Journal of Medical Imaging

MedicalImaging.SPIEDigitalLibrary.org

# Etiology-based classification of brain white matter hyperintensity on magnetic resonance imaging

Mariana Leite
Letícia Rittner
Simone Appenzeller
Heloísa Helena Ruocco
Roberto Lotufo

SPIE.

# Etiology-based classification of brain white matter hyperintensity on magnetic resonance imaging

**Mariana Leite,[a,*] Letícia Rittner,[a] Simone Appenzeller,[b] Heloísa Helena Ruocco,[c] and Roberto Lotufo[a]**
[a]University of Campinas, Faculty of Electrical and Computer Engineering, Department of Computer Engineering and Industrial Automation, Albert Einstein Avenue, 13083-852 Campinas, Brazil
[b]University of Campinas, Faculty of Medical Science, Rheumatology Division, Zeferino Vaz Avenue, 13083-970 Campinas, Brazil
[c]Hospital of Faculty of Medicine of Jundiaí, Multiple Sclerosis Center, Francisco Telles Street, 13202-550 Jundiaí, Brazil

**Abstract.** Brain white matter lesions found upon magnetic resonance imaging are often observed in psychiatric or neurological patients. Individuals with these lesions present a more significant cognitive impairment when compared with individuals without them. We propose a computerized method to distinguish tissue containing white matter lesions of different etiologies (e.g., demyelinating or ischemic) using texture-based classifiers. Texture attributes were extracted from manually selected regions of interest and used to train and test supervised classifiers. Experiments were conducted to evaluate texture attribute discrimination and classifiers' performances. The most discriminating texture attributes were obtained from the gray-level histogram and from the co-occurrence matrix. The best classifier was the support vector machine, which achieved an accuracy of 87.9% in distinguishing lesions with different etiologies and an accuracy of 99.29% in distinguishing normal white matter from white matter lesions.  [DOI: 10.1117/1.JMI.2.1.014002]

Keywords: white matter hyperintensity; brain; magnetic resonance imaging; etiology; texture analysis; classifiers.

Paper 14098PRR received Jul. 22, 2014; accepted for publication Jan. 20, 2015; published online Feb. 19, 2015.

## 1 Introduction

Magnetic resonance imaging (MRI) quality has improved significantly over the last two decades. Now, MRI allows the observation of considerably more subtle and smaller scale abnormalities.[1] White matter hyperintensities (WMHs) are among the most frequently observed incidental MRI findings, and their prevalence increases with age.[1–3] Other common risk factors associated with increased prevalence of WMH include: female gender, atherosclerosis, and elevated systolic blood pressure.[2]

Although WMH is often considered to have uncertain clinical significance, recent studies indicate that they are associated with cognitive impairment, suggestive of a disruption in brain connectivity.[4] Though the etiology of WMH is difficult to determine because of the lack of pathological studies, it is frequently proposed to be of an ischemic or demyelinating nature.[5]

WMH of an ischemic nature occurs when there is an obstruction of a vessel, usually of small caliber, that irrigates the brain. This obstruction is named microangiopathy and it causes a decrease or cessation of blood circulation and a fast degeneration of brain tissue.[6] In contrast, WMH of a demyelinating nature is associated with inflammation that causes destruction of the myelin sheath and compromises neural transmission.[7] This form of WMH is frequently observed in a variety of immune-mediated diseases such as multiple sclerosis (MS). The recognition of the lesion etiology is important for adequate treatment. Depending on the evolution of lesions, demyelinating lesions are frequently treated with immunosuppressants or immunomodulation whereas ischemic lesions are treated with anticoagulants.

The most suitable examination to detect WMH is MRI, as it presents an excellent contrast in soft tissues and provides anatomical images with high resolution.[8] However, the analysis of WMH on MRI is a nontrivial task owing to variations in size, shape, and location of the WMH, as well as the complexity of underlying factors such as different imaging devices. Thus, in order to manually characterize WMH in the human brain, specialists usually need to consider additional clinical information from individuals such as age, physical exams, medical history, and differences in images from different modalities. FLAIR, T2-weighted, and contrast-enhanced images can help to suggest a demyelinating lesion, whereas diffusion images can support the ischemic or inflammatory nature of the lesion.

Therefore, the development of computer-assisted diagnostic tools has aided specialists in diagnostic and medical monitoring by reducing the subjectivity of the procedure and by making it more robust and agile. Further, it allows researchers and practitioners to quantify the severity of WMH and to monitor the evolution of individual WMH over time in longitudinal studies.

For example, Loizou et al.[9] quantitatively analyzed WMH in MS subjects based on texture and shape attributes. In their study, MRI from 22 patients was manually segmented by a neurologist and confirmed by a radiologist. Then the shape and texture analyses were performed at two different time points: initial diagnosis and 6 to 12 months after diagnosis. Texture features were used both in differentiating between normal and abnormal tissues and in assessing disease onset, while shape features were used to evaluate the evolution of the lesions (longitudinal study). Nine texture features were computed from the spatial gray-level dependence matrix, also known as a gray-level co-occurrence matrix (GLCM): contrast, sum of squares, inverse difference moment, sum average, sum variance, sum entropy, entropy, difference variance, and difference entropy. Also, 12 shape features

*Address all correspondence to: Mariana Leite, E-mail: maripb@dca.fee.unicamp.br

2329-4302/2015/$25.00 © 2015 SPIE

were computed: $x$-coordinate maximum length (maximum length in the horizontal coordinate), $y$-coordinate maximum length (maximum length in the vertical coordinate), area, perimeter, perimeter/area, eccentricity, equivalence diameter, major axis length, minor axis length, centroid, convex area, and orientation. After feature extraction for each subject, the mean feature value was computed for lesions at 0 months, lesions at 6 to 12 months, and normal white matter. The results showed that there is no significant difference between most of the shape and texture features extracted at diagnosis and the same features determined 6 months after diagnosis. However, they noticed a significant difference between most of the texture features extracted from the normal white matter tissue and from the lesions at 0 and 6 to 12 months.

A computer-assisted method for performing automatic segmentation of WMH in cranial MRI that uses the k-nearest neighbor (kNN) classifier based on voxel gray-level intensities and spatial information was proposed by Anbeek et al.[10] This method generated probability maps representing the probability that each voxel is a part of a WMH infarct, and experimental analysis was performed in a set of 20 patients with stroke.

A method to detect and quantify WMH using supervised segmentation in subjects with the risk of neurological disorders, especially Alzheimers disease, was proposed by Ithapu et al.[11] They applied two learning models [support vector machines (SVMs) and random forests] to accomplish the proposed task and used texture filters, called textons, and intensity variation as features. Their experiments were performed in a dataset containing T1-weighted and T2-weighted MR scans from a total of 251 subjects: 169 healthy controls, 40 mild cognitively impaired, and the remainder were demented. The authors achieved significant improvement over the current state-of-the-art unsupervised model.

