaging biomarkers in multiple sclerosis. J Magn Reson Imaging.
481   2010;31(4):770–88.

482 4. Sormani MP, Bruzzi P. MRI lesions as a surrogate for relapses in multiple sclerosis: a meta-
483   analysis of randomised trials. Lancet Neurol. 2013 Jul 1;12(7):669–76.

484 5. Absinta M, Sati P, Masuzzo F, Nair G, Sethi V, Kolb H, et al. Association of Chronic Active
485   Multiple Sclerosis Lesions With Disability In Vivo. JAMA Neurol. 2019 Dec 1;76(12):1474–83.

486 6. Frischer JM, Weigand SD, Guo Y, Kale N, Parisi JE, Pirko I, et al. Clinical and pathological
487   insights into the dynamic nature of the white matter multiple sclerosis plaque. Ann Neurol.
488   2015;78(5):710–21.

489 7. Luchetti S, Fransen NL, van Eden CG, Ramaglia V, Mason M, Huitinga I. Progressive multiple
490   sclerosis patients show substantial lesion activity that correlates with clinical disease
491   severity and sex: a retrospective autopsy cohort analysis. Acta Neuropathol (Berl).
492   2018;135(4):511–28.

493 8. Dal-Bianco A, Grabner G, Kronnerwetter C, Weber M, Höftberger R, Berger T, et al. Slow
494   expansion of multiple sclerosis iron rim lesions: pathology and 7 T magnetic resonance
495   imaging. Acta Neuropathol (Berl). 2017;133(1):25–42.

496 9. Absinta M, Sati P, Schindler M, Leibovitch EC, Ohayon J, Wu T, et al. Persistent 7-tesla phase
497   rim predicts poor outcome in new multiple sclerosis patient lesions. J Clin Invest. 2016 Jul
498   1;126(7):2597–609.

499   10. Kaunzner UW, Kang Y, Zhang S, Morris E, Yao Y, Pandya S, et al. Quantitative susceptibility
500   mapping identifies inflammation in a subset of chronic multiple sclerosis lesions. Brain J
501   Neurol. 2019 01;142(1):133–45.

502   11. Chawla S, Kister I, Sinnecker T, Wuerfel J, Brisset J-C, Paul F, et al. Longitudinal study of
503   multiple sclerosis lesions using ultra-high field (7T) multiparametric MR imaging. PLOS ONE.
504   2018 Sep 13;13(9):e0202918.

505   12. Bagnato F, Hametner S, Yao B, van Gelderen P, Merkle H, Cantor FK, et al. Tracking iron in
506   multiple sclerosis: a combined imaging and histopathological study at 7 Tesla. Brain J
507   Neurol. 2011 Dec;134(Pt 12):3602–15.

508   13. Absinta M, Sati P, Gaitán MI, Maggi P, Cortese ICM, Filippi M, et al. Seven-tesla phase
509   imaging of acute multiple sclerosis lesions: A new window into the inflammatory process.
510   Ann Neurol. 2013;74(5):669–78.

511   14. Bian W, Harter K, Hammond-Rosenbluth KE, Lupo JM, Xu D, Kelley DA, et al. A serial in vivo
512   7T magnetic resonance phase imaging study of white matter lesions in multiple sclerosis.
513   Mult Scler Houndmills Basingstoke Engl. 2013 Jan;19(1):69–75.

514   15. Mehta V, Pei W, Yang G, Li S, Swamy E, Boster A, et al. Iron Is a Sensitive Biomarker for
515   Inflammation in Multiple Sclerosis Lesions. PLOS ONE. 2013 Mar 14;8(3):e57573.

516   16. Hammond KE, Metcalf M, Carvajal L, Okuda DT, Srinivasan R, Vigneron D, et al. Quantitative
517   in vivo magnetic resonance imaging of multiple sclerosis at 7 Tesla with sensitivity to iron.
518   Ann Neurol. 2008;64(6):707–13.

519   17. Absinta M, Sati P, Fechner A, Schindler MK, Nair G, Reich DS. Identification of Chronic Active
520   Multiple Sclerosis Lesions on 3T MRI. Am J Neuroradiol. 2018 Jul;39(7):1233–8.

521   18. Sati P, Thomasson DM, Li N, Pham DL, Biassou NM, Reich DS, et al. Rapid, high-resolution,
522   whole-brain, susceptibility-based MRI of multiple sclerosis. Mult Scler Houndmills
523   Basingstoke Engl. 2014 Oct;20(11):1464–70.

524   19. Eskreis-Winkler S, Deh K, Gupta A, Liu T, Wisnieff C, Jin M, et al. Multiple sclerosis lesion
525   geometry in quantitative susceptibility mapping (QSM) and phase imaging. J Magn Reson
526   Imaging. 2015;42(1):224–9.

527   20. Stüber C, Pitt D, Wang Y. Iron in Multiple Sclerosis and Its Noninvasive Imaging with
528   Quantitative Susceptibility Mapping. Int J Mol Sci. 2016 Jan 14;17(1).

529   21. Wisnieff C, Ramanan S, Olesik J, Gauthier S, Wang Y, Pitt D. Quantitative susceptibility
530   mapping (QSM) of white matter multiple sclerosis lesions: Interpreting positive
531   susceptibility and the presence of iron. Magn Reson Med. 2015 Aug;74(2):564–70.

