Int. J. Advance Soft Compu. Appl, Vol. 10, No. 3, November 2018
ISSN 2074-8523

White Matter Lesion Intensity Standardization using Adaptive Landmark based Brain Tissue Analysis on FLAIR MR Image

K. H. Ong[1,2], S. M. Shamsuddin[1,2], S. Z. Mohd Hashim[1,2], S. Sulaiman[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; sitizaiton@utm.my, sarina@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

Image intensity values which are extracted from magnetic resonance imaging (MRI) are not standardised and do not have tissue-specific interpretation due to the limitation of MRI instrumentation. The limitation poses many difficulties on data visualisation and texture feature analysis. Intensity and texture features extracted from MRI are not comparable for each inter-scan and intra-scan. Hence, they are not appropriate to be applied in supervised learning approaches to analyse the texture of white matter lesions. Consequently, this drawback often requires a standardisation method prior to further image analysis, which remains a common problem. In this study, a new automated method for image intensity standardisation is proposed to provide a standard intensity scale. In the proposed method, the landmarks in the intensity scale are automatically detected in the brain tissue intensity distribution using an adaptive outlier detection approach. Subsequently, landmarks are used to transform the brain tissues and lesion intensity into a standard scale by using the proposed transformation method. The method is validated using the cranial

MRIs (FLAIR sequence) that contain the white matter lesions from 10 subjects during their 3-year follow-up study. A paired t-test: t(29) = 2.045 and P(29)=1.42x10^{-15} where P<0.0001 confirms the significant difference in the before and after intensity range. In addition, intensity and texture features between the output images from the proposed approach and a leading intensity standardisation algorithm are further compared using the coefficient of variation, Pearson's correlation coefficient, and Kullback-Leibler divergence. Finally, qualitative evaluation of the MRI intensity is presented using the fixed-window-level method.

Keywords: *Intensity normalisation, intensity standardization, MRI, outlier detections, white matter lesions.*

1 Introduction

Consequent to a revolution in diagnostic imaging modality, magnetic resonance imaging has become an advanced intervention that offers a non-invasive imaging technology enabling visualisation inside the human body without the use of ionising radiation. The magnetic resonance (MR) modality provides high contrast images to allow radiologists to identify abnormality by visualising the soft-tissue intensity differences. Thus, MRI is widely used to diagnose and study the progression of brain diseases that are caused by white matter lesions (WML) such as multiple sclerosis (MS). However, quantitative WML analysis based on a classification model is challenging since MRI does not provide Hounsfield Units (HU) as offered in a computed tomography (CT) scan. Moreover, results of the quantitative WML analysis are easily influenced by the MRI acquisition conditions such as MR protocol, different brands of MRI scanners, and MRI parameters. Consequently, image intensity standardisation is an essential step to visualise the desired tissue uniformity and improve the tissue-specific meaning. Furthermore, intensity standardisation can also enhance the quality of brain lesion segmentation by using a supervised learning algorithm. Thus, it may improve the accuracy of the results in quantification analysis [1].

2 Related Work

There are several intensity standardisation methods that have been reported in the literature [1-8]. An automatic brightness and contrast adjustment approach has been proposed by Wendt [9] to visualise MR images uniformly. The simplistic standardisation approach using minimum and maximum pixel intensities for each image is first determined. These values are then mapped onto grey scale linearly in the 8-bit display. However, this method does not standardise the MR intensity for tissue-specific meaning. Nyu and Udupa [2] extended the method to standardise MRI intensity by using a windowing transformation approach. A set of images was used as input to learn and define the parameters. Mode of the histogram, minimum and maximum percentile intensities, and shoulder of the

"background hump" were the parameters suggested by Nyu and Udupa [2]. The parameters were used as landmarks to determine an intensity-standardised scale histogram. The actual intensities from the original histogram that were generated from input volumes were linearly transformed into the intensity standardised scale histogram. Ge *et al.* [1] applied the intensity standardisation method with new variant parameters such as median and percentiles that were introduced by Nyul *et al.* [3] to determine the standard histogram. In their experiments, the characteristic of healthy white matter tissue and abnormal tissue in Multiple Sclerosis (MS) patients can be distinguished accurately. Furthermore, an intensity standardisation method based on a multiplicative correction field was presented by Weisenfeld and Warfteld [4]. The method transformed MR images to match standard scale images with a minimising Kullback-Leibler divergence. Hence, the quality of MR brain lesions segmentation by using the classification approach could be further improved. To the best of our knowledge, the effectiveness of intensity standardisation on pathological images has been discussed very rarely in the literature. Most of the methods suggested above often did not include abnormalities such as white matter lesions. Researchers tend to remove the lesions on the brain images when evaluating the proposed intensity standardisation algorithm. An interesting patch-based intensity standardisation had been introduced by Roy *et al.* [8], whereby, a set of three-dimensional (3D) patches were stacked into one-dimensional (1D) vectors. These best-matching vectors between patch subject and patch atlas were determined by maximum likelihood and an expectation-maximisation (EM) algorithm. An intensity-standardised image was generated by replacing the centre of the pixel of each best-match subject patch with the atlas patch.

In recent years, image intensity standardisation on brain lesion images has gained more attention than healthy brain images. This is understandable since intensity standardisation allows for many other supervised learning algorithms to be applied to identify brain abnormalities such as white matter lesions and brain tumours. Jäger *et al.* [10] suggested a set of grey values to search and map between a set of probability density functions (PDF) which were generated from MR images and a set PDF generated from reference MR images. The mapping between the joint PDF of both sets of images can be approximated by the minimisation of the distance. Therefore, the benefit of this image intensity standardisation method is independent of the application, region of interest, acquisition protocol, and modality. In another work, MS lesion quantitative analysis based on texture features that were computed based on MR intensity pixels often suffered from inaccurate results due to non-standardisation. An appropriate intensity standardisation method was proposed by Loizou *et al.* [5,11] among six different intensity standardisation methods. These intensity standardisation methods were evaluated based on a texture feature that was extracted from the original and standardised images. Their evaluation method included Wilcoxon rank sum test and Kullback-Leibler Divergence. Thus, the

results of MS lesion classification were significantly improved after intensity standardisation was applied and they were less dependent on the MR acquisition protocol in their findings. In addition, a novel pathology robust intensity standardisation algorithm was presented by Ekin [7]. In their study, standardisation by global constraints (e.g., histogram similarity) and local constraints (e.g., voxel intensity) that improved transformation functions between the input and the reference were proposed. Subsequently, the final transfer function was computed as a weighted combination of the two local and global transfer functions. Thus, the method was successfully applied on pathology MR images.

A thorough review of image intensity standardisation applied to MRI with MS lesion load was first reported by Shah *et al.* [6]. They are investigating the effect of standardisation approaches when the methods are applied on MR brain lesions. Parametric supervised classifications such as a standard Bayesian classifier, an outlier-detection based approach, and a Bayesian classifier with Markov Random Field (MRF) were used to classify brain tissue and MS lesions before and after an intensity standardisation process. In their comprehensive evaluation, they concluded that image intensity standardisation was a significant step towards providing an improved discriminating ability for supervised learning algorithms.