Lao et al.[12] presented a computer-assisted method to segment WMH based on local features and SVM classifier. They used multiple MRI sequences of 45 patients with cerebrovascular disease: T1-weighted, T2-weighted, proton density-weighted, and FLAIR MR scans. They reported that the combination of these four different MR acquisition protocols with SVM makes possible the development of a relatively robust and fully automated segmentation method of white matter abnormalities.

Zimring et al.[13] introduced an automatic algorithm for the detecting and contouring of MS lesions in the brain MRI, described in the paper as brighter regions within the image. The method was based on the adaptive threshold algorithm and artificial neural networks' classifier. It was applied in a set of 45 images acquired from 14 patients with MS.

Another related work based on artificial neural networks' classifier aimed to segment WMH in MRI.[14] The multispectral segmentation technique proposed in this study was quantitatively compared with manually delineated regions. A total of 36 images from six brain volumes were analyzed twice, each by two researchers, under the supervision of a neuroradiologist. Results indicated that the segmentation technique proposed by these researchers facilitates the analysis of MRI. In addition, the method produced similar or lower intra- and inter-rate variabilities when compared with a manual segmentation.

Klöppel et al.[15] compared different methods that detect WMH in the brain white matter in MRI using classifiers. For comparison, experiments were conducted where these methods were applied to a set of 20 patients with dementia and cognitive deficits, and the comparison was based on intensity features combined with both supervised and nonsupervised classifiers. The intensity features were the gray-level value of the index voxel and all voxels within a sphere of 8-mm radius. They also applied Gabor filters (GFs) for feature extraction. They computed features from 32 different GFs for each index voxel. The authors emphasize that the classification methods present a set of internal parameters that must be adhered to, limiting their clinical applicability. Their main conclusion is that the SVM classifier provides the best performance; however, the results can be improved.

The aforementioned works aim to locate and/or segment WMH with a specific etiology. However, none of these studies compared the WMH of different etiologies. It is known that the lesion characterization is usually determined by the patient pathology; however, it is possible to have patients with lesions of different etiologies. A MS patient, for example, may have an ischemic lesion. In cases like this, an image-based classifier may be a fundamental tool in order to identify its etiology and, therefore, decide which treatment to use.

In this paper, we present a computer-assisted method based on texture attributes and classifiers to distinguish WMH based on their etiology. Texture attributes' extraction is based on different approaches, such as the co-occurrence matrix, run-length matrix, gradient, and histogram in order to analyze their texture discrimination. We also compare and/or combine multiple classifiers such as SVM, kNN, linear discriminant analysis (LDA), and optimum path forest (OPF).

## 2 Methodology

The proposed method was initiated upon MRI acquisition and finalized with the classification. The method returns the corresponding class of each region of interest (ROI): normal white matter, WMH with ischemic etiology, or WMH with demyelinating etiology, as the output. The method was subdivided into four main steps. First, the image acquisition procedure was performed using the MRI scanner, followed by the manual ROI extraction and annotation performed by a specialist. Second, texture analysis in which the texture attributes were extracted from each ROI was performed and it was followed by the attribute selection procedure (third step). Finally, the classification step was performed (Fig. 1).

The algorithms comprising this methodology were developed on Adessowiki,[16] a web-based collaborative environment for development and documentation of scientific computing algorithms. All codes were written in the Python language and are available on Adessowiki.

### 2.1 *Image Acquisition and Manual Region of Interest Extraction*

Our image database was generated using the T2-weighted MRI and it was obtained in the axial plane (6-mm thick, flip angle 170 deg, repetition time 4800 ms, echo time 128 ms, matrix size $= 256 \times 256$, and FOV $= 21 \times 23$ cm) on a Elscint 1.5T Prestige scanner at the Faculty of Medical Science (FCM) of University of Campinas (UNICAMP). All 54 patients and 19

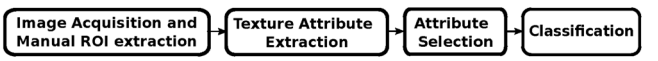


**Fig. 1** Overview of the proposed method: image acquisition and manual region of interest (ROI) extraction, followed by the texture analysis and attributes selection, and finalized by classification.

healthy volunteers were informed in advance and signed a free and informed consent form approved by the Research Ethics Committees of FCM-UNICAMP. The selected group was already diagnosed with MS or stroke, presented a wide range of ages, and both men and women were included. ROIs were manually extracted from two-dimenional slices of the T2-weighted MR images and annotated based on the clinical data of the patients. WMH areas were identified in the manual process as bright areas within the white matter on T2 weighted MR images.

The analysis derived 76 ROIs of normal white matter from 19 healthy volunteers, 64 ROIs of WMH with ischemic etiology from 4 patients with stroke, and 143 ROIs representing WMH with demyelinating etiology from 50 patients with MS. Multiple ROIs were extracted from each patient, usually with irregular sizes and shapes (Fig. 2). There were no visible differences between demyelinating and ischemic WMHs on MR images.

The extracted WMH ROIs did not preserve the lesion boundaries and shapes. This happens because the criterion for bounding these ROIs was to segment only lesion, thus the region localized exactly in the frontier between the lesion and normal tissue (borders) was not used. Normal white matter ROIs were delineated following similar shape, localization, and size of WMH ROIs.

## 2.2 *Attributes Extraction*

Color and shape attributes were not applied, as MRI is gray-scale and our ROIs did not preserve the WMH boundaries and shapes. In addition, since size and shape of the lesions are not regular among classes, it is not possible to use shape attributes to distinguish normal white matter, WMH with demyelinating nature, and WMH with ischemic nature.

Thus, we only used texture attributes. The extracted attributes were determined based on the following texture analysis approaches.

### 2.2.1 *Statistical approach based on the gray-level histogram*

The histogram analyzes the gray-level distribution in an ROI.[17] We extracted nine statistical measures from the histogram: mean, variance, skewness, kurtosis, 1% percentile, 10% percentile, 50% percentile, 90% percentile, and 99% percentile.

### 2.2.2 *Gray-level co-occurrence matrices approach*

The co-occurrence matrix[18] analyzes the occurrence of pairs of pixels with gray levels $i$ and $j$ in an image given a specific offset and orientation between them. We computed the co-occurrence matrix for different distances (1, 2, 3, 4, and 5) and in different orientations (0 deg, 45 deg, 90 deg, and 135 deg). From each co-occurrence matrix, 11 texture attributes were extracted: angular second moment, contrast, correlation, sum of squares, inverse difference moment, sum average, sum variance, sum entropy, entropy, difference variance, and difference entropy. In order to achieve rotation invariant attributes, the mean of attributes extracted in different orientations within a specific distance were computed. Thus, 55 texture attributes were computed from GLCM, since it used five distances.

