Automatic white matter lesion detection and segmentation on Magnetic Resonance Imaging: A review of past and current state-of-the-art

K. H. Ong[1,2], S. M. Shamsuddin[1,2], S. Sulaiman[1,2], A. Ali[1,2], N. R. Mohd Zain[4] and S. P. Hang[3]

[1]*UTM Big,Data Centre*
[2]*School of Computing,Faculty of Engineering*
[3]*Department of Mathematical Sciences, Faculty of Science*,
Universiti Teknologi Malaysia, UTM Skudai, 81310 Johor, Malaysia.
e-mail:ongkokhaur@gmail.com; mariyam@utm.my; sarina@utm.my;
aida@utm.my; sphang@utm.my
[4]*Radiology Department*,
Hospital Kuala Lumpur,50586 Jalan Pahang, Wilayah Persekutuan Kuala
Lumpur, Malaysia.
e-mail:nrosezain@gmail.com

Abstract

White matter lesion (WML) is an abnormal tissue occurring in white matter. It indicated the damage of the myelin sheath that used to surround the axon of a neurone. This resulting neurological and vascular disorder occur in the patient, also commonly developed in the healthy brain of elderly. Magnetic Resonance Imaging is a non-invasive medical equipment preferred choice by the clinician to diagnose and observed the injury of brain tissue. However, WML quantitative assessment and analyse on the large volume of MR imaging is a challenge. In this paper, we provide an intensive review of the past and recent WML delineation and detection methods. This review included visual scoring assessment, a common preprocessing step for WML segmentation, false positive elimination, and the latest automatic WML segmentation approaches will be presented.

1   Introduction

A white matter lesion (WML) is a damaged region in the white matter tissue. It can be observed as a "white spot" or a "cluster of white spots" within the white matter region using magnetic resonance imaging technique. Therefore, it is also commonly seen as white matter hyper-intensity (very bright area). WML are often developed in healthy brains of elderly patients with neurological and vascular disorders. Assessment of WML using MRI is critical to help the clinician a second opinion to provide the right treatment for respective patients. However, visual scoring evaluation and manual delineation on MRI are very challenging as the assessment result is subjective and not comparable to the scoring from another clinician. A quantitative approach such as volume of WML load is a way to resolve the drawback of visual scoring assessment. In our opinion, WML assessment using the advancement of computer aid diagnosis (CAD) application enables fast, consistent, precise and comparable results.

In the recent decade, automatic white matter lesion segmentation algorithms have gained attention from the interdisciplinary researchers. It is becoming a well known technique because the manual delineation and semi automatic white matter lesion segmentation approaches have been always suffering from several drawbacks. Apparently, manual delineation requires a radiologist's intervention to delineate lesions. Hence, this is time consuming and tedious job for radiologists. Therefore, the segmentation is an important technique required to be applied to speed up the analysis of WML volume and provide better understanding of the brain diseases such as multiple sclerosis, vascular dementia [4-8], Ischemic Strokes [7,9] and Alzheimer's disease [6,8,10].The segmented results of these segmentation approaches will be normally led by intra-observer variability (the same subject of study delineated by the same radiologist at different times) and inter-observer variability (the same subject of study delineated by different radiologists). Semi-automatic approaches are proposed to minimise the variability by allocating the "seed" point or region of interest (ROI). The "seed" region will be grown to segment the desired region automatically. The often used existing methods are region growing and level-set algorithm. However, these algorithms are still time consuming and tedious to allocate many lesions in a series of images. Thus, they are not suitable to be performed in a very large scale clinical study. On the other hand, automatic WML segmentation approaches have gained much attention since they are a fast computation approach, give consistent result and do not require user intervention. They are also able to mimic the performance of a neuro-radiologist to identify and quantify white matter lesion load.

In this study, recent work on automatic white matter lesion segmentation approaches is critically reviewed. The aim of this work is to present the existing research on automatic WML segmentation included their pre- and post-processing methods especially false positive reduction method.

2   Computer Aided Detection and Diagnosis System for WML Load Quantification

The advancement of technique and imaging technology that facilitates many automatic WML segmentation approaches have been developed and studied. This is because the clinical assessment of WML using visual scoring procedure is not efficient in large scale WML progression studies. Consequently, WML load quantification using computer-aided detection has become essential to speed up the WML quantitative analysis. Semi-automated and automated WML segmentations are two common approaches used to segment and quantify WML. Usually, the semi-automated approaches involve the intervention of a neuro-radiologist. In this process, a "seed-point" of a WML candidate is allocated by a neuro-radiologist on the desired image. The "seed point" will be calculated based on the image analysis algorithm to "grow" regions and then segment the WML on the target image. The techniques which had been used to "grow" regions were region-growing, level set algorithm and fuzzy connectedness[11-14,5]. However, semi-automatic approaches remain tedious and labour intensive especially for large scale WML progression studies. In contrast, automatic WML segmentation is an alternative approach to segment and calculate the WML load quantitatively without expert intervention. In the next section, the automatic WML segmentation approaches will be further discussed. The pre and post processing (false positive reduction) methods will also be discussed in the common automatic WML segmentation framework.

2.1   Preprocessing

Commonly, automatic white matter lesion segmentation framework consists of preprocessing step, WML detection, WML segmentation method and post processing steps. Preprocessing procedure is the prior action to enhance the raw image that will be used in the detection and segmentation stages. The typical pre-processing methods consist of skull stripping, MR intensity standardisation, MR intensity inhomogeneity correction and image registration. Depending on the WML segmentation framework, these preprocessing procedures vary from one another. Details of each common preprocess method is described in following:

Skull stripping algorithm is also known as brain extraction algorithm. It is mainly applied to T1-w sequence brain imaging because T1-w sequence shows the prominence of brain structure and hard tissues. The principle of skull stripping algorithm is to first remove the intensity voxels of the skull. Hence, intensity voxel of brain tissues can be extracted and this is followed by the classification of brain tissues and brain pathologies that could be further analysed precisely such as white matter lesions and brain tumours. There are three categories of algorithms commonly found in skull stripping algorithm. They are deformable model-based

method [15,5], morphology-based method [16], and histogram-based method [17]. Brain extraction tool based on the deformable model approaches has been investigated and developed by Smith [15]. They proposed to deform a sphere's surface as initial shape using active contours technique and propagate to the boundary of the brain. Hence, the voxel of whole brain tissue was segmented. On the other hand, skull stripping algorithm using level set technique was presented by Zhuang *et al.* [5]. In their framework, a circle shape was initialised in the middle slice of the entire brain slices. The first segmented brain slice was used as the first mask to propagate into entire brain associated with subject age information. This parameter enabled the estimation of the iteration in level-set to remove skull in the following slice accurately. Mathematical morphology was applied to tackle skull stripping problem presented by Dogdas *et al*. [16]. This method combined the four morphology operations namely dilation, erosion, opening and closing. Furthermore, these morphology operation using 3-dimensional structure elements such as cube, cross and octagon were employed to perform skull segmentation. In another work, a robust skull stripping method using histogram analysis was reported by Balan *et al*. [17]. They proposed a new method to partition the histogram and analyse the grey level range of the brain tissue. The method was validated using public available dataset and compared with Brain Extraction Tool (BET) and Brain Surface Extractor (BSE). The authors concluded that the earlier was less specific while the latter was less sensitive.

Intensity inhomogeneity correction is also referred to as bias field correction. The algorithm is mainly used to filter the image intensity non-uniformities caused by sensitivity variations from MR scanner receiver coil. The image intensity with non-uniformities is hardly detected by bare eyes, yet it influences the accuracy of brain segmentation results. Two well-known methods have been employed to overcome this complex problem [18-19].

Intensity standardisation is also commonly known as intensity normalisation. The algorithm is mainly applied to solve the standard image intensity scale in MRI. Intensity voxels of brain tissue in the same region are not comparable for each scan, even when the MR images are acquired from the same MR modality and protocol. Unlike CT images, Hounsfield unit (HU) in density value is provided to differentiate intensity value of various tissues. However, MRI shows the image intensity contrast to distinguish among different tissues based on visual observation. Eventually, the lack of standardisation in intensity scale increases the difficulty in developing a learning model to automatically perform the WML segmentation and analysis. In the recent decade, several image intensity standardisation algorithms have been proposed by many researchers [20-23]. They are landmark-based approach [22,24], kullback-leibler Divergence-based approach [25] and patch matching-based approaches [23]. These image intensity standardisation algorithms

were applied in supervised-based WML segmentation algorithms [26-31]. The main reason was these supervised approaches used features extracted from voxel intensity. Therefore, the image intensity needed to be standardised for each scan. Thus, the accurate and robustness of the results could be obtained. Evaluation of various image intensity standardisation algorithm was reported by Bergeest and J¨ager [32]. More specifically, Shah *et al*. [33] evaluated these methods on MRI with WML analysis study. Coincidentally, the method proposed by Nyu and Udupa [22] achieved the best performance in both review papers. Notably, their method enabled fast computation and was easily customised into various anatomy MRI images. The details of the landmark-based intensity standardisation can be obtained in Nyu and Udupa [22] and Nyul *et al*. [24].

Image registration is the crucial step of success of automated segmentation or classification of brain image using the multi-sequence image. Image registration technique is often required in the task of aligning two or more images based on a spatial correspondence of their common image feature. There are several applications often applied to clinical imaging as listed below:

  a. an image of the same subject that can be combined with different modalities/sequences
  b. aligning images of various subjects in cohort studies
  c. aligning temporal sequences of images in between the scans

A typical study of image registration involves a scene and a reference image. A scene is an image that performs transformation and matches into reference image. In order to align a scene and a reference image, image registration technique consists of two important steps which are transformation model in the first step, followed by similarity metric.