From the existing literature, it is found that the lack of intensity standardisation on MRI still leads to the re-training of many supervised learning algorithms for every new protocol setting on MRI. Obviously, this is a tedious and time-consuming process for clinicians to repeat the algorithm training process every time. Hence, many of discussed supervised learning algorithms are not appropriate give efficient solution to the intensity standardisation issues on MRI. Besides, the lack of intensity standardisation also caused a difficulty for neuro-radiologist in performing a comparable brain lesion assessment study because the intensity contrast of a lesions are vary from time to time when imaging acquired. Therefore, in this study, an improved intensity standardisation method is introduced based on enhanced landmark-based approach with automated outlier detection method. The landmark-based approach has been used in many previous studies [17-20] and its performance and accuracy were evaluated by Bergeest and Jäger [21]. In addition, Shah *et al.* [6] had thoroughly validated its exceptional performances on MRI images to detect MS brain lesions which contributed by its fast computation and ability to reduce the complexity. The integration of the landmark-based approach and automated outlier detection method are able to standardise the brain tissue as well as lesion on MR images. Unlike the existing approaches that had been discussed, only brain tissue is normalised but not the lesion on MR images. In term of WML visual assessment [22-24], the proposed method aims to standardise image intensity contrasts for a comparable brain lesion assessments. This paper consists of Section 3, description of details of MR images and the proposed method used in the study; Section 4, presentation of results and the evaluation

methods; and finally, in Sections 5 and 6, discussion of the findings and conclusion respectively.

3 Materials and Methods

The proposed new image intensity standardisation method is validated using the dataset MRI sequences obtained from the clinical study of the protective effects of palm vitamin E tocotrienols on brain white matter [13]. Pathological cranial MR images with 10 subjects comprising T1-weighted (T1-W) and Fast Fluid Attenuated Inversion Recovery (FLAIR) sequences were randomly selected. All subjects were part of at 3-year follow-up study that was acquired by different parameters in each follow-up year. Therefore, a total of 30 case studies (628 images) were used in this validation of intensity standardisation. In addition, the dataset included the cranial MR images from six healthy subjects that were acquired by different parameters used to construct an image intensity standard scale model; these healthy subjects were between 38 and 55 years of age (mean age 45.00 ± 5.83 yr.). The subjects who were being used for intensity standardisation were between 40 and 62 years of age (mean age 48.90 ± 6.98 yr.). The subjects were scanned with an acquisition matrix of 512 x 512 for axial FLAIR and axial T1-weighted sequences. Both sequences had a slice thickness of 5.0 mm; all MR brain imaging were obtained from a 1.5T Signa HDx GE Scanner. Details of MRI protocol and parameters are given in clinical study of the protective effects of palm vitamin E tocotrienols on brain white matter [13]. The evaluation and experiment were performed on a computer with Intel Core i5-2450M CPU 2.50Ghz and 8.0 Gb installed memory (RAM) on 64-bit windows 7 operating system.

3.1 Manual Delineation and Features of White Matter Lesions

White matter lesions were delineated manually based on original image slices by an experience neuro-radiologist. The lesions annotation created manually was done by using the MIPAV [1] (Medical Image Processing, Analysis, and Visualization) software package. All lesion annotations converted into the binary mask in patches to extract the voxel intensities before and after the standardisation approaches. Grey-level co-occurrence matrix (GLCM) was constructed defined as second-order statistical texture features. Haralick features such as contrast, homogeneity, energy, and correlation are employed to describe the relationship between the grey level intensity in GLCM.

3.2 Preprocessing

Brain tissue intensity standardisation is a crucial step in brain lesion analysis especially when dealing with a supervised segmentation approach. Prior to

[1] http://mipav.cit.nih.gov/

intensity standardisation, skull stripping [14] and inhomogeneity N3 correction [15] are essential preprocessing steps in the proposed method. In this study, inhomogeneity N3 correction proposed by Sied *et al.* [15] is first performed on T1-W and FLAIR sequence images. This approach is mainly to eliminate the MRI artefact, which is caused by the receiver coil sensitivity variation. It is a necessary step that must be performed prior to the standardisation process according to the thorough evaluation by Madabhushi and Udupa [16]. Subsequently, the model-based level set method introduced by Zhuang *et al.* [14] was used to the perform skull stripping process, because T1-W is the best MRI sequence to show the structure of brain tissues. and thus, the preferred sequence for skull stripping. The skull-stripped T1-W sequence was used as a mask to extract the brain in the corresponding FLAIR sequence. Co-registration does not apply because the T1-W and FLAIR sequence images for each patient used in this study were well aligned because they are acquired at the same time during the acquisition process. Furthermore, a small degree of misaligned T1-W's mask applied on the corresponding FLAIR image does not have an impact on the proposed method. The reason for this is that, cerebrospinal-fluid (CSF) in the subarachnoid space in brain tissue is not used further in our processing; a small degree of mis-alignment would cut-off a few voxel of CSF only. Brain tissues such as grey matter and white matter do not affect this case.

3.3 The Proposed MR Intensity Standardisation

The proposed MR intensity standardisation method is flexible and can be customised into various regions of interest including brain and lesion images. The outlier detection method was used to automatically identify the landmarks of normal brain tissue voxel distribution. FLAIR images with skull stripping were used to construct a histogram and perform a smoothing operation with using 1D Gaussian kernel. An initial point is set at the full width at half maximum(FWHM) to perform a gradient descent process. Local minimum points P_{cl} and P_{cr} as shown in Fig. 1 were determined and used to compute the outliers P_{1i} and $P2_i$, respectively. In order to compute these outliers, a box-whisker plot was performed. The outlier, f_3 is defined in as Eq. (1)

$$f_3 = Q_3 + 1.5 \times IQR \qquad (1)$$

where IQR is the inter-quartile range (see Eq. (2)) that denotes the range of values falling within the 25th percentile, Q_1 and 75th percentile, Q_3 of the voxel distribution.

$$IQR = Q_3 - Q_1 \qquad (2)$$

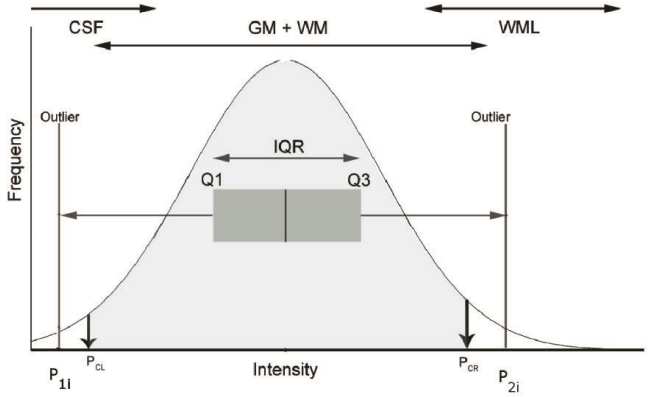

Fig.1 A pair of landmark [P_1i, P_2i] constructed from a histogram based on the outlier detection approach [12].