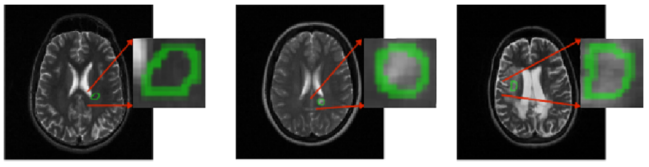


**Fig. 2** Samples of extracted ROIs of different size, shape, and localization: (a) normal white matter ROI, (b) white matter hyperintensity (WMH) with demyelinating etiology ROI, and (c) WMH with ischemic etiology ROI.

### 2.2.3 *Run-length matrix approach*

The run-length matrix[19] analyzes the frequency in which a specific number of pixels with the same gray-level occurs consecutively in a determined direction. We computed this matrix in four different orientations (0 deg, 45 deg, 90 deg, and 135 deg) and extracted five texture attributes from each run-length matrix: run length nonuniformity, grey-level nonuniformity, long run emphasis, short run emphasis, and run percentage. In order to achieve rotation invariant attributes, the means of attributes extracted in different orientations were computed.

### 2.2.4 *Statistical approach based on the gradient*

The gradient presents directional changes on the image gray level. We extracted five statistical measures based on the gray-level distribution of each ROI's gradient:[17] mean, variance, skewness, kurtosis, and percentage of pixels with nonzero gradient.

A total of 74 texture attributes were computed for each ROI through these approaches, and were then normalized between 0 and 1. This feature normalization was performed population wise.

## 2.3 *Attributes Selection*

We applied two different methods of attribute selection: principal component analysis (PCA) and the decision tree algorithm. For both methods, we used a python library called scikits-learn.[20] PCA is a mathematical procedure proposed in 1901 by Karl Pearson to reduce a dataset by discarding redundant information. The result is a lower number of noncorrelated attributes called principal components. This compact representation makes the following classification step faster.[21]

In our experiments, we used PCA as an attribute transformation in which the principal components' vectors were used as new attributes. In order to compute the number of principal components to retain, we used an automatic method to choose the PCAs' dimensionality proposed by Minka.[22] In this method, PCA is interpreted as a density estimation, and a Bayesian model selection is used to determine the dimensionality of the data.

The decision tree algorithm[23] was originally intended for classification, but recently, it has been largely used as an attribute selection method. The main goal of this technique is to generate a tree structure that summarizes the relevant information of the input data, enabling its visualization and interpretation. Each internal node in the tree structure represents an attribute test in which possible results are indicated in the edges. Finally, the external nodes present the final prediction. After decision tree construction, it is possible to choose the most discriminant attributes.

In order to choose the most discriminant attributes, we first compute the discrimination degree of each attribute by using a method called the mean decrease impurity, also known as the Gini importance.[24] For each internal node that splits on a specific feature, the error reduction of that node weighted by the probability of reaching that node is computed. This probability is represented by the number of samples that was routed to the node over all trees of the ensemble. Thereafter, all features whose discrimination degree is lower than the mean discrimination degree are discarded. It is important to notice that these discrimination degree values are relative to a specific dataset. Thus, it is not possible to compare discrimination degrees between different datasets.

### 2.4 *Classification*

The final step of the proposed methodology aimed to distinguish the previous extracted WMH ROIs according to their etiology by using two different approaches: to distinguish normal white matter from WMH as a first step and then to distinguish ischemic WMH from demyelinating WMH; or to distinguish normal white matter, ischemic WMH, and demyelinating WMH at once. The classifiers SVM, OPF, LDA, and kNN were designed based on texture features extracted from ROIs of normal white matter, ischemic WMH, and demyelinating WMH.

SVM is a supervised learning method that can be applied to classification or regression. It performs classification by constructing a set of hyperplanes in a high-dimensional space that optimally separates the data into two categories.[25] The OPF[26] classifier models the data classification task as a partition problem in a graph and can be used as a supervised or unsupervised classification method. LDA is a parametric and statistical method that can be used not only in the attributes' selection step, but also in the classification step. The kNN decision rule assigns the points to an unclassified sample with the most frequent label of its k-nearest previously classified points.[27] The 1NN classifier, kNN classifier using $k = 1$, presents many conceptual similarities to the OPF classifier, and it is possible to obtain the 1NN classifier by considering all training samples of the OPF as prototypes.[28]

We executed the proposed tasks using SVM, OPF, LDA, and kNN classifiers in order to compare their accuracy rates by using the scikits-learn library. The method used to assess the classifier accuracy through randomly sampled partitions of data was the 10-fold cross-validation, in which 10% of the ROIs goes into the test set, while 90% of the ROIs goes into the training set in each fold.

#### 2.4.1 *Fusion of classifiers*

The fusion of classifiers aimed to achieve better accuracy rates through the combination of the results of each classifier using two distinct approaches: bagging method (equally weighted voting) and fusion by SVM.

In fusion by SVM, the individual classifiers' results were used to train and test an SVM classifier which represented the final result of the fusion procedure.[29] In order to perform this fusion, the dataset was initially split into three different sets: training, testing, and validation. The training set was used to train the individual classifiers, and the testing set was used to test them. Then an SVM classifier was trained with the individual classifiers' results in the testing set. Finally, the SVM classifier was tested with the result of the individual classifiers on the validation set, generating the fusion result.

**Table 1** Matrix $M$ representing the percentage of agreement between two classifiers: $c_i$ and $c_j$.

| | Positive $c_i$ | Negative $c_i$ |
|---|---|---|
| Positive $c_j$ | a | b |
| Negative $c_j$ | c | d |

In order to optimize the fusion of the classifiers' procedure, only the classifiers with high diversity measures should be combined.[29] Each diversity measure computes the agreements or disagreements between pairs of classifiers when dealing with the same dataset. Thus, the diversity measures assess the correlation of classifiers and this information can be used to determine the most appropriate ones to be combined.

Diversity measures are based on a matrix that comprises the relationship between pairs of classifiers, denoted by $M$, with the percentage of hits or misses of the classifiers $c_i$ and $c_j$ (Table 1). The value $a$ reflects the percentage of times that both classifiers present correct results. The values $b$ and $c$, on the other hand, present the percentage in which $c_i$ misses the classification while $c_j$ produces a hit classification or vice-versa. Finally, the value $d$ represents the percentage in which both classifiers present incorrect results.