In the transformation model step, different transformations such as translate, rotate, scale and affine are considered as rigid transformation. Whereas, the B-spline and thin-plate splines transformations which can be deformed into the content is considered as non-rigid (or known as elastic) transformation. Therefore, it is an important step to define what transformation model used to align. Similarity, metric will be performed subsequently, where it measures the degree of matching scene and reference image.

Alignment between two images requires a spatial transformation to map the corresponding specified area to another one. Therefore, features are the essential elements in image registration. Features are defined as the properties that are able to describe and represent image content in a specific area. For instance, shape, spatial information, and edge are the common features employed in image registration, which are commonly known as feature-based approaches. Besides, image registration techniques using pixel intensity alone are considered as intensity-based approaches. Image registration techniques have been applied as preprocessing step in white matter lesions segmentation [31,34,29,35-37]. Many of

them often employed image registration to multi-sequences (T1-w, T2-w, Proton Density and FLAIR) and standard brain atlas to determine true lesion from various sources.

2.2   Algorithms of Automatic White Matter Lesion Segmentation

There are two comprehensive reviews of automated white matter lesion segmentation methods which have been reported very recently [38-41]. The supervised learning methods that employ features using intensity values and spatial information are often applied for automated white matter hyperintensities segmentation concluded by Goceri *et al.* [41]. White matter hyperintensities segmentation can be classified into supervised learning algorithm, unsupervised algorithm and semi-automated algorithm as suggested by Caligiuri *et al.* [39]. A good algorithm should incorporate a proper preprocessing stage, utilise the multimodal image data, make use of spatial information and good automated method to remove the false positive WML concluded in their study.

Based on the literature review presented by Garc´ıa-Lorenzo *et al.* [40], automated segmentation approaches are divided into two groups. They are supervised learning and unsupervised learning methods. Supervised learning is defined as the "learn" method or mathematical model constructed based on white matter lesion definition from a set of training database (manually segmented lesion images). Wheras, unsupervised learning is defined as clustering methods to separate voxel images into several tissue clusters such as White Matter(WM), Grey Matter(GM), Cerebrospinal Fluid(CSF) and WML. However, Llad´o *et al.* [38] concluded that supervised learning methods are further divided into two sub-groups, namely supervised strategies based on Atlas and learning from manual segmentation. Supervised strategies are based on Atlas using standard brain atlas spatial information; for example, coordinate of various brain tissues. Therefore, the supervised segmentation process usually requires a registration process to outline the atlas into the target image. The accurate spatial information allows the feature such as intensity and texture obtained certainly to perform voxel classification on the target image.

Supervised strategies based on learning from manual segmentation require intervention from radiologist to label the brain tissues and WML manually. Thus the features information can be collected to construct the classification model accurately. For the unsupervised learning method, Llad´o *et al.* [38] suggest two sub-group can also be further divided. They are unsupervised strategies segmenting tissue and unsupervised strategies segmenting only lesions. Voxels of brain tissues were first clustered, and white matter lesion detected as outliers on each brain

tissues described on unsupervised strategies segmenting tissue. Whereas, unsupervised strategies segmenting only lesions using strategies to classify white matter lesion from other brain tissues by enhancing the features of the lesion in the target image.

Based on the review in our study, the automatic white matter segmentation can be categorised into three different approaches, and they are FLAIR histogram threshold methods, supervised learning methods and unsupervised learning methods. The metrics definition and summary of each study evaluation are reported in Tables 1 and 2 respectively.

Metric	Formula	Good	Worse
DSI	$\dfrac{2 \times TP}{FP + FN + (2 \times TP)}$	1	0
TPR	$\dfrac{TP}{TP + FN}$	1	0
FPR	$\dfrac{FP}{FP + TN}$	0	1
PPV	$\dfrac{TP}{TP + FP}$	1	0

Table 1: The common evaluation metric true positive rate (TPR), false positive rate (FPR), dice similarity index (DSI), and positive predictive value (PPV) are selected to summaries the accuracy for each study in this review. They are defined using notations: true positive (TP), true negatives (TN), false positives (FP) and false negative (FN).

Table 2: Summary of evaluation results from the methods described in the review.

Approaches	Articles	subjects	No# subjects	Ground truth preparation	Evaluation results	TPR	FPR	DSI	PPV
FLAIR histogram threshold approach	Roura *et al.* [42]	Subjects with clinically isolated syndrome (CIS)	70	Manual segmentation	$R_{FLAIR} = 0.95$ $R_{PD} = 0.80$	0.36 0.5	- -	0.3 0.33	0.53 0.62
	Cabezas *et al.* [43]	Subjects with MS lesions	45	Semi-automated segmentation	DSI = 0.50	0.45	0.41	0.50	-
	Yoo *et al.* [44]	Subject with Cognitive Aging and Dementia	48	Manual segmentation	ICC= 0.996	-	-	0.93	-
	Ong *et al.* [1]	Subjects with normal aging	19	Semi-automated segmentation & visual rating	$R=0.85^h$ $R=0.96^f$	0.67	0.53	0.46	-
	De Boer *et al.* [28]	Subjects with normal aging	215	Manual segmentation	EF=0.5	0.79	-	0.72	-
	Souplet *et al.* [30]	Subjects with MS lesions	24	Manual segmentation	VD = 86.5%^{d} AD = 8.2mm^d VD = 55.8%^{e} AD = 7.4mm^e	0.58^d 0.49^e	0.69^d 0.76^e	- -	- -
	Wen and Sachdev [45]	Subjects with normal aging, dementia and other neuropsychiatric disorders.	477	Visual rating	ICC = 0.43 R = 0.79	-	-	-	-
	Jack *et al.* [46]	Subjects with Alzheimer's Disease	10	Manual segmentation	MAE = 6.6% CV=1.4%	-	-	-	-

Table 2: Summary of evaluation results from the methods described in the review. (Continued)

Approaches	Articles	subjects	No# subjects	Ground truth preparation	Evaluation results	TPR	FPR	DSI	PPV
Supervised learning approach	Ghafoorian et al. [47]	Subjects with ageing brain and cerebral small vessel disease	420	Manual segmentation	DSI = 0.79[f] DSI = 0.78[g]	-	-	0.79	-
	Valverde et al. [48]	Subjects with MS lesions	35	Semi-automated segmentation	R = 0.97	0.79	0.36	0.51	65.3
	Zhan et al. [49]	Subjects with diabetes	50	Manual segmentation	TNR = 0.997	0.83	-	0.76	-
	Rincón et al. [50]	Subjects with brain infarct and mild cognitive impairment	28	Manual segmentation	OSR = 0.37 USR = 0.28	-	-	0.69	-
	Griffanti et al. [51]	Subjects with Alzheimer'sDisease and cognitive impairment	85	Manual segmentation	ICC=0.99	-	0.22	0.75	-
	Roy et al. [52]	Subjects with hypertension	24	Manual segmentation	DSI = 0.61[a] DSI= 0.71[b] DCI= 0.76[c]	-	-	0.61[a] 0.71[b] 0.76[c]	-
	Steenwijk et al. [2]	Subjects with hypertension	20	Manual segmentation	ICC = 0.92	0.73	-	0.73	-
	Geremia et al. [53]	Subjects with MS lesions	20	Manual segmentation	-	0.39	-	-	0.4
	Yamamoto et al. [54]	Subjects with MS lesions	6	Manual segmentation	JI = 0.64	0.82	-	0.77	-
	Zacharaki et al. [31]	Subjects with MS lesions	42	Manual segmentation	Visual Quality Evaluation	-	-	-	-
	Akselrod-Ballin et al. [55]	Subjects with MS lesions	25	Semi-automated segmentation	R^2=0.90	0.55	0.02	-	-

Table 2: Summary of evaluation results from the methods described in the review. (Continued)

Approaches	Articles	subjects	No# subjects	Ground truth preparation	Evaluation results	TPR	FPR	DSI	PPV
Supervised learning approach	Kroon *et al.* [56]	Subjects with MS lesions	24	Manual segmentation	VD= 402.5%[d] AD = 9.5mm[d] VD = 469.6%[e] AD = 10mm[e]	0.44[d] 0.41[e]	0.89[d] 0.91[e]	- -	- -
	Lao *et al.* [29]	Subjects with diabetes	50	Manual segmentation	R=0.85[f] SC=0.77[f] R=0.88[g] SC=0.73[g]	- -	- -	- -	- -
	Anbeek *et al.* [57]	Subjects with arterial vascular disease	20	Manual segmentation	DSI=0.8	0.97	0.03	0.80	-

Table 2:Summary of evaluation results from the methods described in the review. (Continued)