An overview of the proposed method is illustrated in Fig. 2. Generally, the proposed method consists of two main steps. They are called the training step and the transformation step [2]. Details of the standard intensity scale formation are described in the following section.

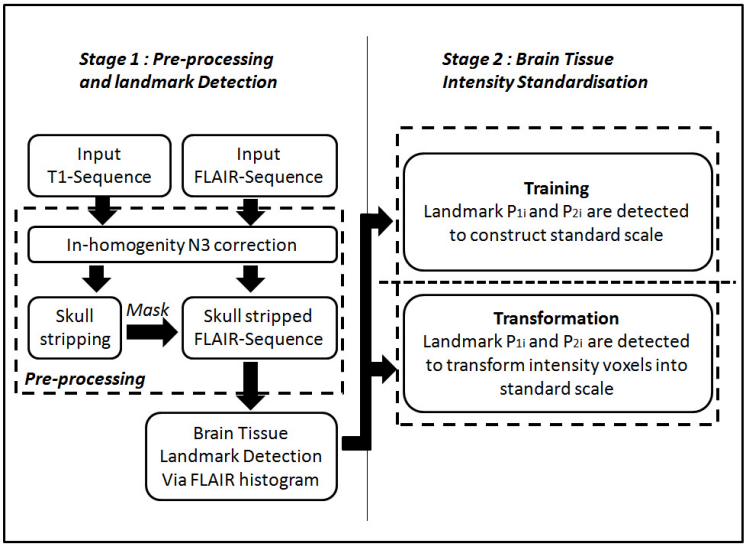

Fig.2 An overview of the proposed MR Intensity standardisation framework using landmark-based brain tissue analysis.

Training step : A set of preprocessed healthy brain images $v_i = g_i(x; y) \in V_{FLAIR}$ where V_{FLAIR} is the volume image of the FLAIR MR sequence and g is an intensity function in two dimensions that assign an integer intensity value for each v_i are used as input. The outlier detection method [12] that is based on the gradient descent approach is applied to automatically compute the landmarks (parameters) P_{1i} and P_{2i} from the H_i intensity histograms that were obtained from each v_i image. These landmarks are important keys in this study because they are used to separate the voxels of Cerebrospinal fluid (CSF), voxels of normal brain tissues (which include white matter (WM), and grey matter (GM)), and the voxels of hyperintensity (regions with high intensity) that are potentially related to image artefacts or white matter lesions.

A standard intensity scale [*L1*,*L2*] is constructed from landmarks P_{1i} and P_{2i} that are obtained from the intensity histogram where the histogram is generated using preprocessed healthy brain images as demonstrated in Fig. 1. The detected parameters are first computed into a range of intensity of interest (IOI) which is denoted as [H'_{1i}, H'_{2i}]. In this study, a range of [0, 4095] that is denoted as [G'_1, G'_2] are selected as IOI to ensure the range is under a lossless condition as reported by Nyu and Udupa [2]. A set of parameters $P_1'_i$ and $P_2'_i$ are computed using Eq. (3) and Eq. (4), respectively.

$$P'_{1i} = H'_{1i} + (P_{1i} - G'_1)\frac{H'_{1i} - H'_{2i}}{G'_1 - G'_2} \qquad (3)$$

$$P'_{2i} = H'_{1i} + (P_{2i} - G'_1)\frac{H'_{1i} - H'_{2i}}{G'_1 - G'_2} \qquad (4)$$

In the final step of the training process, the mean of a set of landmark parameters *L1* and *L2* is computed as shown in Eq. (5) and Eq. (6).

$$L1 = \frac{P'_{1i}}{n}, i = 1,2...n \qquad (5)$$

$$L2 = \frac{P'_{2i}}{n}, i = 1,2...n\,. \qquad (6)$$

where n is total number of training image slice.

Transformation step: The core idea of the transformation step is to transform all intensity voxels of each preprocessed brain image slice into a standard intensity scale. In other words, it is the deformed histogram H_i from each preprocessed brain image slice that is matched to the standard histogram. All slices of preprocessed brain images $v_i = g_i(x; y) \in V_{FLAIR}$. The landmark (parameters) P_{1i}

and P_{2i} for each histogram H_i are detected and mapped onto L1 and L2 of the standard intensity scale. Hence, every voxel is computed and transformed using three piecewise linear functions. The first linear function is transformed from [G'_{1i},$P1_i$] to [H'_{1i},$L1$]; the second linear function is transformed from [P_{1i},P_{2i}] to [$L1$,$L2$]; and the third linear function is transform from [P_{2i},G'_{2i}] to [$L2$,H'_{2i}]. The transformation process in linear mapping using the piecewise linear function can be illustrated in Fig. 3.

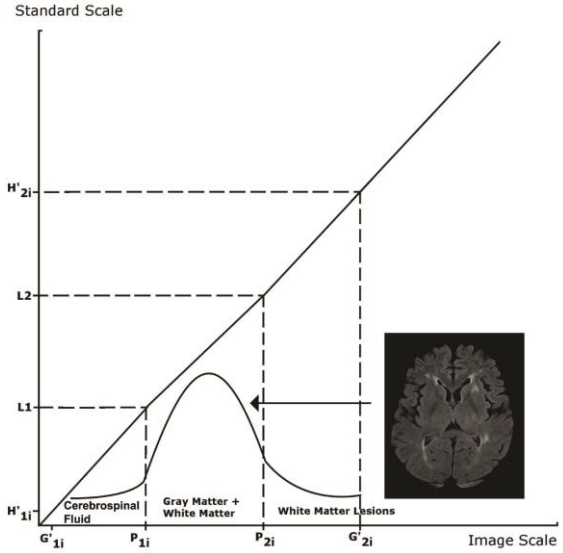

Fig.3 Intensity standardisation scale constructed using a landmark-based brain tissue analysis.

The proposed method is different from Nyu and Udupa [2] since the 0.02 percentile at the right-most part of the histogram is intended to be cut off. In this study, all of the intensity values will be fully utilised since the right-most tail of the histogram is important for white matter lesion detection and segmentation analysis. Hence, the right-most point of the histogram will be first estimated based on Eq. 7 instead of being limited to a 4095 intensity range. This is mainly because, for many of the abnormal brain images detected by outlier method, the right-most point will normally exceed the 4095 intensity range. In the proposed transformation step, the intensity of each voxel will then be mapped and transformed based on Eq. 8.

$$H'_{2i}(x) = L2 + (G'_{2i} - P_{2i})\frac{L1 - L2}{P_{1i} - P_{2i}} \qquad (7)$$

$$T(x)=\begin{cases} L1+(x-P_{1i})\dfrac{H'_{1i}-L1}{G'_{1i}-P_{1i}}, G'_{1i} < x \le P_{1i} \\ L2+(x-P_{2i})\dfrac{L1-L2}{P_{1i}-P_{2i}}, P_{1i} < x \le P_{2i} \\ H'_{2i}+(x-G_{2i})\dfrac{L2-H'_{2i}}{P_{2i}-G'_{2i}}, P_{2i} < x \le G'_{2i} \end{cases} \quad (8)$$

3.4 Validation Method

In the experiment, validation is done to depict the standardisation results obtained from the proposed method, which are compared with the original intensity value and existing standardisation methods named as histogram normalisation and decile-based histogram standardisation proposed by Loizou *et al.* [5,11] and Nyul *et al.* [3] respectively. The aim of the evaluation is to verify the following:

- The consistency of intensity level after the standardisation process by using the proposed method and existing standard methods.
- Minimisation of the change in intensity distribution after employing the proposed method and existing standardisation methods.
- Minimisation of the change in texture structure after employed proposed method and existing standardisation methods.