We considered as diversity measures double fault (DFM), Q statistic (QSTAT), and iterative agreement (IA),[30] defined as follows:

$$\mathrm{DFM}_{i,j} = d, \tag{1}$$

$$\mathrm{QSTAT}_{i,j} = \frac{ad - bc}{ad + bc}, \tag{2}$$

$$\mathrm{IA}_{i,j} = \frac{2(ac - bd)}{(a+b)(c+d) + (a+c)(b+d)}. \tag{3}$$

## 3 Experimental Results and Discussion

Experiments were conducted using texture attributes and classifiers to distinguish WMH based on their etiology using different approaches: to distinguish normal white matter from WMH (L/NL task) and then to distinguish WMH with ischemic etiology from WMH with demyelinating etiology (D/I task), or to design a classifier to distinguish three classes at once: normal white matter, WMH with ischemic etiology, and WMH with demyelinating etiology (NL/D/I task). Experiments measured the texture discrimination of the attributes, the effectiveness of the attribute selection methods, and the classifiers' accuracy during the performance of these different tasks.

Classification tasks were accomplished by SVM, kNN, LDA, and OPF using different parameter configurations. For SVM, we tested both linear and RBF kernels, and the parameters were selected through a grid-search technique. We also simulated different configurations of OPF by changing the distance metric (Euclidean or Manhattan) and kNN by varying $k$ values ($k = 1$, $k = 3$, and $k = 5$). Finally, we performed experiments with different attribute selection techniques, such as PCA and

**Table 2** Best accuracy rates achieved for the proposed tasks by varying the parameter setting and combining available techniques. Indication of the applied attribute selection: (*) decision tree; (**) PCA; () none.

| Task | SVM | kNN | OPF | LDA | Bagging |
|---|---|---|---|---|---|
| L/NL | **99.29 ± 0.01** | 98.23 ± 0.02* | 98.23 ± 0.01* | 98.94 ± 0.02 | 98.59 ± 0.02 |
| D/I | 86.5 ± 0.09 | 83.57 ± 0.07* | 81.23 ± 0.06* | 83.52 ± 0.06** | **87.9 ± 0.05** |
| NL/D/I | **90.09 ± 0.05** | 87.63 ± 0.07* | 84.82 ± 0.06** | 87.61 ± 0.06 | 86.93 ± 0.06 |

Note: SVM, support vector machine; kNN, k-nearest neighbor; OPF, optimum path forest; LDA, linear discriminant analysis, bold value represents the highest accuracy rates.

decision tree, in order to find the best set of attributes to distinguish the different classes.

Among all these possible combinations of techniques and parameter configurations, the best results achieved for each one of the tasks are shown in Table 2. The classifier with the best accuracy rate for the L/NL task was SVM with an RBF kernel ($C = 1000$ and $\gamma = 0.01$), without any attribute selection method. LDA presented an accuracy rate of 98.94%, whereas kNN ($k = 1$) and OPF (distance = Manhattan) achieved identical results with an accuracy rate of 98.23%. In both cases, the classifier was combined with the decision tree for attribute selection.

The best classifier for the D/I task was SVM using an RBF kernel ($C = 10$ and $\gamma = 0.1$) with an accuracy rate of 86.5%. kNN combined with the decision tree presented an accuracy rate of 83.57% using $k = 3$. The LDA classifier with PCA achieved an accuracy rate of 83.52%. The worst result was observed for an OPF classifier with the Manhattan distance combined with the decision tree.

Finally, the SVM classifier with an RBF kernel ($C = 10$ and $\gamma = 0.1$) without any attributes selection technique achieved an accuracy rate of 90.09% when performing the NL/D/I task. Similar accuracy rates were achieved for kNN ($k = 5$) combined with the decision tree and LDA (approximately 87.6%), whereas OPF with the Manhattan distance achieved an accuracy rate of 84.82% combined with PCA.

In addition to the comparison between the achieved accuracy rates, we also analyzed the confusion matrix for each task. When performing the L/NL task, the confusion matrix for SVM showed that two ROIs were incorrectly classified (Table 3), whereas the OPF confusion matrix indicated that only five samples were incorrectly classified; two WMH ROIs were classified as normal white matter, whereas three ROIs were classified as WMH; however, they were actually normal white matter (Table 4). Thus, the difference between the classifiers that achieved the highest and the lowest accuracy rates in this task was not representative.

**Table 3** Normalized confusion matrix of L/NL task using the SVM classifier which achieved the highest accuracy rate.

| | | Predicted classes | |
|---|---|---|---|
| | | Lesion (L) | Nonlesion (NL) |
| Actual classes | Lesion (L) | 0.7279 (206/283) | 0.0035 (1/283) |
| | Nonlesion (NL) | 0.0035 (1/283) | 0.2651 (75/283) |

The confusion matrix of the classifiers with higher and lower accuracy rates when performing the D/I task was compared (Tables 5 and 6). The confusion matrix for SVM presented a different error distribution as there was a higher number of ROIs with ischemic WMH classified as demyelinating WMH than vice versa. On the other hand, OPF errors occurred with similar probabilities between classes since the number of demyelinating WMH classified as ischemic was similar to the number of ischemic WMH classified as demyelinating.

Finally, we analyzed the confusion matrix for SVM when performing the NL/D/I task with no attributes' selection method (Table 7). The results suggested that the D/I task was more complex than the L/NL task, since the major number of misclassified samples occurs between demyelinating and ischemic classes.

**Table 4** Normalized confusion matrix of L/NL task using the OPF classifier which achieved the lowest accuracy rate.

| | | Predicted classes | |
|---|---|---|---|
| | | Lesion (L) | Nonlesion (NL) |
| Actual classes | Lesion (L) | 0.7244 (205/283) | 0.0070 (2/283) |
| | Nonlesion (NL) | 0.0106 (3/283) | 0.258 (73/283) |

**Table 5** Normalized confusion matrix of D/I task using the SVM classifier which achieved the highest accuracy rate.

| | | Predicted classes | |
|---|---|---|---|
| | | Demyelinating (D) | Ischemic (I) |
| Actual classes | Demyelinating (D) | 0.6425 (133/207) | 0.0483 (10/207) |
| | Ischemic (I) | 0.0869 (18/207) | 0.2223 (46/207) |

**Table 6** Normalized confusion matrix of D/I task using the OPF classifier which achieved the lowest accuracy rate.

| | | Predicted classes | |
|---|---|---|---|
| | | Demyelinating (D) | Ischemic (I) |
| Actual classes | Demyelinating (D) | 0.5894 (122/207) | 0.1014 (21/207) |
| | Ischemic (I) | 0.0869 (18/207) | 0.2223 (46/207) |

**Table 7** Normalized confusion matrix of the NL/D/I task using SVM classifier which achieved the highest accuracy rate.

| | | Predicted classes | | |
|---|---|---|---|---|
| | | Nonlesion (NL) | Demyelinating (D) | Ischemic (I) |
| Actual classes | Nonlesion (NL) | 0.2685 (76/283) | 0 (0/283) | 0 (0/283) |
| | Demyelinating (D) | 0.0035 (1/283) | 0.4665 (132/283) | 0.0353 (10/283) |
| | Ischemic (I) | 0 (0/283) | 0.0601 (17/283) | 0.1661(47/283) |

Only 1 of the 28 errors occurs between normal white matter and WMH. In addition, it is possible to check the similarities between performing WMH etiology-based classification in one step (NL/D/I task) or two steps (L/NL and D/I tasks). For example, the number of demyelinating WMH classified as ischemic WMH is 10 on both D/I (Table 5) and NL/D/I (Table 7) tasks when using the SVM classifier.