Approaches	Articles	subjects	No# subjects	Ground truth preparation	Evaluation results	TPR	FPR	DSI	PPV
Unsupervised learning approach	Zhao et al. [58]	-	-	-	Visual assessment	-	-	-	
	Sudre et al. [59]	Subjects with Type 2 Diabetes	19	Manual segmentation	$R^2 = 0.96$ Lin=0.88 VD = 0.52 AD =6.8	0.37	0.1	0.46	-
	Jain et al. [60]	Subjects with MS lesions	20	Manual segmentation	ICC=0.8	0.57	-	0.69	-
	Simões et al. [3]	Subjects with ageing brain and dementia	40	Manual segmentation	R=0.99 OF = 0.65 EF=0.34	-	-	0.68	-
	Vald´esHern´andez et al. [61]	Subject with mild stroke and ageing brain	20	Manual segmentation	$JI_{MCMxxxVI}$=0.61 $JI_{thresholding}$=0.31	-	-	-	-
	Schmidt et al. [62]	Subjects with MS lesions	70	Manual segmentation	Acc = 0.9995	0.8	0.0003	0.75	-
	Garc´ıa-Lorenzo et al. [63]	Subjects with MS lesions	10	Manual segmentation	R = 0.97 ICC = 0.91	-	-	0.63	-
	Seghier et al. [64]	Subjects with normal aging, stroke & simulated data	64	Manual segmentation	DSI=0.64	-	-	-	0.64
	Khayati et al. [65]	Subjects with MS lesions	7[a] 10[b] 3[c]	Manual segmentation	R=0.93[a] R=0.95[b] R=0.98[c]	0.74	-	0.75	-
	Bricq et al. [27]	Subjects with MS lesions	24	Manual segmentation	VD = 73.0%[d] AD = 6.7mm[d] VD = 51.3%[e] AD = 6.6mm[e]	0.46[d] 0.4[e]	0.51[d] 0.61[e]	- -	- -

Table 2: Summary of evaluation results from the methods described in the review. (Continued)

Approaches	Articles	subjects	No# subjects	Ground truth preparation	Evaluation results	TPR	FPR	DSI	PPV
Unsupervised learning approach	Prastawa and Gerig[66]	Subjects with MS lesions	24	Manual segmentation	VD = 92.7%[d] AD = 32.7mm[d] VD = 92.5%[e] AD = 33.4mm[e]	0.11[d] 0.08[e]	0.62[d] 0.66[e]	- -	- -
	Garc´ıa-Lorenzo et al. [67]	Subjects with MS lesions	24	Manual segmentation	VD = 67.5%[d] AD = 18.3mm[d] VD = 41.4%[e] AD = 19.9mm[e]	0.35[d] 0.38[e]	0.47[d] 0.48[e]	- -	- -
	Admiraal-Behloul et al. [68]	Subjects with higher prevalence of smoking, diabetes, hypertension, and a history of vascular disease.	100	Manual segmentation	DSI=0.75	-	-	0.75	-
	Van Leemput et al. [37]	Subjects with MS lesions	20	Manual segmentation	R=0.98	-	-	0.58	-

[a]: Mild lesion load

[b]: Moderate lesion load

[c]: Severe lesion load

[d]: Evaluated with ground truth created by rater from Boston Children's Hospital

[e]:Evaluated with ground truth created by rater from The University of North Carolina

[f]: Evaluated with ground truth created by human observer 1

Acc: Accuracy

CV: Coefficient of variation

ICC: Intraclass Correlation Coefficient

JI: Jacard Index

Lin: Corresponding Linear Coefficient

EF: Extra fraction

MAE: Mean Absolute Error

OF: Overlap fraction (a.k.a Jacard Index)

OSR: Over segment rate

USR: Under segment rate

[g]: Evaluated with ground truth created by human observer 2

[h]: Evaluated with visual scale rated by human observe

R: Correlation coefficient (Pearson's correlation)

TNR: True negative rate

AD: Average Distance

VD: Volume different

2.2.1 FLAIR histogram threshold approaches

FLAIR histogram threshold approaches usually assemble the intensity distribution to find the best cutoff point from intensity histogram to segment a white matter lesion. In the first study of this strategy, semi-automatic segmentation using FLAIR histogram was suggested by Hirono *et al.* [69]. They used the intensity distribution of white matter voxel to determine white matter lesions. A threshold value defined with 3.5 Standard Deviations (SD) was applied to segment a WML from WM intensity distribution. Besides, false positive of a WML was further removed manually because some voxels intensity of grey matter was present higher than the proposed threshold value. Subsequently, an adaptive threshold cut-off value has been developed by Jack *et al.* [46]. The threshold value to automatic WML segmentation used sophisticated regression model on FLAIR images. In their framework, preprocessing such as mean filter and anisotropic filter were used to compensate for the inhomogeneity and remove noise respectively. Histogram distribution was constructed based on preprocessed images; the left tail of intensity histogram indicated CSF and middle of distribution represented normal brain tissues (WM and GM). Potential WML and image artefacts were presented in the right tail of intensity FLAIR histogram. Properties of intensity distribution such as sum of voxels intensity distribution, kurtosis, skewness, mean, and standard deviation to create the regression model to determine threshold value. Another similar work using the threshold approach was presented by Wen and Sachdev [45]. They designed an automatic WML segmentation based on the mean and standard deviation ($\mu + \sigma$) of white matter intensity distribution to determine WML loads. For example, $\mu_{WM} + 6\sigma_{WM}$ was defined to threshold severe WML loads, while $\mu_{WM} + 3\sigma_{WM}$ was defined to threshold mild and moderate WML loads.

Souplet *et al.* [30] introduced the threshold-based technique for MS lesion segmentations in the Medical Image Computing and Computer Assisted Intervention (MICCAI) (Challenge, 2008). Maximum likelihood estimation was employed to classify ten classes (GM, WM, CSF, the outlier, and six GM/CSF partial volume classes) of brain tissues based on T1-w and T2-w sequence. The mask of GM was used to superimpose on FLAIR images, and the threshold value of $T = \mu_{WM} + 2\sigma_{WM}$ was computed to determine potential lesions. The final lesion was determined after post processing by reducing false positive of the lesion. Notably, this is the best approach compared to eight MS lesion segmentation approaches proposed in the challenge. Coincidentally, there was another threshold-based technique suggested by Bricq *et al.* [27] in the same challenge. In this work, brain tissues such as WM, GM and CSF were first classified using Hidden Markov Chain (HMC) approach with probabilistic atlas (constructed from 31 healthy brains) on FLAIR and T2-w images. Trimmed Likelihood Estimator was used to optimise the parameters of HMC model to detect the outliers. The detected outliers with probability value lower than the threshold value were defined as MS lesions as concluded by the authors. Another thresholding approach based on intensity histogram was presented by de Boer *et al.* [28]. The threshold value $T = \mu_{GM} +$

$\alpha\sigma_{GM}$ was suggested to segment a WML based on voxels intensity of grey matter distribution. GM, CSF and WM were segmented using kNN classifier and α adaptively optimised threshold parameter that was obtained from the experiment using six subjects with segmented WML manually by a neuro-radiologist.

A box-plot method with trim mean approach was introduced by Ong *et al.* [1] to determine outliers from normal brain tissues for each FLAIR image. Authors suggested that WML were detected and segmented adaptively based on box-plot analysis to define outliers and extreme outliers[70] using histogram generated from skull striped FLAIR images. The extreme outliers of box-plot were used to detect WML and outliers of box-plot which were used to threshold and segment WML. Outlier, $f_3 = Q_3 + 1.5 \times IQR$ and extreme outlier, $F_3 = Q_3 + 3.0 \times IQR$ where $IQR = Q_3 - Q_1$ was the Inter Quartile Range that denoted the range of values falling within the 25th percentile and 75th percentile of the intensity distribution (see Fig. 1). The final lesions were determined after false positive reduction using morphology operation. The method was robust to images that were acquired from multicenter and fast and efficient computation using boxplot analysis. The method validated with publicly available benchmark dataset (MICCAI challenge 2008) showed high accuracy.
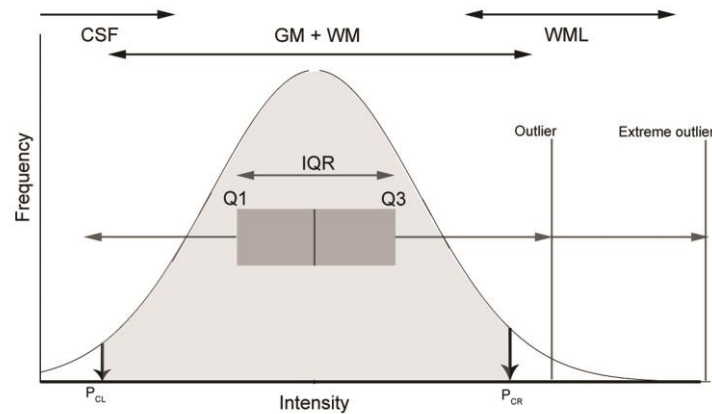

Fig.1 Graphical scheme of the box-plot method with trim mean approach introduced by Ong et al. [1]. The outliers and extreme outliers are determined using the Box and Whisker plot using the intensity distribution of the GM and WM voxels. Hence, the WML is detected based on extreme outlier value and then WML segmented based on adaptive thresholding using outlier value. Reproduced from Ong et al. [1].

A threshold value was extended based on Jack *et al*.[46] with Bayesian decision rule presented by Yoo *et al*.[44]. In their study, the *IG-peak*, peak of intensity FLAIR histogram and the standard deviation(σG) were obtained from the fitted Gaussian. The final lesions could be determined based on *Io = σ_GZo+IG-peak* where *Zo* was adaptive parameter Z-score. In another recent work, brain tissues were segmented using a modified expectation-maximisation algorithm, and a WML was then segmented using thresholding onFLAIR images[43]. The adaptive threshold value, *T = μGM + $\gamma\sigma GM$* where *μGM* and *σGM* were obtained GM distribution on FLAIR images and γ was empirical parameter used to define outliers. Subsequently, a set of rules were defined to remove false positive and determine WML as suggested by Cabezas *et al*.[43].