Therefore, we selected the appropriate evaluation method for the above verification. There are four types of evaluation methods in this study:

- Coefficient of variation (CV): Repeatability of intensity level;
- Kullback-Leibler divergence: similarity measure of intensity distribution;
- Pearson's correlation coefficient [26]: Change in texture structure;
- Fixed image contrast visualisation.

4 Results

4.1 Quantitative Evaluation

In order to measure the difference before and after standardisation results using the proposed method, a paired sample t-test was done. The significance test is performed based on the intensity distribution of a total of 30 case studies, 628 MR images for 10 subjects in this study. From the statistical analysis, the p-value is 1.42×10^{-15}, which is less than 0.0001 (conventional criteria). Therefore, the evaluation shows that there is a significant difference between the intensity results before and after standardisation process. The changes of intensity distribution for each subject and each time point are shown significant before and after the standardisation on the histogram in Fig. 4. Before the standardisation, the image intensity ranges of the distribution are varying from one to another (row (a)). After the proposed method is applied, the image intensity values of distribution in each study were changed according to the standard scale (row (b)). Subsequently, the proposed method was also compared with the decile-based standardisation method [3] and histogram normalisation(HN) [27]. In our studies, MR images for each of the 10 subjects with their 3-year follow-up studies were computed by using these standardisation methods. The decile-based standardisation was widely used in recent research because it was easy to customised into various anatomical regions and fast computation could be achieved due to less computational complexity. Lately, this method was thoroughly evaluated and used by Shah *et al.* [6] to investigate the significance of image intensity standardisation on the study of multiple sclerosis segmentation and classification of MRI.

In our evaluation, the proposed method is compared with the decile-based standardisation method proposed by Nyul *et al.* [3] using software called Computer Aided Visualization and Analysis Software System (CAVASS)[2]. Furthermore, the accuracy of white matter lesion segmentation and classification would be significantly affected by the computation of texture features which were mainly calculated from the intensity value of MR images. Therefore, the minimum change in texture information during transformation intensity into a standard scale was always critical. In recent literature, histogram normalisation [27] was evaluated by Loizou *et al.* [11] and it was proven that texture features were not affected after the standardisation process. Therefore, the histogram normalisation was also used in our results evaluation and comparison.

[2] http://www.mipg.upenn.edu/cavass/

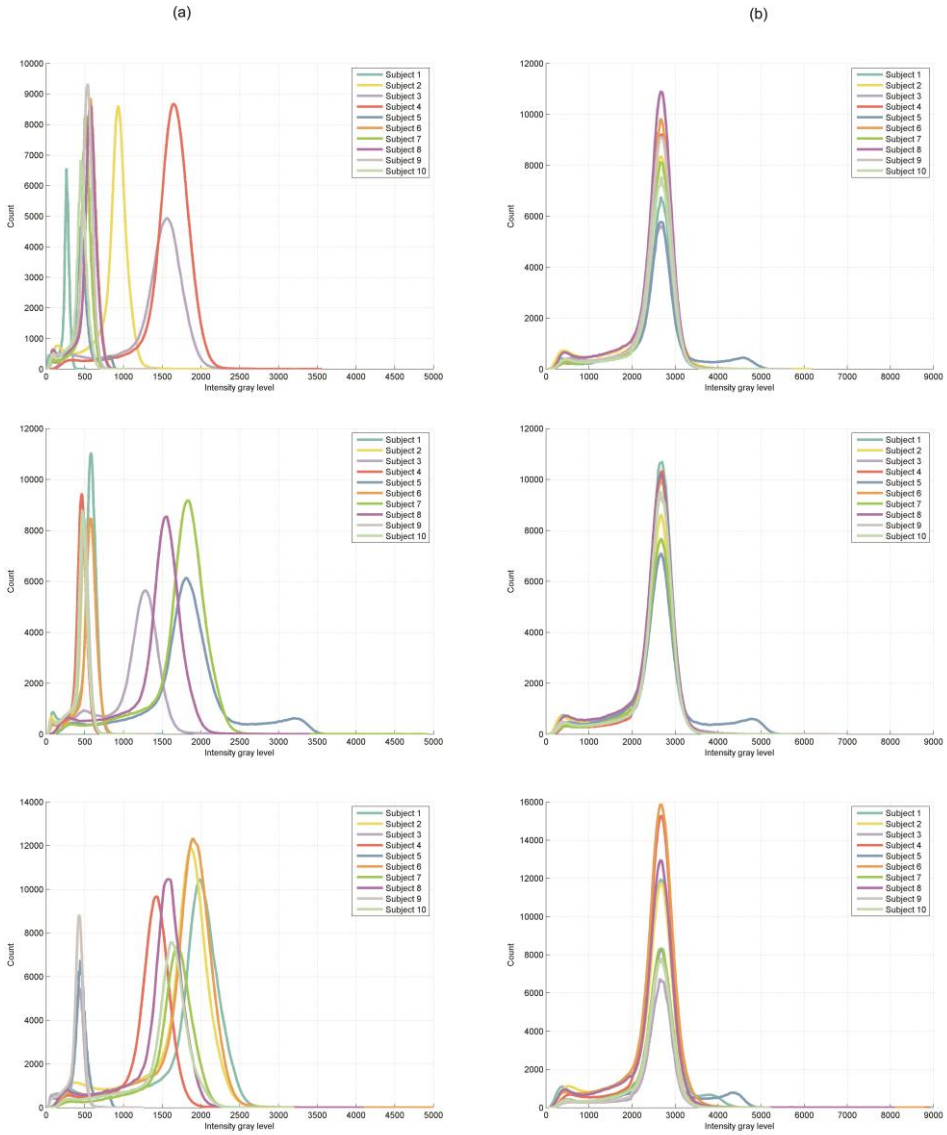

Fig.4 Image intensity distribution of each subject before (row (a)) and after (row (b)) standardisation. The first, second and third column are the base-year, 1-year and 2-year follow-up study.