All these results were achieved using 10-fold ROI-wise cross-validation, but we also performed experiments that include a patient-wide folding. In this experiment, 75% of the brains were used in the training and the remained 25% of the brains generated the testing set. This fourfold patient-wide was chosen in order to guarantee a reasonable number of patients in each fold. The results using ROI-wise and patient-wise folding were similar. This suggests that the usage of different regions of the same subject into testing and training sets is valid and does not represent an over fitting.

The combination of image processing and pattern recognition techniques presented a high accuracy rate in the solution of an L/NL task, providing a good start point to the following classification of WMH ROIs according to their etiology. This indicates that the texture attributes extracted from ROIs were representative and made classification easier to accomplish. The classifier with the worst result presented an accuracy rate of 98.23%, approximately 1% lower than the accuracy rate obtained by the best classifier. Indeed, in a practical application based on texture attributes that aims to distinguish normal white matter from WMH as a first step, any proposed classifier (SVM, LDA, kNN, or OPF) would be suitable, since they all presented high accuracy rates. In this case, the choice of the method to be used is based on other criteria such as algorithm implementation complexity and processing time, among others.

The difference between the best and the worst classifiers in the D/I task was approximately 5%, indicating that the most suitable classifier for this task is SVM without any attribute selection method (86.5%). The LDA classifier achieved the second highest accuracy rate (approximately 3% lower than the SVM accuracy rate), which is similar to SVM and has been widely implemented and documented in the literature. Thus, our results indicated that SVM is the optimal method to accomplish the D/I task.

The results also confirmed that OPF and 1NN classifiers present similar accuracy rates.[28] For the NL/D/I task, we obtained an accuracy of 83.75% for the OPF classifier using the Euclidean distance and 84.11% for the 1NN classifier. These rates confirmed the similarity between classifiers, with the 1NN accuracy rate slightly higher.

Experiments on the NL/D/I task, in which the L/NL and the D/I tasks were performed simultaneously, presented a similar number of errors as they did when tasks were performed independently. On both approaches, it was possible to verify the same error distribution by analyzing the corresponding confusion matrices. Thus, this combination represents a comprehensive, efficient approach that is suitable for practical applications assessing WMH etiology.

In addition to the analysis of the most effective results achieved across these separate tasks, we also studied the influence of preprocessing, attributes' quantization, attribute selection, and the fusion of classifiers on the results. These results are presented in the next subsections.

### 3.1 *Influence of Region of Interest Preprocessing*

The brightness variation issues in MR images are often due to a variation in the acquisition parameters and in the MRI field strength. Since our dataset was acquired with the same MRI equipment using the same acquisition parameters, the preprocessing step was not considered a fundamental part of the developed method. Nevertheless, we experimented with a preprocessing technique since it is a step typically applied before texture analysis.

Before the second step of the proposed method (Fig. 1), we applied a preprocessing method through the normalization of the ROI into the range of values $[\mu \pm 3\sigma]$, where $\mu$ represents the mean of gray levels, and $\sigma$ represents the standard deviation of the analyzed ROI. This normalization modifies the gray-level distribution inside the ROI, but does not preserve the relative variation between two gray levels, thus causing significant variations in the texture attributes.

We achieved the best results using no preprocessing step (Table 8). This indicates that the variation in gray levels of the pixels causes a modification in the texture attributes, making them less effective in differentiating normal white matter and WMH and in the distinguishing different types of WMH, since this normalization reduces the sensitivity which degrades the texture information.[31] We conclude that this normalization makes more difficult to accomplish the classification step,

**Table 8** Accuracy rate achieved by the proposed tasks with or without the preprocessing step (normalization of regions of interest).

| Task | Classifier | No preprocessing | Normalization |
|---|---|---|---|
| L/NL | SVM | **99.29 ± 0.01** | 96.1 ± 0.02 |
| D/I | SVM | **86.5 ± 0.09** | 82.57 ± 0.08 |
| NL/D/I | SVM | **90.09 ± 0.05** | 83.78 ± 0.05 |

Note: Bold values represent the highest accuracy rates.

**Table 9** Accuracy rates achieved for the proposed tasks by changing the quantization parameter using SVM classifier.

| Task | 2 bits/pixel | 3 bits/pixel | 4 bits/pixel | 5 bits/pixel | 6 bits/pixel |
|---|---|---|---|---|---|
| L/NL | $96.46 \pm 0.02$ | $99.28 \pm 0.01$ | $\mathbf{99.29 \pm 0.01}$ | $98.94 \pm 0.02$ | $98.22 \pm 0.02$ |
| D/I | $75.88 \pm 0.01$ | $85.5 \pm 0.05$ | $\mathbf{86.5 \pm 0.09}$ | $84.52 \pm 0.09$ | $82.12 \pm 0.08$ |
| NL/D/I | $82.35 \pm 0.04$ | $89.02 \pm 0.04$ | $\mathbf{90.09 \pm 0.05}$ | $87.63 \pm 0.06$ | $86.89 \pm 0.07$ |

Note: Bold values represent the highest accuracy rates.

because classifier efficiency is based on the representative degree of texture information.

### 3.2 *Influence of Attributes Quantization*

The dimension of the co-occurrence matrix and run-length matrix is intrinsically related to the quantization of the analyzed image. Quantized images at a rate of 8 bits/pixel present 256 possible gray levels, thus their co-occurrence matrix dimension is $256 \times 256$ pixels. Such matrices are usually largely made up of values of zero, since the majority of combinations of gray level do not occur, and the statistical attributes extracted from those matrices lose their capability to represent texture patterns.

The co-occurrence matrix approximates the joint probabilities of all pair-wise combinations of gray levels. Owing to the extreme number of elements with a value of zero, it produces a poor approximation.[32] A similar analysis could be applied for the run-length matrix.

As some statistical attributes extracted from those matrices are not efficient for classification using a large number of gray levels, we performed a study in order to analyze the ideal rate of quantization for the proposed tasks. We performed experiments varying the quantization rate from 2 to 6 bits/pixel. Results indicated that the quantization rate that generated the best result was 4 bits/pixel (Table 9) and that quantization parameters equal to two, three, five, or six generated texture attributes with a lower representation capability.