An MS lesion segmentation tool has been developed by Roura *et al*. [42]. The method was based on two main steps. The brain tissues were first classified into three main tissues (CSF, GM, WM) based on T1w images using expectation maximisation method. Voxel intensity of grey matter from FLAIR imageswas then extracted using the GM mask. Thus, the lesions with brighter intensity on FLAIR can be determined based on T = μ_{GM} + $\alpha\sigma_{FWHM-GM}$. While the μ_{GM} was the means of grey matter distribution, the $\sigma_{FWHM-GM}$ was the standard deviation defined from FWHM, and α was the parameter used to adjust the threshold level to segment the WML.

2.2.2   Supervised Learning Approaches

Supervised learning approaches are also known as machine learning approaches which require labelled information to assemble a mathematical model ("knowledge") from a set of features during the training phase. Subsequently, features obtained from each targeted image are used to perform lesion classification based on the mathematical model in the testing phase. There are two main strategies used in labelling process [38]. First, the labelled class can be collected manually based on neuro-radiologist delineation. Second, the labelled class can be done automatically based on brain atlas. The second strategy usually involves registration process to map the region between brain atlas and analysed brain image.

Feature is the first and critical component to be investigated before the supervised learning model is constructed. Feature extraction and selection is a challenging study and a crucial step to achieve accurate WML classification. The feature is a prominent attribute of each brain component on the images such as GM, WM, WML, etc. The characteristics of a good feature are listed below:
1. Consistent in the series of images of the same region.
2. Not sensitive to transformations on the images.
3. Not sensitive to image noise.
4. Easy and obvious to be found with bare eyes.

Based on the characteristics listed above, MRI provides an excellent image contrast to differentiate between brain tissues and abnormal tissues. Hence, the most common feature used in lesion segmentation and classification is the voxel intensity [26,29,31]. In many active WML segmentation research, researchers often consolidate several MRI sequences to extract the voxel intensity. The benefit is to distinguish the brain tissues and abnormal tissues. For instance, a white matter lesion appears as the brightest (hyperintensity) intensity voxel in FLAIR and T2-w sequence, but the darkest (hypointensity) in T1-w sequence. Whereas, the intensity of white matter tissues appears as the hyperintensity in T1-w images, and it presents hypointensity in T2-w images. Gray level intensity is a popular feature that can be extracted from a single sequence [29,3] or multi-sequences [49,35,31,52,51,48,50]as reported in the literature. Alternatively, Spatial information is another useful feature in white matter lesion segmentation [49,51,26]. Spatial feature is the information related to space in x-, y-, and z-coordinates of each voxel in the brain images. The voxels are computed using Euclidean distance to represent the spatial feature. For instance, the center of gravity and location of training point. Besides, the position and direction (degree) are defined as polar coordinates which have been applied in WML segmentation [26]. Normalised spatial coordinate is another method to obtain spatial information based on the standard atlas for instance, Montreal Neurological Institute, (MNI brain template) has been utilised in WML detection [52,2,47]. In the recent study, shape information is applied to identify WML[50,48].

Several supervised learning algorithms based on Atlas have been applied to lesion segmentation [52,34,29,31,2]. A well known k-Nearest Neighbor (kNN) was improved by Steenwijk *et al.* [2] to segment and classify the lesion. There were a total of eight important features used for kNN classification proposed by the author (see Fig. 2). Furthermore, the authors employed multi-atlas segmentation [71] to construct Tissue Type Priors (TTPs). The performance of kNN used to segment lesions increased after adding TTPs as concluded by authors. Another Atlas-based approach using decision forest classifier associated with Fisher linear discriminant analysis has been presented by Akselrod-Ballin *et al*. [55]. The method classified and segmented lesions using multi-sequences MR images. Hence, the targeted image was required to register them using Statistical Parametric Mapping (SPM[1]) software map into brain Atlas probability maps. Subsequently, a set of rich features such as neighborhood relations, location, anatomical context, intensity and shape extracted and employed in the decision forest classifier to classify lesions.
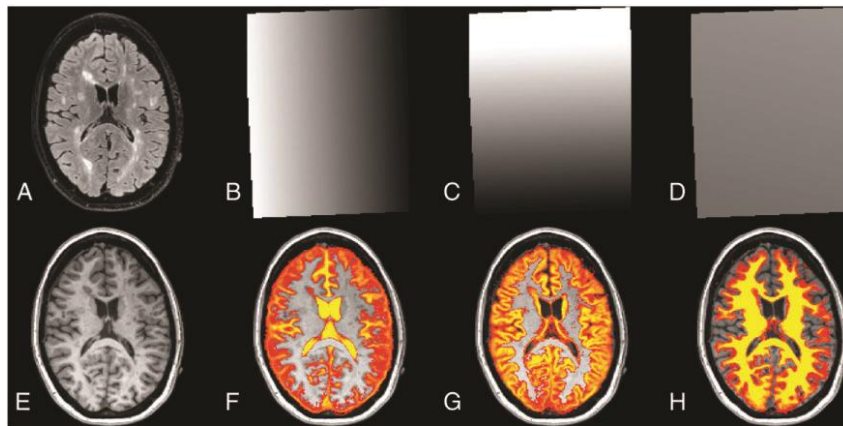
---

[1]http://www.fil.ion.ucl.ac.uk/spm/

Fig.2 Features used for the kNN classification as suggested by Steenwijk et al. [2], They include 3DFLAIR intensity (A),MNI-normalised spatial coordinate x (B), spatial coordinate y (C), spatial coordinate z (D), 3DT1 intensity (E), pCSF (F), pGM (G), and pWM (H). Reproduced from Steenwijk et al. [2].

In another supervised learning strategy, a classifier model was constructed based on features that were extracted from delineated image and labelled manually by a neuro-radiologist. The model was then used to classify the voxels of lesions or brain tissues to segment the lesions ultimately. Notably, this approach did not need registration algorithm to align and map the voxels to reference image template or Atlas. Based on the literature, supervised lesion segmentation uses prior knowledge annotated from human expert to increase accuracy and robustness of the algorithm, where the volume of the false positive lesion is reduced significantly. The classifier often employed in the segmentation includes kNN, Support Vector Machine (SVM), decision forest [53,72] and Neural Network classifier [61]. In a study reported by Anbeek *et al.* [26], two neuro-radiologists were employed to delineate white matter lesions and organise them into deep white matter lesions (DWML) and periventricular white matter lesions (PVWML). The final output delineated by the neuro-radiologists was further validated in a consensus meeting to avoid intra-observer variability. A set of features that consist of voxels intensity features and spatial features were generated. Then, kNN classifier was performed on analysed image to produce probability image map. Hence, by applying the different values of the threshold, the binary of WML could be obtained from the probability image map. In another work, the classifier applying decision forest was presented by Geremia *et al.*[53]. In their proposed method, a neuro-radiologist was required to perform the labelling on lesions and background images. Three sets of features namely intensity, prior and context-rich information were extracted from multi-sequence images (T1-w, T2-w, and FLAIR). Thus, the decision forest classifier model was constructed. In their experiment, decision forest classifier was able to select the most discriminative features to achieve the best white matter lesion segmentation result.

A comparison of five multi-sequence techniques on WML segmentation in normal ageing was presented by Vald´esHern´andez *et al*.[61]. These techniques included one unsupervised technique known as Minimum Variance Quantisation (MVQ) and four supervised techniques namely Back-Propagated Neural Networks, Nearest Neighbor, Gaussian classifier, and Parzen Windows. All presented techniques in their comparison study were used to segment WM, CSF and WML on brain MRI. Based on the comparison among supervised techniques, Parzen Windows was considered to have achieved better segmentation results. However, unsupervised MVQ method was found to achieve the best segmentation compared to other supervised methods by using the color fusion of two or more structural MRI sequences. Researchers concluded that the only drawback of the MVQ was it needed further methodological developments to automate and improve its reliability.

In a recent work, several deep convolution neural network (CNN) architectures have been proposed by Ghafoorian *et al*.[47]. Their network architectures consisted of a single scale model, multi-scale early fusion, multi-scale late fusion with independent weights and multi-scale late fusion with weight sharing. Also, Patches-based training was considered in the framework because it required less memory and was easy to optimise in the imbalanced classification problem. Thus, patches with multi-scale that extracted the spatial information (location) features were implemented during the training step. The eight features utilised in this study were the x, y, and z coordinates of the corresponding voxels in MNI atlas space, distances from the left ventricle, right ventricle, brain cortex and mid-sagittal brain surface and the prior probability of WMH occurring in that location. In their observation, CNN that was associated with location information was out performed by a random forest classifier using hand-crafted features, and CNN without associated anatomical location formation concluded in their study.

2.2.3   Unsupervised Learning Approaches

Unsupervised learning approach is an approach to label each voxel of MRI without an expert intervention. Based on the literature, probability and clustering are the two unsupervised learning approach often used. The probability approach is the method that computes the likelihood of each voxel belongs to desired classes. Whereas, the clustering approach is the method to group a set of the data point in the same group/cluster based on their similarity. Similarity can be defined using a distance measure such as Euclidean. Apparently, the approach does not require expert intervention for the class labelling. However, prior knowledge such as the number of clusters is required. The number of clusters is defined from the image content based on human understanding to construct the region clusters. Furthermore, an initial guess of mu and sigma will be required for probability approach such as expectation–maximisation (EM) algorithm.