The repeatability of the intensity scale was crucial to ensure the consistency of image features in MR images processing. Hence, with a good consistency in the intensity scale, the accuracy of the classification model in the white matter lesions segmentation was improved. Base on this fact, the CV was calculated based on intensity distribution for each of the 10 subjects during 3-year follow-up studies. The CV comparison among FLAIR MR Images before the standardisation process, the existing standardisation method, and our proposed method can be illustrated in Fig. 5. From the CV comparison within methods, a huge variation is shown in FLAIR MR Images before the standardisation process for each subject and their

follow-up studies. On the other hand, image intensity generated from HN, decile-based standardisation and proposed method showed a good consistency and trivial variation after the standardization process. In addition, the mean and standard deviation of CV for each method in percentages were calculated. Among the CV comparison of methods, the proposed method showed a good consistency with the smallest variation (1.25 ± 0.89) among other method. The proposed method showed the lowest CV measures with 1.25% in averaged over 30 studies as it was efficient to maintain the consistency of all images as shown in Table 1. The proposed method also showed small dispersion compared across 30 different studies where the standard deviation of CV is 0.89%. In addition, HN showed 2.03% of CV averaged over 30 studies, which were considered slightly higher than the proposed method. The decile-based standardization method shown is 91.59% lower than FLAIR MR Images before standardization process and 5.88% higher compared to the proposed method.

The minimum change in image distribution during the transformation process is essential to prevent inaccurate computation of intensity features and texture features. The Kullback-Leibler(KL) divergence [25] is employed to evaluate the distance in between the distribution of the standardised image and non-standardised image. In our observation, HN show a KL divergence $3.13 \times 10^{-5} \pm 2.85 \times 10^{-5}$ which was good and similar to the image before standardisation as shown in Table 2. It is the lowest value compared to the proposed method and decile-based standardisation as shown in Fig. 6. Apparently, decile-based standardisation received the highest value of KL divergence (1.04 ± 0.20), which indicated that an image intensity distribution change occurs during the transformation process. Our proposed method received a good similarity, which is seven times better than the decile-based standardisation method.

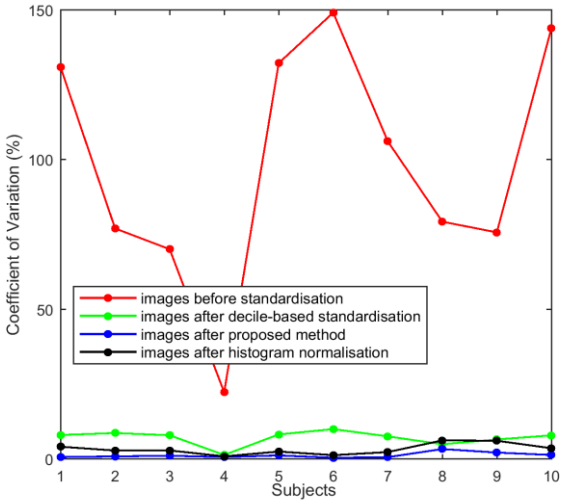

Fig.5 Coefficient of variation of all standardised FLAIR MR Images from each subject with their 3-year follow-up study.

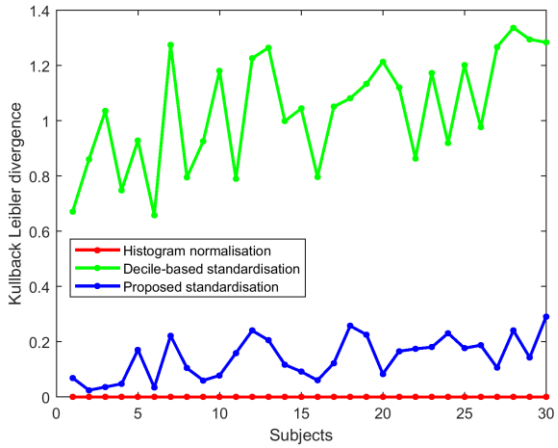

Fig.6 Kullback-Leibler(KL) divergence of all standardised FLAIR MR Images from each subject with their 3-year follow-up study.

Table 1: Average coefficient of variation, comparison before and after intensity standardisation among histogram normalisation, decile-based standardisation and the proposed method (mean ± std).

FLAIR MR Image Type	**Coefficient of Variation**
IBS	98.72±40.60
IAHN	3.28±1.80
IADS	7.13±2.42
IAPM	1.25±0.89

IBS : images before standardisation; IAHN: images after histogram normalisation; IADS: images after decile-based standardisation; IAPM: images after proposed method.

Table 2: Average Kullback-Leibler divergence, comparison of difference in image distribution before and after the standardisation process, use of histogram normalisation, decile-based standardisation and proposed method (mean ± std).

FLAIR MR Image Type	**Kullback-Leibler Divergence**
IAHN	$3.13 \times 10^{-5} \pm 2.85 \times 10^{-5}$
IADS	1.04 ± 0.20
IAPM	0.14 ± 0.08

IAHN: images after histogram normalisation; IADS: images after decile-based standardisation; IAPM: images after proposed method.

The accuracy of the classification model was highly associated with the image features. Hence, it is important to ensure the image features extracted are comparable before and after the standardisation process. In this evaluation, the WML were delineated from six different case studies, which were randomly selected from 3-year follow-up studies. A comparison correlation coefficient of each texture's relationship to FLAIR MR images for each method before standardisation is illustrated in Fig. 7. We report that the texture features of WML extracted based on images after the process with our proposed method and HN received the highest correlation compared to decile-based standardization. Furthermore, mean and standard deviation of each feature extracted from WML were also calculated and illustrated in Table. 3. The mean of four features computed from the image before standardisation and image after the proposed method and HN showed trivial difference. They are 0.003 (Contrast), 0.001 (Homogeneity), 0.000 (Energy), and 0.001 (Correlation). On the other hand, mean of each feature extracted from decile-based standardisation showed a huge difference compared to the image before standardisation. They are 2.014 (Contrast), 0.025 (Homogeneity), 0.064 (Energy) and 0.036 (Correlation). Based on this fact, the results of Pearson's correlation as shown in Fig. 7 are further confirmed.