### 3.3 *Influence of Attribute Selection*

We performed experiments using two different attribute selection methods: decision tree and PCA. Results indicated that the use of an attribute selection method did not increase the accuracy of SVM, which is the best classifier for each proposed task (Table 10). This could be explained by the fact that SVM assigns weights to the most relevant attributes, thus no relevant attributes have a lower influence on the trained classifier.

**Table 10** Accuracy rates achieved for the best classifiers of each proposed task by using a different attribute selection method or even using no attribute selection method.

| Task | Classifier | No attribute selection | Decision tree | PCA |
|---|---|---|---|---|
| L/NL | SVM | $\mathbf{99.29 \pm 0.01}$ | $99.29 \pm 0.01$ | $99.29 \pm 0.01$ |
| D/I | SVM | $\mathbf{86.5 \pm 0.09}$ | $86.5 \pm 0.09$ | $86 \pm 0.09$ |
| NL/D/I | SVM | $\mathbf{90.09 \pm 0.05}$ | $87.58 \pm 0.06$ | $90.09 \pm 0.06$ |

Note: Bold values represent the highest accuracy rates.

Based on the results achieved for all classifiers performing the D/I task (Table 11), it is evident that the accuracy rate of the classifiers LDA, OPF, and kNN improved when the attribute selection method was used.

There was no overall best attribute selection method as the results varied according to the classifier and the performed task. While OPF and kNN classifiers achieved the best accuracy rates with the decision tree, LDA classifier presented the best results using PCA (Table 11).

We also analyzed the discrimination degree of each texture attribute extracted from ROIs, while performing the tasks L/NL (Fig. 3), D/I (Fig. 4), and NL/D/I (Fig. 5). It was possible to verify through discrimination analysis that the texture attributes extracted from the co-occurrence matrix and the histogram were

**Table 11** Accuracy rates achieved on the D/I task when varying the attribute selection method.

| Classifier | No attribute selection | Decision tree | PCA |
|---|---|---|---|
| SVM | $\mathbf{86.5 \pm 0.09}$ | $86.5 \pm 0.09$ | $86 \pm 0.09$ |
| kNN | $81.19 \pm 0.09$ | $\mathbf{83.57 \pm 0.07}$ | $81.19 \pm 0.09$ |
| OPF | $80.26 \pm 0.09$ | $\mathbf{81.23 \pm 0.06}$ | $77.21 \pm 0.09$ |
| LDA | $79.11 \pm 0.07$ | $81.57 \pm 0.08$ | $\mathbf{83.52 \pm 0.06}$ |

Note: Bold values represent the highest accuracy rates.

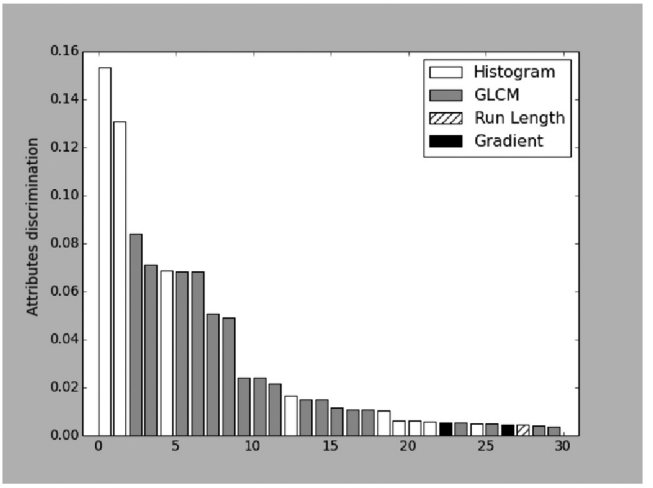


**Fig. 3** Representation of the attributes discrimination computed by the decision tree algorithm while performing the L/NL task (in decreasing order). Only the 30 most discriminating attributes are shown for visualization purposes.

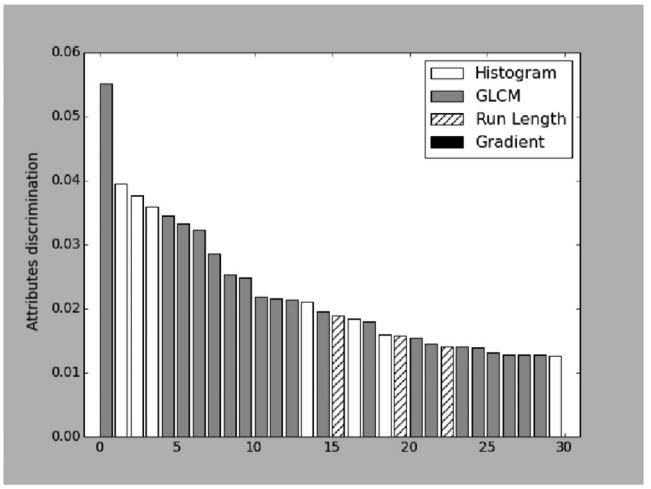


**Fig. 4** Representation of the attributes discrimination computed by the decision tree algorithm while performing the D/I task (in decreasing order). Only the 30 most discriminating attributes are shown for visualization purposes.

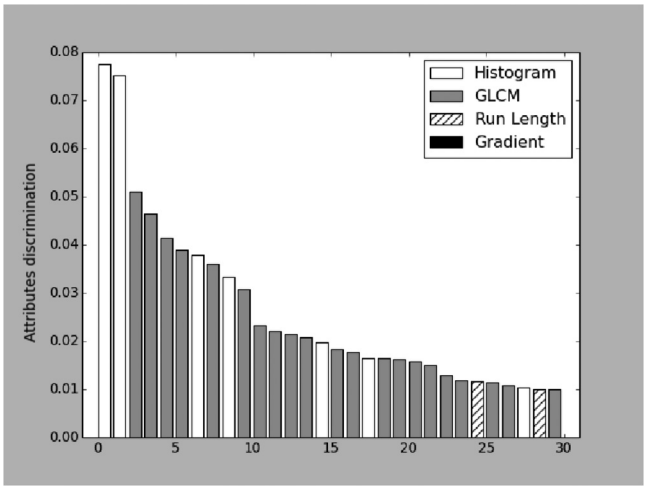


**Fig. 5** Representation of the attributes discrimination computed by the decision tree algorithm while performing the NL/D/I task (in decreasing order). Only the 30 most discriminating attributes are shown for visualization purposes.