In the past, expectation–maximisation (EM) algorithm was a popular method used to classify the brain tissues based on the statistical models. The stochastic approximation model using EM algorithm has been proposed by Van Leemput *et al*. [37]; the initial parameter used for the classification of WM, GM, and CSF which were first obtained from digital brain Atlas. Thus, the lesion voxels that were not described by the model could be estimated. A rule- based approach using two-level (reasoning and adaptive) segmentation technique was developed by Admiraal-Behloul *et al*. [68]. The reasoning level applied the fuzzy inference system where each voxel was assigned into linguistic values (BRIGHT, MEDIUMBRIGHT and DARK) by using fuzzy if-then rules. At adaptive level, the intensity value was transformed into linguistic information. The lesion can be detected at the adaptive level with fuzzy inference rules. For instance:

"If voxel_position is WM and t2_intensity is BRIGHT and
flair_intensity is BRIGHT, then segmented_voxel is WMH."

Another interesting work was presented by Bricq S. [73], in which the author kept the neighborhood information of voxel intensities using a Hidden Markov chain model. The prior information was obtained from a probabilistic brain Atlas. Consequently, the outlier (lesions) could be detected using the trimmed likelihood estimator. Prastawa and Gerig [74] extended the method proposed by Van Leemput *et al.* [37], where the method was based on an Atlas of healthy brain images, and lesions were subsequently detected as outliers. In their proposed method, a group of the segmented voxels of the potential lesions was further validated among neighboring regions using Kullback-Leibler divergence. Ultimately, the proposed method was performed without required delineation by a neuro-radiologist at specific brain regions. Besides, the method can be performed on the images acquired from the various scanner with scanning parameter since it did not require training process. Unified model using mixtures of Gaussians (MOG) and fuzzy clustering performed on T1-w images was presented by Seghier *et al.* [64]. Outlier voxels were detected in each tissue by comparing WM and GM segmented voxels under the unified model. Thus, outlier voxels in each tissue were defined as the lesions that were represented in fuzzy membership value in the interval of [0, 1]. The fuzzy membership value indicated the degree of abnormality of every single voxel. The authors concluded that this fuzzy membership was critical in generating lesion-deficit mappings. In another work which used a mixture clustering model was proposed by Khayati *et al.* [65]. Voxels of brain image of FLAIR images were clustered into four groups(GM, CSF, WM and "others") by using Adaptive Mixtures Model (AMM), Bayesian classifier and also MRF, where "others" were the outliers of the brain cluster model. Hence, a cluster of "others" were analysed as the final brain lesions. Similarly, three main classes of brain tissues were clustered into GM, WM and CSF using a robust Expectation-Maximisation (EM) algorithm presented by Garc´ıa-Lorenzo *et al.* [67]. These voxels of tissue clusters

were used to construct multidimensional feature space. The lesions were determined from outliers rejected from their modified EM algorithm based on Trimmed Likelihood (TL) Estimator and Mahalanobis distance. This work was further improved and investigated by Garc´ıa-Lorenzo *et al.* [63]. The GMM was applied on a T1-w image to extract the three brain tissues (CSF, GM and WM) presented by Schmidt *et al.* [62]. The brain tissues information was then co-registered to voxel intensities of FLAIR images to compute the lesion belief map. True lesions can be defined by a threshold (k) value. The optimal k value can be estimated from a set of a binary reference image delineated manually by a clinical expert. The highest dice coefficient representing the optimal threshold k value was compared with a lesion segmented with threshold k value and a binary reference image. Besides, Gaussian Mixture Model which was combined with context sensitive EM algorithm was proposed by Simões *et al.* [3]. The WML can be defined with threshold after the GMM is convergence (See Fig. 3). Notably, the approach uses a single sequence (3D FLAIR) as their input. The method shows a promising result in their WML segmentation study.

In another work, Expectation Maximisation algorithm with neighborhood consistency constraints was suggested by Sudre *et al.* [59]. The coherence between neighborhood voxel information could be improved with Markov Random Field(MRF). In their three-tier hierarchy framework. A total of four tissue types (GM, WM, CSF and Non-brain) were determined in the initial Gaussian Mixture Model. Subsequently, the number of cluster for each component was automatically determined through a split and merge strategy. Hence, the lesion could be decided from the outlier portion of the final model. A similar approach using GMM was presented by Jain *et al.* [60]. The outlier class (partial volume effects, artefacts and MS brain lesions) could be estimated from FLAIR images by using the brain tissue segmented region obtained from T1-w image as prior information. So, voxel intensities from FLAIR images that corresponded to the healthy brain tissues were modelled as the normal distribution. Subsequently, the voxel intensities of FLAIR which were not under the normal distribution were estimated as outlier belief map. The final lesions were then segmented from the outlier belief map based on the anatomical information.

Recently, an energy minimisation method for MS lesion segmentation was presented by Zhao *et al.* [58]. The method was an extension of MICO algorithm which was originally designed for healthy brain tissue segmentation that involved intensity inhomogeneity correction. Therefore, the method did not require training procedure and intensity inhomogeneity correction in their framework. A larger weight parameter is assigned to energy formulation on FLAIR images. Thus, the preliminary MS lesion could be detected. The accuracy of the method was further improved with the region growing algorithm; for instance, active contour models

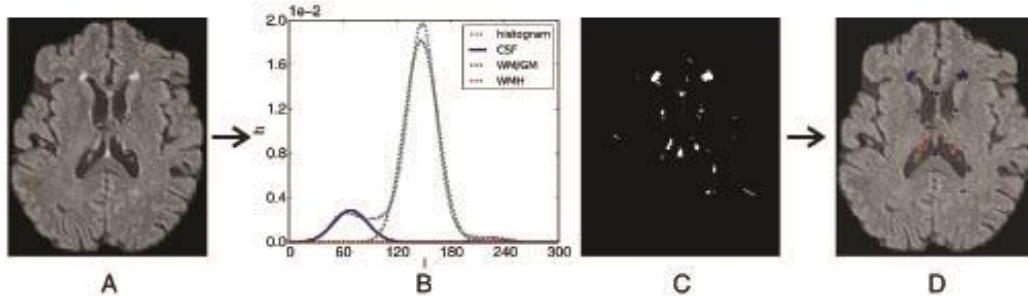or level set method as suggested by Zhao *et al.* [58].

Fig.3 An overview of the segmentation framework proposed by Simões *et al.* [3]: A) skull-stripped and bias field-corrected FLAIR image and B) fit a 3-class context-sensitive GMM based on (A). Subsequently, a threshold to the WMH class probability map, obtains C) an initial lesion segmentation. D) A post-processing step removes false positive in the (C); the removed false positive in read; the final WML segmentation in blue. Reproduced from Simões *et al.* [3].

2.3   False Positive Elimination

False positive (FP) is defined as voxels hyper-intensity region which seems like WML but it is not. The false positive consists of image artefact, voxel of incomplete skull stripping and image noise. Besides, the framework with miss-registration process may increase false positive (mainly on eyeballs). In the previous study, false positive elimination is included in the post-processing step to identify the final WML and remove the fake lesions. The overview of false positive reduction techniques is listed in Table 3.

Generally, the morphology operation is often applied to reduce the FP based on their sizes. The size of each study varies from 3-10 voxel intensities. Additionally, the anatomical information is found useful to define the true positive where the location of WML only occurs in white matter regions. Some of the studies utilise the classification method to reduce the FP occurrence in their study. The hyper-intense region like eyes and fat that incompletely removed by automatic skull stripping algorithm results in false positive as reported by Lao *et al.* [29]. These false positive can be eliminated in a few iterations using a technique that consists of morphological operations and adaptive thresholding in skull removed FLAIR images. On the other hand, false positive comprising of dura and skull showing hyper-intensity region which is required to be removed after WML segmentation was reported by de Boer *et al.* [28]. Each of the connected voxels labelled background or non-background was first identified with classified brain tissues and WML. In their rule-based method, false positive would be eliminated if the ratio of component labelled background to component labelled non-background was larger than 0.4 (chosen based on observation of a neuro-radiologist during the experiment), then components labelled 'non-background' were labelled as 'background'. In

another work presented by Ong *et al.* [1], the morphology operation and anatomical information on brain tissue were utilised to eliminate false positive. The brain tissue clusters such as WM, GM, and CSF were extracted from the T1-w image using FCM algorithm. The segmented WML that did not occur within the WM region would be removed as false positive since WML only appear in white matter region. Besides, false positive caused by the "shine through effect in periventricular" could be removed using CSF cluster (included periventricular region). The CSF cluster would be dilated with three voxels and superimposed on top of the segmented lesion to further reduce the false positive. Thus, the final WML was then identified.

The prior information refined lesions by using the rule-based method as reported by Cabezas *et al.* [43]. They set the minimum size of the lesion in their experiments with ten voxels. This parameter was meant to remove approximate lesion load with 30 mm or less to be excluded from their binary image. Another similar work by Zhao *et al.* [58] proposed that false positive WML could be removed according to the areas (size) of all connected components in the lesion binary mask. In conclusion, image artefact, incomplete skull stripping algorithm and incorrect performance of image registration in preprocessing step which often lead to false positive are generated. False positive elimination techniques vary from simple use of morphology operation to advance employed supervised learning algorithm depending on how false positive is presented. Hence, false positive elimination is an important post processing study because false positive result in huge influence to the WML segmentation accuracy. Hence, the automated WML segmentation employed for quantification of lesion load with less false positive rate is the preferred method to be used in a clinical practice.

Table 3: Summary of false positive reduction methods described in the review.