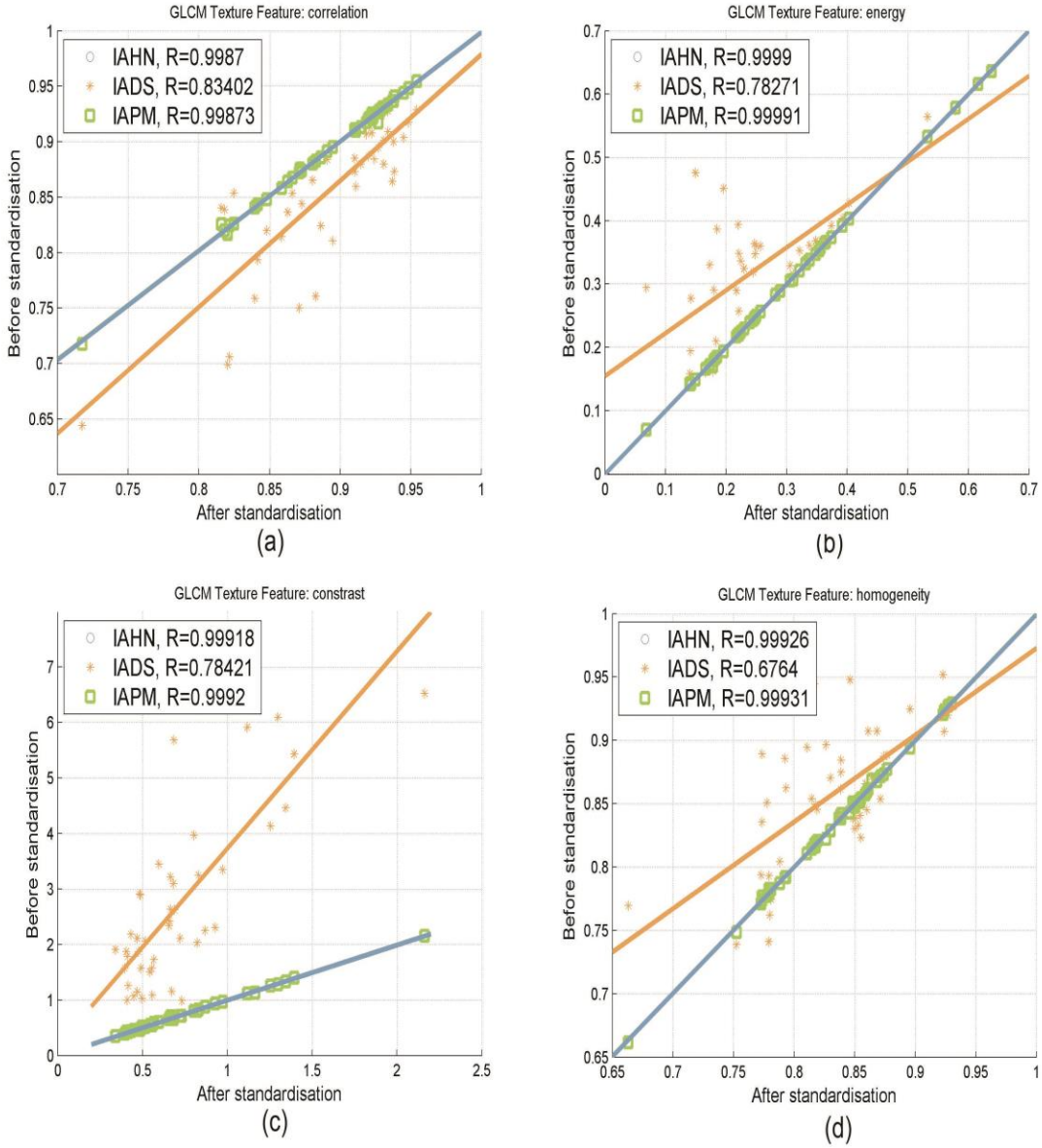

Fig.7 Pearson's correlation coefficient of four texture features of WML compared to each standardisation method, where IAHN is images after histogram normalisation; IADS is images after decile-based standardisation; IAPM is images after proposed method; std: standard deviation

Table 3: Average WML texture features delineated by experience radiologist computed using a GLCM approach, comparison before and after intensity standardisation among histogram normalisation, decile-based standardisation and proposed method (mean ± std).

FLAIR MR Image Type	GLCM Texture features			
	Contrast	Homogeneity	Energy	Correlation
IBS	0.722 ± 0.355	0.832 ± 0.052	0.281 ± 0.127	0.892 ± 0.049
IAHN	0.719 ± 0.352	0.831 ± 0.052	0.281 ± 0.127	0.893 ± 0.048
IADS	2.736 ± 1.609	0.857 ± 0.053	0.345 ± 0.11	0.856 ± 0.066
IAPM	0.719 ± 0.352	0.831 ± 0.052	0.281 ± 0.127	0.893 ± 0.048

IAHN: images after histogram normalisation; IADS: images after decile-based standardisation; IAPM: images after proposed method

4.2 Qualitative Evaluation

The quality of the MR image standardisation can be judged by using the naked eye as suggested by Nyu and Udupa [2] and Nyul *et al.* [3]. Fig. 8 shows an example of a 2D axial image slice from the different three-year follow-up studies with fixed window level. Four axial images in the first row before standardisation show different contrast at fixed window with a window level of 2602 and window width of 3659, this is mainly because of inconsistency in the grey-level intensity value. Apparently, the radiologist might need more time and effort to re-adjust the window level in order to visualise lesions accurately for each subject for a comparable lesions image assessment. This is worse especially when the image showed black indicating that the intensity value is out of the fixed window level. For example, in the first column with the first row as shown in Fig. 8. These axial images in the first row were then performed with the proposed standardisation as shown in the second row, decile-based histogram standardisation [3] is in the third row and histogram normalisation [11] is shown in the fourth row. Direct visual comparison at a fixed window level is obviously shows that our proposed method can be archived with better repeatability performance. In this experiment, these images were generated by using a decile-based standardisation histogram showing poor quality where the huge change in image texture could be observed. On the other hand, images after a HN process [11] show inconsistency among four axial images. We further judged the quality of the processed image from each method by generating these images as shown in Fig. 8 into intensity distribution. From Fig. 9b, it is seen that the intensity distribution that were processed by the proposed method were well aligned at the centre of distribution. This indicated that these distribution are well located at the standard scale. Apparently, the image before the standardisation process had been shown at nonstandard scale can be noticed as illustrated in Fig. 9a. It is noticed that with decile based standardisation, all

intensities were mapping well into 10 different landmarks [3] as demonstrated in Fig. 9c. However, these histogram plot do not form a normal distribution which reveal the inconsistency of the texture of brain. Hence, we do believe this finding might make a great impact on the structure of image content such as image texture. Thus, this is well explained by the worst correlation of texture features before and after the decile-based standardisation process. It is worthwhile to notice that the image (K) processed with histogram normalisation was out of the standard scale as reported in Fig. 9d. This is mainly because voxels of lesions have extreme hyperintensity, which were developed in the image (K) as shown in Fig. 8. Therefore, the right tail of the intensity distribution will shift entire distribution, which indicates a subject with moderate or severe lesion volume can cause the distribution to become non-standardised. On the other hand, this did not affected the proposed method since it was adapted to extend the right tail accordingly.