532   22. Rizzo S, Botta F, Raimondi S, Origgi D, Fanciullo C, Morganti AG, et al. Radiomics: the facts
533   and the challenges of image analysis. Eur Radiol Exp [Internet]. 2018 Nov 14 [cited 2020 Jun
534   30];2. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6234198/

535   23. Coroller TP, Agrawal V, Narayan V, Hou Y, Grossmann P, Lee SW, et al. Radiomic phenotype
536   features predict pathological response in non-small cell lung cancer. Radiother Oncol. 2016
537   Jun 1;119(3):480–6.

538   24. Liu Y, Kim J, Balagurunathan Y, Li Q, Garcia AL, Stringfield O, et al. Radiomic Features Are
539   Associated With EGFR Mutation Status in Lung Adenocarcinomas. Clin Lung Cancer. 2016
540   Sep 1;17(5):441-448.e6.

541   25. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, et al. Advancing The Cancer
542   Genome Atlas glioma MRI collections with expert segmentation labels and radiomic
543   features. Sci Data. 2017 Sep 5;4(1):170117.

544   26. Li Y, Qian Z, Xu K, Wang K, Fan X, Li S, et al. MRI features predict p53 status in lower-grade
545   gliomas via a machine-learning approach. NeuroImage Clin. 2018 Jan 1;17:306–11.

546   27. Yao B, Bagnato F, Matsuura E, Merkle H, van Gelderen P, Cantor FK, et al. Chronic multiple
547   sclerosis lesions: characterization with high-field-strength MR imaging. Radiology. 2012
548   Jan;262(1):206–15.

549   28. Muschelli J, Sweeney E, Lindquist M, Crainiceanu C. fslr: Connecting the FSL Software with
550   R. R J. 2015 Jun;7(1):163–75.

551   29. Jenkinson M, Beckmann CF, Behrens TEJ, Woolrich MW, Smith SM. FSL. NeuroImage. 2012
552   Aug 15;62(2):782–90.

553   30. Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TEJ, Johansen-Berg H, et al.
554   Advances in functional and structural MR image analysis and implementation as FSL.
555   NeuroImage. 2004 Jan 1;23:S208–19.

556   31. Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, et al. User-guided 3D active
557   contour segmentation of anatomical structures: significantly improved efficiency and
558   reliability. NeuroImage. 2006 Jul 1;31(3):1116–28.

559   32. Tustison NJ, Avants BB, Cook PA, Zheng Y, Egan A, Yushkevich PA, et al. N4ITK: Improved N3
560   Bias Correction. IEEE Trans Med Imaging. 2010 Jun;29(6):1310–20.

561   33. Doshi J, Erus G, Ou Y, Gaonkar B, Davatzikos C. Multi-Atlas Skull-Stripping. Acad Radiol. 2013
562   Dec 1;20(12):1566–76.

563   34. Shinohara RT, Sweeney EM, Goldsmith J, Shiee N, Mateen FJ, Calabresi PA, et al. Statistical
564   normalization techniques for magnetic resonance imaging. NeuroImage Clin. 2014 Aug
565   15;6:9–19.

566   35. Valcarcel AM, Linn KA, Vandekar SN, Satterthwaite TD, Muschelli J, Calabresi PA, et al.
567   MIMoSA: An Automated Method for Inter-Modal Segmentation Analysis of Multiple
568   Sclerosis Brain Lesions. J Neuroimaging Off J Am Soc Neuroimaging. 2018 Jul;28(4):389–98.

569   36. Valcarcel AM. mimosa: "MIMoSA": A Method for Inter-Modal Segmentation Analysis.
570   [Internet]. 2017. Available from: https://github.com/avalcarcel9/mimosa

571   37. Valcarcel AM, Muschelli J, Pham DL, Martin ML, Yushkevich P, Brandstadter R, et al. TAPAS:

572   A Thresholding Approach for Probability Map Automatic Segmentation in Multiple Sclerosis.

573   NeuroImage Clin. 2020 Jan 1;27:102256.

574   38. Dworkin JD, Linn KA, Oguz I, Fleishman GM, Bakshi R, Nair G, et al. An Automated Statistical

575   Technique for Counting Distinct Multiple Sclerosis Lesions. Am J Neuroradiol. 2018

576   Apr;39(4):626–33.

577   39. Kolossváry M, Kellermayer M, Merkely B, Maurovich-Horvat P. Cardiac Computed

578   Tomography Radiomics: A Comprehensive Review on Radiomic Techniques. J Thorac

579   Imaging. 2017 Mar 1;33:1.

580   40. Kolossváry M, Karády J, Szilveszter B, Kitslaar P, Hoffmann U, Merkely B, et al. Radiomic

581   Features Are Superior to Conventional Quantitative Computed Tomographic Metrics to

582   Identify Coronary Plaques With Napkin-Ring Sign. Circ Cardiovasc Imaging. 2017 Dec;10(12).

583   41. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic Minority Over-

584   sampling Technique. J Artif Intell Res. 2002 Jun 1;16:321–57.

585   42. Kuhn M. Building Predictive Models in R Using the caret Package. J Stat Softw. 2008 Nov

586   10;28(1):1–26.

587   43. Wright MN, Ziegler A. ranger: A Fast Implementation of Random Forests for High

588   Dimensional Data in C++ and R. J Stat Softw. 2017 Mar 31;77(1):1–17.

589