Approaches	False positive elimination method	Article
FLAIR histogram threshold approach	• Two level threshold to remove false positive voxels	Roura *et al.* [42]
	• Removing false positive based on their size & prior Knowledge of anatomical information	Cabezas *et al.* [43]
	• Removing false positive based on their size & prior Knowledge of anatomical information	Yoo *et al.* [44]
	• Removing false positive based on their size & prior Knowledge of anatomical information	Ong *et al.* [1]
	• Removing false positive based on their size & classification algorithm	De Boer *et al.* [75]
	• Prior Knowledge of anatomical information to exclude the false positive voxels	Souplet *et al.* [30]
	• Removing false positive based on their size	Wen and Sachdev [45]
	• Manually removing false positive voxels	Jack *et al.* [46]

Approaches	False positive elimination method	Article
Supervised learning approach	• Removing false positive based on their size	Valverde *et al.* [48]
	• Removing false positive based on spatial probability equation	Zhan *et al.* [49]
	• N/A	Ghafoorian *et al.* [47]
	• Removing false positive based on classification algorithm	Rincón *et al.* [50]
	• Prior Knowledge of anatomical information to exclude the false positive voxels	Griffanti *et al.* [51]
	• Prior Knowledge of anatomical information to exclude the false positive voxels	Roy *et al.* [52]
	• Removing false positive based on their size	Steenwijk *et al.* [2]
	• N/A	Geremia *et al.* [53]
	• Removing false positive based on classification algorithm	Yamamoto *et al.* [54]
	• Prior Knowledge of anatomical information to exclude the false positive voxels	Zacharaki *et al.* [31]
	• Removing false positive based on clustering algorithm	Akselrod-Ballin *et al.* [55]
	• N/A	Kroon *et al.* [56]
	• Removing false positive based on classification algorithm & prior Knowledge of anatomical information	Lao *et al.* [29]
	• N/A	Anbeek *et al.* [26]

Table 3: Summary of false positive reduction methods described in the review. (Continued)

Table 3: Summary of false positive reduction methods described in the review. (Continued)

Approaches	False positive elimination method	Article
Unsuperised learning approach	• Removing false positive based on their size	Zhao *et al.* [58]
	• Rules based method to remove false positive voxels	Sudre *et al.* [59]
	• Prior Knowledge of anatomical information to exclude the false positive voxels	Jain *et al.* [60]
	• Prior Knowledge of anatomical information to exclude the false positive voxels	Simões *et al.* [3]
	• Manually removing false positive voxels	Vald´esHern´andez *et al.* [61]
	• Prior Knowledge of anatomical information to exclude the false positive voxels	Schmidt *et al.*[62]
	• Rules based method to remove false positive voxels	Garc´ıa-Lorenzo *et al.* [63]
	• N/A	Seghier *et al.* [64]
	• Rules based method to remove false positive voxels	Khayati *et al.* [65]
	• Removing false positive based on their size	Bricq *et al.* [27]
	• Rules based method to remove false positive voxels	Prastawa and Gerig [74]
	• Rules based method to remove false positive voxels	Garc´ıa-Lorenzo *et al.* [67]
	• Prior Knowledge of anatomical information to exclude the false positive voxels	Admiraal-Behloul *et al.* [68]
	• Removing false positive based on their size and connectivity rules	Van Leemput *et al.* [37]

3   Discussions

The advantage of fully automated WML segmentation using supervised learning approach is that the classification model is constructed based on the feature input from an experienced neuro-radiologist. In short, WML is classified and segmented based on the "knowledge" from neuro-radiologists. Therefore, the WML segmentation is more meaningful compared to unsupervised learning and histogram-based thresholding approaches. However, there are three possible drawbacks that can be concluded for supervised learning. First, most of the supervised classification algorithms are computationally expensive to build the optimal classification model during the training phase. The second drawback is that the labelling of WML and brain tissues need to be performed carefully by an experienced neuro-radiologist during the training phase. Thus, the process is time-consuming and tedious. Lastly, images from different hospitals will need to consider re-training of the learning model depending on the selected features that are potentially influenced by acquisition protocol. Automatic labelling is possible for supervised learning approach because several studies have shown that standard brain Atlas is an alternative method used to construct classification model without clinician manual delineation. The only drawback applied on this method is that the registration algorithm needs to be performed accurately because the accuracy of registration will give a tremendous impact to WML segmentation accuracy. Specifically, miss-registration during the training and target image transformation into standard brain atlas often increase the false positive rate as reported in the literature.

Unsupervised learning approaches are another famous technique applied on WML segmentation. The advantage of unsupervised learning approaches is that they do not require human intervention and the image data set acquired from multi-institute is easy to adapt. The classification either applies Gaussian mixture model or grouping of the adjacent features to distinguish the brain lesions and various brain tissues. Therefore, they do not require a training process to construct the learning model. Hence, the approach is able to classify the image content adaptively without the need of the learning model. In our opinion, the histogram-based thresholding approaches are widely applied to WML segmentation because the method does not require training phase and it is easily implemented. Furthermore, the adaptive threshold value can be determined based on intensity distribution using simple statistical analysis which have been shown in many previous studies. Hence, this approach is robust and computation is efficient to be applied on MR image acquired from the multi-centre. The drawback of thresholding approaches is that the threshold value is determined based on intensity distribution. Normally the right tail of intensity distribution will be retained. Thus, these segmented WML are considered as the outliers of healthy tissues. It is a subject of many debates because the segmented WML might include voxels of incomplete skull stripping process

and flow artefact such as "shine through effect". Hence, false positive of WML will be increased indirectly.

Based on the literature, the accuracy of classification and segmentation for age-related or vascular WML is not promising as compared to multiple sclerosis which can be observed in the literature [39]. There are two possible reasons; First, the degree of small vessels ischemia varies and is presented in intensity and the boundaries to determine WML is fuzzier. Second, the white matter may present normal (healthy) on FLAIR image, but the actual region has become damaged and presented in low contrast area which is known as abnormalities in the normal appearing white matter.

4   Conclusion

In this paper, three main approaches (FLAIR histogram threshold, supervised learning, and unsupervised learning approaches) from the past to the latest of fully automated WML segmentation on MRI have been reviewed. Their segmentation accuracy and false positive reduction methods have been summarised in Tables 2 and 3 respectively. The main advantage of WML assessment based on the visual scoring method and manual segmentation is its easy implementation. However, human intervention is required to evaluate image data set with slice by slice basis. Hence, it is labour intensive, tedious and time-consuming. Furthermore, the output is inconsistent and shows high variation intra and inter-reader agreement. Apparently, quantitative WML load based on automated WML segmentation is superior to visual scoring and manual segmentation, because the results are fast, precise, consistent and comparable in the assessment of WML load.

Overall, the approaches discussed are still suffering from a major research question addressed by Yamauchi *et al.*[9]. These approaches still generate false positive that are caused by image artefact and tiny digital noise(likely to be lesion). These false positive are potentially tiny lesions that are miss-identified by human observers during the ground truth preparation. Therefore, more research study is needed to investigate on false positive detection [47]. Usually, most of the authors intend to segment the WML accurately but have less intention to remove the false positive. These approaches are not effective to reduce the false positive caused by image artefact and tiny noise. A simple morphology operation methods such as dilate, erosion, closing and opening have been applied and tended to remove false positive in post processing step [26-27,43,28,67,46,65,29,1,64,30,45,44,31]. Also, some of them remove the false positive based on the size of hyper-intensity region which is not more than 3 mm. In this context, no advance method has been proposed in the literature. It is an open problem which is gaining more attention.

References

[1]. 1. Ong KH, Ramachandram D, Mandava R, Shuaib IL (2012) Automatic white matter lesion segmentation using an adaptive outlier detection method. Magnetic Resonance Imaging 30 (6):807-823

[2]. 2. Steenwijk MD, Pouwels PJW, Daams M, van Dalen JW, Caan MWA, Richard E, Barkhof F, Vrenken H (2013) Accurate white matter lesion segmentation by k nearest neighbor classification with tissue type priors (kNN-TTPs). NeuroImage: Clinical 3 (0):462-469. doi:http://dx.doi.org/10.1016/j.nicl.2013.10.003

[3]. 3. Simões R, Mönninghoff C, Dlugaj M, Weimar C, Wanke I, van Cappellen van Walsum A-M, Slump C (2013) Automatic segmentation of cerebral white matter hyperintensities using only 3D FLAIR images. Magnetic Resonance Imaging 31 (7):1182-1189. doi:http://dx.doi.org/10.1016/j.mri.2012.12.004

[4]. 4. Peters N, Dichgans M (2010) Vascular dementia. Vaskulre Demenz 81 (10):1245-1255

[5]. 5. Zhuang AH, Valentino DJ, Toga AW (2006) Skull-stripping magnetic resonance brain images using a model-based level set. NeuroImage 32 (1):79-92

[6]. 6. Cavalieri M, Enzinger C, Petrovic K, Pluta-Fuerst A, Homayoon N, Schmidt H, Fazekas F, Schmidt R (2010) Vascular dementia and Alzheimer's disease - Are we in a dead-end road? Neurodegenerative Diseases 7 (1-3):122-126

[7]. 7. Debette S, Markus HS (2010) The clinical importance of white matter hyperintensities on brain magnetic resonance imaging: systematic review and meta-analysis. BMJ 341. doi:10.1136/bmj.c3666