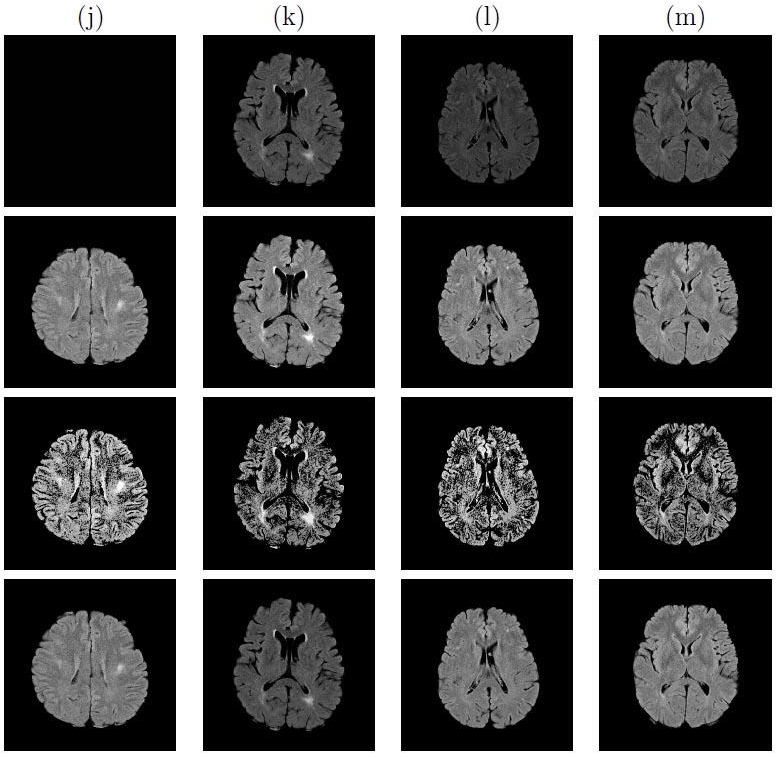

Fig.8 Individual slices from four different studies acquired using FLAIR sequence with variable lesion load. First row is the original FLAIR before the standardisation process. Second, third, and fourth rows are images after standardisation based on proposed method, deciles based histogram standardisation [3] and histogram normalisation [11], respectively. All image slices are adjusted at a fixed window with window level of 2602 and window width of 3659.

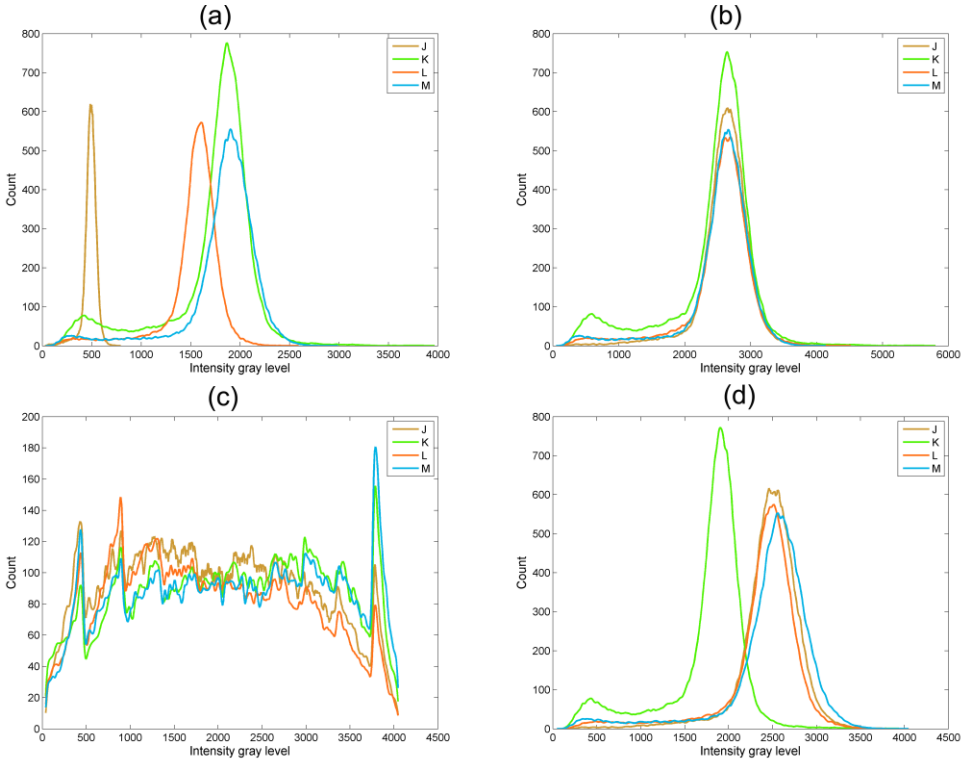

Fig.9 Intensity distribution generated from J, K, L, and M of individual image slices in Fig. 8. where (a) show the image intensity distribution before the standardisation process, (b) is image intensity distribution after the proposed standardisation process. (c) is image intensity distribution after decile-based standardisation process. (d) is image intensity distribution after histogram normalisation process.

5 Discussions

In this study, our aim is to standardise the image intensity scale of MRI for white matter lesions analysis. From the analysis and results, the method is applied on 660 MRI slices that consist of 10 subjects in a 3-year follow-up study. Images standardised using the proposed method have the relatively smallest value of coefficient variation compared with histogram normalization and deciles based histogram standardization, for each of the 30 studies as reported in Fig. 5. The standardised image processed by using the proposed method demonstrate a trivial variation for different subjects and different time points. This is shown that the proposed standardisation method able to correct the scanner sensitivity variations

and variations due to repeatability studies. Therefore, comparisons between MRI dataset from different time point and different subject become meaningful.

Furthermore, Fig. 6 has shown that the distribution of MRI intensity for each subject is similar to the distribution of intensity before standardisation using the proposed method since the mean and standard deviation of KL divergence is 0.14 ± 0.08. This is indicates the method shows the change of intensity level without influent the image data. The proposed method considers shows a good similarity to the image before standardisation. Therefore, it is proved that the method is an appropriate method for MR intensity standardisation on FLAIR sequence.

Texture features of WML delineated by experience radiologist computed using a GLCM approach between the image before and after standardisation using the proposed method, which shows a good correlation as illustrated in Fig. 7. Moreover, it is noticed that mean and standard deviation of contrast, homogeneity, energy and correlation extracted from the image after the proposed process were relatively close to the image before standardisation process as demonstrated in Table 3. These evaluations suggest that the proposed method is capable of a standardized MRI intensity scale and minimises the change in image texture during the transformation process. Hence, it is suitable as a key preprocessing step for white matter lesions analysis.

The proposed method is found robust comparing to existing methods in literature. The reason is our method is to include the landmarks L1 (Eq. 5) and L2 (Eq.6) computation that using the intensity of specific brain tissues namely white matter and grey matter on FLAIR sequence. The landmarks L1 (Eq. 5) and L2 (Eq.6) are essential to improve the consistency of the intensity scale from various studies with different time points on FLAIR sequence in MRI. Unlike other method in literature [1-3], the percentile or mode of intensity distribution was used as the main landmark without considering the information of specific brain structure. Furthermore, the exiting methods [1-3] is only suitable for small amount of lesion load images with assumption these amount of lesion voxel is above the 99.8 percentile value. However, a study conducted to show that standardised MS lesion image with decile-based histogram standardisation added advantage to increase the MS lesion segmentation accuracy [6].