**Table 12** Identification of the three most discriminating texture attributes of each task, extracted from Figs. 3–5.

| Relevance | L/NL task | D/I task | NL/D/I task |
|---|---|---|---|
| First | Percentile 99% (histogram) | Entropy (GLCM) | Percentile 50% (histogram) |
| Second | Percentile 50% (histogram) | Percentile 1% (histogram) | Percentile 99% (histogram) |
| Third | Homogeneity (GLCM) | Mean (histogram) | Sum of squares (GLCM) |

the most frequent attributes of groups in all the proposed tasks, stressing their relevance in the classification of normal white matter, ischemic WMH, and demyelinating WMH (Figs. 3, 4, and 5, respectively).

The three most important texture attributes for each task were extracted from the histogram and from the GLCM (Table 12). In addition, it is important to note that the same histogram attributes (percentiles 99% and 50%) are the two most relevant texture attributes for both the L/NL and NL/D/I tasks.

### 3.4 *Influence of Fusion of Classifiers*

Experiments were performed using all classifiers and by selecting some classifiers through the diversity metrics and performing the fusion of classifiers using a bagging method or fusion by SVM. We tested the fusion of classifiers only on the D/I task because it was the most complex task that was implemented. The diversity metrics DFM and QSTAT indicated LDA and kNN as the most diverse pair of classifiers. Further, the IA metric indicated the classifiers SVM and kNN as the most diverse pair (Fig. 6).

In summary, LDA, kNN, and SVM were selected as the most diverse classifiers for at least one metric, whereas OPF was not selected. In addition, kNN presented the highest degree of diversity because it was selected by all metrics. The combination of these diverse classifiers (LDA, kNN, and SVM) using the bagging method achieved an accuracy rate of 87.9%, the maximum

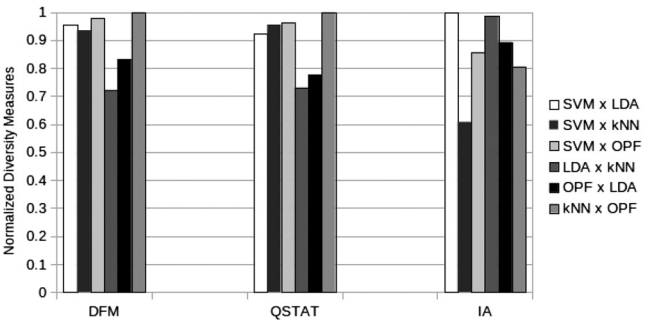


**Fig. 6** Selection of the most diverse pair of classifiers by using the metrics double fault (DFM), iterative agreement (IA), and Q statistic (QSTAT). The lowest value of each metric indicates the pair of classifiers with highest degree of diversity.

**Table 13** Accuracy rates obtained on D/I task when performing fusion of classifiers using the best configuration of each individual classifier.

| Fusion method | LDA | OPF | kNN | SVM | Accuracy |
|---|---|---|---|---|---|
| Bagging | x | x | x | x | $85.52 \pm 0.06$ |
| SVM | x | x | x | x | $83.14 \pm 0.08$ |
| Bagging | x | o | x | x | $\mathbf{87.9 \pm 0.05}$ |
| SVM | x | o | x | x | $83.09 \pm 0.06$ |
| Bagging | x | x | o | x | $85.54 \pm 0.06$ |
| Bagging | o | x | x | x | $84.09 \pm 0.07$ |

Note: Bold values represent the highest accuracy rates.

accuracy rate reached when using the fusion of classifiers' methods. The best accuracy rates obtained for the fusion of all classifiers or by selecting some classifiers through the diversity metrics are summarized in Table 13.

The fusion by SVM generated lower results when compared with bagging results, possibly because it was used for a limited number of classifiers. It is important to note that for some classifiers' combinations, such as the combination of SVM, OPF, and kNN, fusion produced a lower accuracy rate than the individual classifiers.

Other combinations, such as the fusion of LDA, kNN, and SVM, presented a higher accuracy rate than individual classifiers, highlighting the efficacy of using fusion and diversity metrics. This combination also achieved a better performance than the combination of all classifiers, indicating that the inclusion of OPF diminished fusion performance. This is probably due to the majority vote (bagging) in the inclusion of a fourth classifier (even number of classifiers) producing an inconsistent decision.

## 4 Conclusion and Perspectives

The main goal of this study was to determine if it is possible to distinguish WMH based on their etiology (i.e., demyelinating or ischemic nature) using image processing and pattern recognition techniques. The proposed method comprises the application of texture analysis to the ROI, followed by a classification step to distinguish the classes: normal white matter, WMH with demyelinating nature, or WMH with ischemic nature.

Our experiments have shown that the combination of texture analysis and classifiers is suitable to perform the proposed task. We carried out two different approaches to solve this problem: to distinguish normal white matter, ischemic WMH, and demyelinating WMH at once (NL/D/I task) or to perform it in separated steps (L/NL and D/I tasks). The results suggest that the combination of the L/NL and D/I tasks in one single step is suitable for practical applications of WMH analysis, as the NL/D/I task presents a similar number of errors compared with performance of the tasks independently, and thus, represents an efficient approach.

Among the classifiers compared in the study, SVM presented the best accuracy rates in the execution of all proposed tasks. In addition, through discrimination degree analysis, we can conclude that the texture attributes extracted from the co-occurrence matrix and from the histogram are the most relevant ones for accomplishing the proposed tasks.

In addition, it can also be noted that an attribute selection step is desirable, as all classifiers increased their accuracy rates when combined with an attribute selection method except the SVM classifier. This may be explained by the fact that the SVM performed this procedure intrinsically during the construction of the classifier model.

In future works, we intend to eliminate the manual ROI extraction step by developing an automatic method for WMH segmentation based on texture classification motivated by the high accuracy rates achieved to distinguish normal white matter from WMH (L/NL task). A WMH segmentation method based on texture classification would require almost no parameters to be tunned, and would neither require additional MRI modalities nor atlases. We are also planning to expand the image database by including images acquired with different MRI equipment in order to validate the method and to verify its robustness. In this case, it will be necessary to analyze and minimize the influence of MRI acquisition conditions in the texture analysis by using preprocessing techniques such as inhomogeneity correction, image normalization, and quantization. The chosen technique must preserve the discrimination degree of the attributes. Finally, note that the classifiers used to perform the D/I task presented the worst accuracy rates, and in order to improve this method and to better understand the problem, we are planning further investigation on the extraction of other attributes such as scale invariant feature transformation and local binary patterns.

## Acknowledgments

The authors would like to thank the Federal Agency of Support and Evaluation of Postgraduate Education (CAPES) and the São Paulo Research Foundation (FAPESP) processes 2012/21826-1 and CEPID2013/07559-3 for providing financial support.