[8]. 8. Diniz BS, Butters MA, Albert SM, Dew MA, Reynolds Iii CF (2013) Late-life depression and risk of vascular dementia and Alzheimer's disease: Systematic review and meta-analysis of community-based cohort studies. British Journal of Psychiatry 202 (5):329-335

[9]. 9. Yamauchi H, Fukuda H, Oyanagi C (2002) Significance of white matter high intensity lesions as a predictor of stroke from arteriolosclerosis. Journal of Neurology, Neurosurgery & Psychiatry 72 (5):576-582. doi:10.1136/jnnp.72.5.576

[10]. 10. Park D-C Intuitive Fuzzy C-Means Algorithm for MRI Segmentation. In: Information Science and Applications (ICISA), 2010 International Conference on, 2010. pp 1-7

[11]. 11. Gibou F, Fedkiw R (2005) A fast hybrid k-means level set algorithm for segmentation. 4th Annual Hawaii International Conference on Statistics and Mathematics:281-291

[12]. 12. Herman GT, Carvalho BM (2001) Multiseeded segmentation using fuzzy connectedness. IEEE Transactions on Pattern Analysis and Machine Intelligence 23 (5):460-474

[13]. 13. Liu Z, Lin J, Zou Y, Chen K, Yin G Automatic 3D segmentation of MRI brain images based on fuzzy connectedness. In: 2nd International Conference on Bioinformatics and Biomedical Engineering, iCBBE 2008, Shanghai, 2008. 2nd International Conference on Bioinformatics and Biomedical Engineering, iCBBE 2008. pp 2561-2564

[14]. 14. Udupa JK, Saha PK, Lotufo RA (2002) Relative fuzzy connectedness and object definition: Theory, algorithms, and applications in image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (11):1485-1500

[15]. 15. Smith SM (2002) Fast robust automated brain extraction. Human Brain Mapping 17 (3):143-155

[16]. 16. Dogdas B, Shattuck DW, Leahy RM (2005) Segmentation of skull and scalp in 3-D human MRI using mathematical morphology. Human Brain Mapping 26 (4):273-285

[17]. 17. Balan AGR, Traina AJM, Ribeiro MX, Marques PMA, Traina Jr C (2012) Smart histogram analysis applied to the skull-stripping problem in T1-weighted MRI. Computers in Biology and Medicine 42 (5):509-522. doi:http://dx.doi.org/10.1016/j.compbiomed.2012.01.004

[18]. 18. Sied JG, Zijdenbos AP, Evans AC (1998) A nonparametric method for automatic correction of intensity nonuniformity in mri data. IEEE Transactions on Medical Imaging 17 (1):87-97

[19]. 19. Van Leemput K, Maes F, Vandermeulen D, Suetens P (1999) Automated model-based bias field correction of MR images of the brain. IEEE Transactions on Medical Imaging 18 (10):897-908

[20]. 20. Ge Y, Udupa JK, Nyul LG, Wei L, Grossman RI (2000) Numerical tissue characterization in MS via standardization of the MR image intensity scale. Journal of Magnetic Resonance Imaging 12 (5):715-721

[21]. 21. Loizou CP, Pantziaris M, Seimenis I, Pattichis C Brain MR image normalization in texture analysis of multiple sclerosis. In: Information Technology and Applications in Biomedicine, 2009. ITAB 2009. 9th International Conference on, 2009. IEEE, pp 1-5

[22]. 22. Nyú LG, Udupa JK (1999) On standardizing the MR image intensity scale. Magnetic Resonance in Medicine 42 (6):1072-1081

[23]. 23. Roy S, Carass A, Prince JL Patch based intensity normalization of brain MR images. In: Biomedical Imaging (ISBI), 2013 IEEE 10th International Symposium on, 2013. IEEE, pp 342-345

[24]. 24. Nyul LG, Udupa JK, Zhang X (2000) New variants of a method of MRI scale standardization. IEEE Transactions on Medical Imaging 19 (2):143-150

[25]. 25. Weisenfeld NL, Warfteld S Normalization of joint image-intensity statistics in MRI using the Kullback-Leibler divergence. In: Biomedical Imaging: Nano to Macro, 2004. IEEE International Symposium on, 2004. IEEE, pp 101-104

[26]. 26. Anbeek P, Vincken KL, Van Osch MJP, Bisschops RHC, Van Der Grond J (2004) Probabilistic segmentation of white matter lesions in MR imaging. NeuroImage 21 (3):1037-1044

[27]. 27. Bricq S, Collet C, Armspach J MS Lesion Segmentation based on Hidden Markov Chains. In, 2008.

[28]. 28. De Boer R, Vrooman HA, van der Lijn F, Vernooij MW, Ikram MA, van der Lugt A, Breteler MMB, Niessen WJ (2009) White matter lesion extension to automatic brain tissue segmentation on MRI. NeuroImage 45 (4):1151-1161

[29]. 29. Lao Z, Shen D, Jawad A, Karacali B, Liu D, Melhem ER, Bryan RN, Davatzikos C Automated segmentation of white matter lesions in 3D brain MR images, using multivariate pattern classification. In: 2006 3rd IEEE International Symposium on Biomedical Imaging: From Nano to Macro - Proceedings, 2006. pp 307-310

[30]. 30. Souplet J, Lebrun C, Ayache N, Malandain G An Automatic
       Segmentation of T2-FLAIR Multiple Sclerosis Lesions. In, 2008.

[31]. 31. Zacharaki EI, Kanterakis S, Bryan RN, Davatzikos C (2008) Measuring
       brain lesion progression with a supervised tissue classification system.
       Medical image computing and computer-assisted intervention : MICCAI
       International Conference on Medical Image Computing and Computer-
       Assisted Intervention 11 (Pt 1):620-627

[32]. 32. Bergeest J-P, Jäger F (2008) A comparison of five methods for signal
       intensity standardization in MRI. In: Bildverarbeitung für die Medizin 2008.
       Springer, pp 36-40

[33]. 33. Shah M, Xiao Y, Subbanna N, Francis S, Arnold DL, Collins DL, Arbel
       T (2011) Evaluating intensity normalization on MRIs of human brain with
       multiple sclerosis. Medical Image Analysis 15 (2):267-282

[34]. 34. Kroon D. VOE, Slump K (2008) Multiple Sclerosis Detection in
       Multispectral Magnetic Resonance Images with Principal Components
       Analysis. MIDAS Journal

[35]. 35. Maillard P, Delcroix N, Crivello F, Dufouil C, Gicquel S, Joliot M,
       Tzourio-Mazoyer N, Alpe¨¨??rovitch A, Tzourio C, Mazoyer B (2008) An
       automated procedure for the assessment of white matter hyperintensities by
       multispectral (T1, T2, PD) MRI and an evaluation of its between-centre
       reproducibility based on two large community databases. Neuroradiology
       50 (1):31-42

[36]. 36. Wu M, Rosano C, Butters M, Whyte E, Nable M, Crooks R, Meltzer
       CC, Reynolds Iii CF, Aizenstein HJ (2006) A fully automated method for
       quantifying and localizing white matter hyperintensities on MR images.
       Psychiatry Research: Neuroimaging 148 (2-3):133-142

[37]. 37. Leemput KV, Maes F, Vandermeulen D, Colchester A, Suetens P (2001)
       Automated segmentation of multiple sclerosis lesions by model outlier
       detection. IEEE Transactions on Medical Imaging 20 (8):677-688.
       doi:10.1109/42.938237

[38]. 38. Lladó X, Oliver A, Cabezas M, Freixenet J, Vilanova JC, Quiles A, Valls
       L, Ramió-Torrentà L, Rovira À (2012) Segmentation of multiple sclerosis
       lesions in brain MRI: A review of automated approaches. Information
       Sciences 186 (1):164-185. doi:http://dx.doi.org/10.1016/j.ins.2011.10.011

[39]. 39. Caligiuri ME, Perrotta P, Augimeri A, Rocca F, Quattrone A, Cherubini
       A (2015) Automatic Detection of White Matter Hyperintensities in Healthy
       Aging and Pathology Using Magnetic Resonance Imaging: A Review.
       Neuroinformatics 13 (3):261-276. doi:10.1007/s12021-015-9260-y

[40]. 40. García-Lorenzo D, Francis S, Narayanan S, Arnold DL, Collins DL
       (2013) Review of automatic segmentation methods of multiple sclerosis

white matter lesions on conventional magnetic resonance imaging. Medical Image Analysis 17 (1):1-18. doi:http://dx.doi.org/10.1016/j.media.2012.09.004

[41]. 41. Goceri E, Dura E, Gunay M Review on Machine Learning Based Lesion Segmentation Methods from Brain MR Images. In: 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA), 18-20 Dec. 2016 2016. pp 582-587. doi:10.1109/icmla.2016.0102

[42]. 42. Roura E, Oliver A, Cabezas M, Valverde S, Pareto D, Vilanova JC, Ramió-Torrentà L, Rovira À, Lladó X (2015) A toolbox for multiple sclerosis lesion segmentation. Neuroradiology 57 (10):1031-1043. doi:10.1007/s00234-015-1552-2

[43]. 43. Cabezas M, Oliver A, Roura E, Freixenet J, Vilanova JC, Ramió-Torrentà L, Rovira À, Lladó X (2014) Automatic multiple sclerosis lesion detection in brain MRI by FLAIR thresholding. Computer Methods and Programs in Biomedicine 115 (3):147-161. doi:http://dx.doi.org/10.1016/j.cmpb.2014.04.006