In Fig. 9d, the histogram has explained that the HN method is not proper to be applied to FLAIR sequence but only appropriate to be implemented on T2-w sequence images which have been described by [11]. It had been observed that an image intensity distribution curve (curve K) is located away from the centre of standard scale which is processed using HN [27] method. This had shown that the method does not become robust to standardised image intensity on FLAIR sequence images. On the contrary, in our proposed method has implemented the retention of the information from the most-right tail of the image intensity

distribution which is essential for WML analysis. Therefore, our proposed method has been designed by the estimation of the right tail of the image intensity distribution adaptively (refer Eq. 7). As a result, most the intensity pixels on FLAIR images are taken into account in order to identify wider dynamic range on lesion images.

The proposed method has a limitation where it only works for FLAIR sequences of the skull stripped MR brain images. FLAIR sequence is chosen because it is the promising sequence used by various automated segmentation methods to detect and visualise white matter lesions in clinical practice. One of the advantages is that the proposed method works without the need to adjust any parameters manually. Landmarks parameters are detected automatically based on the specific tissues during outlier detection as described. Furthermore, the method is proved to be used for standardising the progression lesions dataset from the different time point. Hence, the proposed method is an essential preprocessing method which can be applied in an automatic procedure in white matter lesion analysis.

In addition, the proposed method also indirectly improves the work efficiency for visual assessment [22-24] in clinical practice, since the method enables radiologists to visualise the comparable lesions under one standard scale for each subject and every time point studied, for example standardised image demonstrated in Fig. 8.

6 Conclusion

Variations in acquisition protocols over time, especially in clinical follow-up studies can lead to non-standardisation intensity in MRI. Meaningful results which construct from these non-standardised images is not possible. Therefore, image intensity standardisation method used to WML identification with a supervised learning approach that relies heavily on intensity feature become critical. In this work, a new adaptive landmark based on brain tissue-specific standardisation method for FLAIR MR brain lesions images is presented. The method is enhanced based on the standardisation method proposed by Nyu and Udupa [2]. The main difference in our proposed method is that the landmarks (parameters) of the standard scale are automatic identified based on brain tissue information using outlier detection. The method, which is validated and evaluated based on a total of 660 MR images that consist of 10 subjects with their 3-year follow-up study are reported. The proposed methodology in this study was designed to standardise the FLAIR images without user intervention effectively. Furthermore, the proposed method does not involve a complex mathematical computation; it is proven to be fast and robust in successfully standardising the brain MR with WML images from the different subject and follow-up studies over

the time. The proposed method demonstrates a superior approach and achieves good results in comparison to existing intensity standardisation methods.

In our opinion, the proposed method can provide a better way to advance the accuracy of classification models to identify WML. However, a WML segmentation investigation is required to conduct in the near future to review the performance of proposed method to improve accuracy of WML identification and segmentation. Besides, the proposed method could also improve the visual assessment [22-24] performance and automatic quantitative assessment of WML progression for drug discovery and early diagnosis of WML treatments with a comparable image intensity result.

ACKNOWLEDGEMENTS

The authors would like to thank the Universiti Teknologi Malaysia (UTM) for their support in Research and Development, UTM Big Data Centre (UTM BDC) and the Soft Computing Research Group (SCRG) for the inspiration in making this study a success. This work is supported by Ministry of Higher Education (MOHE) under Research University Grant Q.J130000.2509.13H82 GUP TIER 1 and Q.J130000.2528.16H65 GUP TIER 1. The authors are thankful to Prof. Yuen Kay Hay at the School of Pharmaceutical Sciences, Universiti Sains Malaysia for his support and for providing the MRI datasets used in the study.

References

[1] 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

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

[3] 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

[4] 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

[5] 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

[6] 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

[7] Ekin A Pathology-robustmr intensity normalizationwith global and local constraints. In: Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on, 2011. IEEE, pp 333-336

[8] 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

[9] Wendt RE (1994) Automatic adjustment of contrast and brightness of magnetic resonance images. Journal of Digital Imaging 7:95-97

[10] Jäger F, Deuerling-Zheng Y, Frericks B, Wacker F, Hornegger J (2006) A new method for MRI intensity standardization with application to lesion detection in the brain. Procs 1010:269-276

[11] Loizou CP, Petroudi S, Seimenis I, Pantziaris M, Pattichis CS (2015) Quantitative texture analysis of brain white matter lesions derived from T2-weighted MR images in MS patients with clinically isolated syndrome. Journal of Neuroradiology 42 (2):99-114. doi:http://dx.doi.org/10.1016/j.neurad.2014.05.006

[12] 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

[13] Gopalan Y, Shuaib IL, Magosso E, Ansari MA, Bakar MRA, Wong JW, Khan NAK, Liong WC, Sundram K, Ng BH (2014) Clinical Investigation of the Protective Effects of Palm Vitamin E Tocotrienols on Brain White Matter. Stroke:STROKEAHA. 113.004449

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

[15] 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

[16] Madabhushi A, Udupa JK (2005) Interplay between intensity standardization and inhomogeneity correction in MR image processing. Medical Imaging, IEEE Transactions on 24 (5):561-576. doi:10.1109/tmi.2004.843256

[17] 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

[18] Khan AR, Wang L, Beg MF (2008) FreeSurfer-initiated fully-automated subcortical brain segmentation in MRI using Large Deformation Diffeomorphic Metric Mapping. NeuroImage 41 (3):735-746. doi:http://dx.doi.org/10.1016/j.neuroimage.2008.03.024

[19] Karimaghaloo Z, Shah M, Francis SJ, Arnold DL, Collins DL, Arbel T (2010) Detection of gad-enhancing lesions in multiple sclerosis using

conditional random fields. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2010. Springer, pp 41-48

[20] Scully M, Anderson B, Lane T, Gasparovic C, Magnotta V, Sibbitt W, Roldan C, Kikinis R, Bockholt HJ (2010) An automated method for segmenting white matter lesions through multi-level morphometric feature classification with application to lupus. Frontiers in Human Neuroscience 4. doi:10.3389/fnhum.2010.00027

[21] 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

[22] Van Straaten ECW, Fazekas F, Rostrup E, Scheltens P, Schmidt R, Pantoni L, Inzitari D, Waldemar G, Erkinjuntti T, Mantyla R, Wahlund LO, Barkhof F (2006) Impact of white matter hyperintensities scoring method on correlations with clinical data: The LADIS study. Stroke 37 (3):836-840

[23] Fazekas F, Chawluk JB, Alavi A (1987) MR signal abnormalities at 1.5 T in Alzheimer's dementia and normal aging. American Journal of Roentgenology 149 (2):351-356

[24] Wahlund LO, Barkhof F, Fazekas F, Bronge L, Augustin M, Sjo?gren M, Wallin A, Ader H, Leys D, Pantoni L, Pasquier F, Erkinjuntti T, Scheltens P (2001) A new rating scale for age-related white matter changes applicable to MRI and CT. Stroke 32 (6):1318-1322

[25] Kullback S, Leibler RA (1951) On information and sufficiency. The Annals of Mathematical Statistics:79-86

[26] Pearson K (1895) Note on regression and inheritance in the case of two parents. Proceedings of the Royal Society of London 58:240-242

[27] Nixon M, Nixon MS, Aguado AS (2012) Feature extraction & image processing for computer vision. Academic Press.