## References

1. M. W. Vernooij et al., "Incidental findings on brain MRI in the general population," *N. Engl. J. Med.* **357**, 1821–1828 (2007).
f A. Spilt et al., "Age-related changes in normal-appearing brain tissue and white matter hyperintensities: more of the same or something else?," *Am. J. Neuroradiol.* **26**(4), 725–729 (2005).
3. F. M. G. Dixon and N. Raz, "The cognitive correlates of white matter abnormalities in normal aging: a quantitative review," *Neuropsychology* **14**, 224–232 (2000).
4. R. P. Kloppenborg et al., "Presence and progression of white matter hyperintensities and cognition: a meta-analysis," *Neurology* **82**, 2127–2138 (2014).
5. K. W. Kim, J. R. MacFall, and M. E. Payne, "Classification of white matter lesions on magnetic resonance imaging in elderly persons," *Biol. Psychiatry* **64**(4), 273–280 (2008).
6. M. D. Ginsberg, "Neuroprotection for ischemic stroke: past, present and future," *Neuropharmacology* **55**, 363–389 (2008).
7. A. Compston and A. Coles, "Multiple sclerosis," *Lancet* **372**, 1502–1517 (2008).
8. S. Appenzeller et al., "Quantitative magnetic resonance imaging analyses and clinical significance of hyperintense white matter lesions in systemic lupus erythematosus patients," *Ann. Neurol.* **64**, 635–643 (2008).
9. C. Loizou et al., "Quantitative analysis of brain white matter lesions in multiple sclerosis subjects," in *Int. Conf. Information Technology and Applications in Biomedicine-IEEE*, pp. 1–4, IEEE, Larnaca (2009).
10. P. Anbeek et al., "Probabilistic segmentation of white matter lesions in MR imaging," *NeuroImage* **21**, 1037–1044 (2004).
11. V. Ithapu et al., "Extracting and summarizing white matter hyperintensities using supervised segmentation methods in Alzheimer's disease risk and aging studies," *Hum. Brain Mapp.* **35**(8), 4219–4235 (2014).
12. Z. Lao et al., "Computer-assisted segmentation of white matter lesions in 3D MR images, using support vector machine," *Acad. Radiol.* **15**, 300–313 (2008).
13. D. G. Zimring et al., "Automatic detection and characterization of multiple sclerosis lesions in brain MR images," *Magn. Reson. Imaging* **16**, 311–318 (1998).
14. A. Zijdenbos et al., "Morphometric analysis of white matter lesions in MR images: method and validation," *IEEE Trans. Med. Imaging* **13**, 716–724 (1994).
15. S. Klöppel et al., "A comparison of different automated methods for the detection of white matter lesions in MRI data," *NeuroImage* **57**, 416–422 (2011).
16. R. Lotufo et al., "Adessowiki: on-line collaborative scientific programming platform," in *5th Int. Symp. on Wikis and Open Collaboration*, ACM, New York (2009), http://adessowiki.fee.unicamp.br/adesso/.
17. R. Woods and R. C. Gonzalez, *Digital Image Processing*, Edgard Blucher, Ed., Prentice Hall, New Jersey (2000).
18. R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst., Man Cybern.* **3**, 610–621 (1973).
19. M. M. Galloway, "Texture analysis using gray level run lengths," *Comput. Graphics Image Process.* **4**, 172–179 (1975).

20. F. Pedregosa et al., "Scikit-learn: machine learning in python," *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
21. K. Pearson, "On lines and planes of closest fit to systems of points in space," *Philos. Mag.* **2**(11), 559–572 (1901).
22. T. P. Minka, "Automatic choice of dimensionality for PCA," in *NIPS*, pp. 598–604 (2000).
23. J. Han and M. Kamber, *Data Mining: Concepts and Techniques*, Elsevier, San Francisco, California (2006).
24. L. Breiman et al., *Classification and Regression Trees*, Wadsworth International Group, Belmont, California (1984).
25. C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.* **20**(3), 273–297 (1995).
26. F. Papa, A. X. Falcão, and C. T. N. Suzuki, "Supervised pattern classification based on optimum-path forest," *Int. J. Imaging Syst. Technol.* **19**, 120–131 (2009).
27. R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, Wiley, New York (2001).
28. R. Souza, L. Rittner, and R. Lotufo, "A comparison between k-optimum path forest and k-nearest neighbors supervised classifiers," *Pattern Recognit. Lett.* **39**, 2–10 (2014).
29. F. A. Faria et al., "Automatic classifier fusion for produce recognition," in *XXV SIBGRAPI—Conf. Graphics, Patterns and Images*, pp. 252–259, IEEE, Ouro Preto (2012).
30. L. I. Kuncheva and C. J. Whitaker, "Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy," *Mach. Learn* **51**, 181–207 (2003).
31. R. A. Lerski et al., "MR image texture analysis—an approach to tissue characterization," *Magn. Reson. Imaging* **11**, 873–887 (1993).
32. D. A. Clausi, "An analysis of co-occurrence texture statistics as a function of grey level quantization," *Can. J. Remote Sens.* **28**(1), 45–62 (2014).

**Mariana Leite** is a PhD student in the Department of Computer Engineering and Industrial Automation at the State University of Campinas, Campinas, Brazil. In 2013, she received a master's degree in the same department. Her research interests include image processing, medical imaging, and pattern recognition.

**Letícia Rittner** graduated in electrical engineering (1994), earning MSc (2004) and PhD (2009) degrees from the University of Campinas, with a 6-month internship at the Montreal Neurological Institute, McGill University, Canada. She was a postdoc at the School of Medicine at the University of Pennsylvania. Presently, she is an assistant professor at the School of Electrical and Computer Engineering, University of Campinas, Brazil.

**Simone Appenzeller** received her MD and PhD degrees from the Faculty of Medical Science, State University of Campinas. She was a postdoc at McGill University, Canada, and at Stavanger Hospital, Norway. Currently, she is an associate professor in the Rheumatology Department, Faculty of Medical Science, State University of Campinas. Her research focuses on neurologic manifestations and neuroimaging in rheumatic diseases.

**Heloísa Helena Ruocco** graduated in medicine from the University of Uberlândia (1990). She received her PhD degree from the University of Campinas (2000), with fellowship at the Université Louis Pasteur de Strasbourg. She finished her second postdoc in neuroscience at the University of Campinas in 2005. She is currently a medical assistant at the Faculty of Medicine of Jundiaí, acting as a multiple sclerosis clinic coordinator.

**Roberto Lotufo** obtained an electronic engineering degree from Instituto Tecnológico de Aeronáutica, Brazil, in 1978; his MSc degree from the University of Campinas, Brazil, in 1981; and his PhD degree from the University of Bristol, U.K., in 1990 in electrical engineering. He has been a professor at the School of Electrical and Computer Engineering, University of Campinas, Brazil, since 1981. He has published over 73 international works with 1040 citations.