[44]. 44. Yoo BI, Lee JJ, Han JW, Oh SYW, Lee EY, MacFall JR, Payne ME, Kim TH, Kim JH, Kim KW (2014) Application of variable threshold intensity to segmentation for white matter hyperintensities in fluid attenuated inversion recovery magnetic resonance images. Neuroradiology 56 (4):265-281. doi:10.1007/s00234-014-1322-6

[45]. 45. Wen W, Sachdev P (2004) The topography of white matter hyperintensities on brain MRI in healthy 60- to 64-year-old individuals. NeuroImage 22 (1):144-154

[46]. 46. Jack CR, O'Brien PC, Rettman DW, Shiung MM, Xu Y, Muthupillai R, Manduca A, Avula R, Erickson BJ (2001) FLAIR histogram segmentation for measurement of leukoaraiosis volume. Journal of Magnetic Resonance Imaging 14 (6):668-676

[47]. 47. Ghafoorian M, Karssemeijer N, Heskes T, van Uden IWM, Sanchez CI, Litjens G, de Leeuw F-E, van Ginneken B, Marchiori E, Platel B (2017) Location Sensitive Deep Convolutional Neural Networks for Segmentation of White Matter Hyperintensities. Scientific Reports 7 (1):5110. doi:10.1038/s41598-017-05300-5

[48]. 48. Valverde S, Cabezas M, Roura E, González-Villà S, Pareto D, Vilanova JC, Ramió-Torrentà L, Rovira À, Oliver A, Lladó X (2017) Improving automated multiple sclerosis lesion segmentation with a cascaded 3D

convolutional neural network approach. NeuroImage 155:159-168. doi:http://dx.doi.org/10.1016/j.neuroimage.2017.04.034

[49]. 49. Zhan T, Yu R, Zheng Y, Zhan Y, Xiao L, Wei Z (2017) Multimodal spatial-based segmentation framework for white matter lesions in multi-sequence magnetic resonance images. Biomedical Signal Processing and Control 31:52-62. doi:http://dx.doi.org/10.1016/j.bspc.2016.06.016

[50]. 50. Rincón M, Díaz-López E, Selnes P, Vegge K, Altmann M, Fladby T, Bjørnerud A (2017) Improved Automatic Segmentation of White Matter Hyperintensities in MRI Based on Multilevel Lesion Features. Neuroinformatics. doi:10.1007/s12021-017-9328-y

[51]. 51. Griffanti L, Zamboni G, Khan A, Li L, Bonifacio G, Sundaresan V, Schulz UG, Kuker W, Battaglini M, Rothwell PM, Jenkinson M (2016) BIANCA (Brain Intensity AbNormality Classification Algorithm): A new tool for automated segmentation of white matter hyperintensities. NeuroImage   141:191-205. doi:http://dx.doi.org/10.1016/j.neuroimage.2016.07.018

[52]. 52. Roy PK, Bhuiyan A, Janke A, Desmond PM, Wong TY, Abhayaratna WP, Storey E, Ramamohanarao K (2015) Automatic white matter lesion segmentation using contrast enhanced FLAIR intensity and Markov Random Field. Computerized Medical Imaging and Graphics 45:102-111. doi:http://dx.doi.org/10.1016/j.compmedimag.2015.08.005

[53]. 53. Geremia E, Clatz O, Menze BH, Konukoglu E, Criminisi A, Ayache N (2011) Spatial decision forests for MS lesion segmentation in multi-channel magnetic resonance images. NeuroImage 57 (2):378-390

[54]. 54. Yamamoto D, Arimura H, Kakeda S, Magome T, Yamashita Y, Toyofuku F, Ohki M, Higashida Y, Korogi Y (2010) Computer-aided detection of multiple sclerosis lesions in brain magnetic resonance images: False positive reduction scheme consisted of rule-based, level set method, and support vector machine. Computerized Medical Imaging and Graphics 34 (5):404-413. doi:http://dx.doi.org/10.1016/j.compmedimag.2010.02.001

[55]. 55. Akselrod-Ballin A, *, Galun M, Gomori JM, Filippi M, Valsasina P, Basri R, Brandt A (2009) Automatic Segmentation and Classification of Multiple Sclerosis in Multichannel MRI. IEEE Transactions on Biomedical Engineering 56 (10):2461-2469. doi:10.1109/tbme.2008.926671

[56]. 56. Kroon D. VOE, Slump K (2008) Multiple Sclerosis Detection in Multispectral Magnetic Resonance Images with Principal Components Analysis. MIDAS Journal

[57]. 57. Anbeek P, Vincken KL, Viergever MA Automated MS-Lesion Segmentation by K-Nearest Neighbor Classification. In, 2008.

[58]. 58. Zhao Y, Guo S, Luo M, Liu Y, Bilello M, Li C (2017) An energy minimization method for MS lesion segmentation from T1-w and FLAIR images. Magnetic Resonance Imaging 39:1-6. doi:http://dx.doi.org/10.1016/j.mri.2016.04.003

[59]. 59. Sudre CH, Cardoso MJ, Bouvy WH, Biessels GJ, Barnes J, Ourselin S (2015) Bayesian model selection for pathological neuroimaging data applied to white matter lesion segmentation. IEEE Transactions on Medical Imaging 34 (10):2079-2102

[60]. 60. Jain S, Sima DM, Ribbens A, Cambron M, Maertens A, Van Hecke W, De Mey J, Barkhof F, Steenwijk MD, Daams M, Maes F, Van Huffel S, Vrenken H, Smeets D (2015) Automatic segmentation and volumetry of multiple sclerosis brain lesions from MR images. NeuroImage: Clinical 8:367-375. doi:http://dx.doi.org/10.1016/j.nicl.2015.05.003

[61]. 61. Valdés Hernández MdC, Gallacher PJ, Bastin ME, Royle NA, Maniega SM, Deary IJ, Wardlaw JM (2012) Automatic segmentation of brain white matter and white matter lesions in normal aging: comparison of five multispectral techniques. Magnetic Resonance Imaging 30 (2):222-229. doi:10.1016/j.mri.2011.09.016

[62]. 62. Schmidt P, Gaser C, Arsic M, Buck D, Förschler A, Berthele A, Hoshi M, Ilg R, Schmid VJ, Zimmer C, Hemmer B, Mühlau M (2012) An automated tool for detection of FLAIR-hyperintense white-matter lesions in Multiple Sclerosis. NeuroImage 59 (4):3774-3783. doi:http://dx.doi.org/10.1016/j.neuroimage.2011.11.032

[63]. 63. Garcia-Lorenzo D, Prima S, Arnold DL, Collins DL, Barillot C (2011) Trimmed-Likelihood Estimation for Focal Lesions and Tissue Segmentation in Multisequence MRI for Multiple Sclerosis. IEEE Transactions on Medical Imaging 30 (8):1455-1467. doi:10.1109/tmi.2011.2114671

[64]. 64. Seghier ML, Ramlackhansingh A, Crinion J, Leff AP, Price CJ (2008) Lesion identification using unified segmentation-normalisation models and fuzzy clustering. NeuroImage 41 (4):1253-1266

[65]. 65. Khayati R, Vafadust M, Towhidkhah F, Nabavi M (2008) Fully automatic segmentation of multiple sclerosis lesions in brain MR FLAIR images using adaptive mixtures method and markov random field model. Computers in Biology and Medicine 38 (3):379-390. doi:http://dx.doi.org/10.1016/j.compbiomed.2007.12.005

[66]. 66. Prastawa M. GG (2008) Automatic MS Lesion Segmentation by Outlier Detection and Information Theoretic Region Partitioning. MIDAS Journal

[67]. 67. Garcia-Lorenzo D, Prima S, Morrissey SP, Barillot C A robust Expectation-Maximization algorithm for Multiple Sclerosis lesion segmentation. In, 2008.

[68]. 68. Admiraal-Behloul F, Van Den Heuvel DMJ, Olofsen H, Van Osch MJP, Van Der Grond J, Van Buchem MA, Reiber JHC (2005) Fully automatic segmentation of white matter hyperintensities in MR images of the elderly. NeuroImage 28 (3):607-617

[69]. 69. Hirono N, Kitagaki H, Kazui H, Hashimoto M, Mori E (2000) Impact of white matter changes on clinical manifestation of Alzheimer's disease: A quantitative study. Stroke 31 (9):2182-2188

[70]. 70. Tukey JW (1977) Exploratory data analysis.

[71]. 71. Aljabar P, Heckemann RA, Hammers A, Hajnal JV, Rueckert D (2009) Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy. NeuroImage 46 (3):726-738. doi:http://dx.doi.org/10.1016/j.neuroimage.2009.02.018

[72]. 72. Mitra J, Bourgeat P, Fripp J, Ghose S, Rose S, Salvado O, Connelly A, Campbell B, Palmer S, Sharma G, Christensen S, Carey L (2014) Lesion segmentation from multimodal MRI using random forest following ischemic stroke. NeuroImage 98:324-335. doi:http://dx.doi.org/10.1016/j.neuroimage.2014.04.056

[73]. 73. Bricq S. CC, Armspach J. (2008) MS Lesion Segmentation based on Hidden Markov Chains. MIDAS Journal

[74]. 74. Prastawa M, Gerig G Automatic MS Lesion Segmentation by Outlier Detection and Information Theoretic Region Partitioning. In, 2008.

[75]. 75. De Boer R, Van Derlijn F, Vrooman HA, Vernooij MW, Ikram MA, Breteler MMB, Niessen WJ Automatic segmentation of brain tissue and white matter lesions in MRI. In: 2007 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro - Proceedings, 2007. pp 652-655