Data Augmentation Using Synthetic Lesions Improves Machine Learning Detection of Microbleeds from MRI

Saba Momeni[1,2(✉)], Amir Fazllolahi[1], Pierrick Bourgeat[1], Parnesh Raniga[1], Paul Yates[3], Nawaf Yassi[4], Patricia Desmond[4], Jurgen Fripp[1], Yongsheng Gao[2], and Olivier Salvado[5]

[1] CSIRO Health and Biosecurity, Australian E-Health Research Centre, Brisbane, Australia
saba.momeni@csiro.au
[2] Department of Engineering, Griffith University, Brisbane, Australia
[3] Department of Nuclear Medicine and Centre for PET, Austin Health Heidelberg, Heidelberg, Australia
[4] University of Melbourne, Parkville, Australia
[5] Data61, Brisbane, Australia

Abstract. Machine learning applied to medical imaging for lesions detection, such as cerebral microbleeds (CMB) from Magnetic Resonance Imaging (MRI), is challenged by the relatively small datasets available for which only subjective and tedious visual reading is available, and by the low prevalence of lesions (a few in $\sim$10% of a typical elderly cohort) resulting in unbalanced classes. Moreover, the lack of actual ground truth might limit the performance of any machine learning method to that of human performance. Yet, the automatic identification of those lesions is relevant to quantify cerebrovascular burden associated with dementia, such as identifying co-morbidity for Alzheimer's disease. In this paper, we investigated a novel approach consisting of simulating synthetic CMB on SWI MRI scans from healthy individuals to create a large and well characterized training dataset, as a data augmentation strategy. Firstly, we characterized actual CMBs from MRI SWI scans and designed a method to create realistic synthetic CMBs whose location, shape, appearance, and size are similar to actual CMBs. We then tested a supervised neural network classifier using various combinations of actual CMB and synthetic CMBs for training. Augmenting data with synthetic CMBs resulted in a large improvement over training on only actual CMBs only when tested on unseen lesions, and provided better results than other standard data augmentation approaches. Our results suggest that data augmentation using synthetic lesions can address the lack of ground truth and low prevalence limitations for medical imaging analysis allowing the deployment of data hungry supervised learning techniques such as deep learning.

Keywords: Cerebral microbleeds · Susceptibility weighted imaging SWI · Machine learning · Data augmentation · Synthetic data

A. Gooya et al. (Eds.): SASHIMI 2018, LNCS 11037, pp. 12–19, 2018.
https://doi.org/10.1007/978-3-030-00536-8_2

1 Introduction

Cerebral microbleeds (CMBs) are hemosiderin deposit caused by structural abnormalities of the blood vessels [1]. They are prevalent in people suffering from cognitive and cerebrovascular disease such as stroke and Alzheimer's disease. They are also asymptomatic and present in cognitively normal individuals. The detection of CMB from MRI is thus clinically important to assess cerebrovascular burden.

Magnetic susceptibility is affected by the presence of CMB in brain parenchyma [2], and susceptibility-weighted imaging (SWI) [2] has been shown to image CMB with excellent sensitivity, limited mostly by the imaging resolution (typically $1 \times 1 \times 2$ mm^3). In SWI scans, CMBs appear as small spherical hypo-intense areas in tissue with a similar appearance than the numerous blood vessels, also seen in SWI, when observed in cross-section. For this reason, the clinical observation of CMB, often based on the Microbleed Anatomical Rating Scale (MARS) [3], is tedious and time-consuming because of the large number of similar features from blood vessels. Indeed, CMB mimics include vessel cross section, calcification, and cavernous malformation on SWI.

Automated methods are also challenged by the large number of mimics resulting in numerous false positives while there is a relatively small number of true positives since the prevalence is about a few CMB in 10% of asymptomatic elderly subjects [4]. This creates a large unbalance classes hampering supervised training, further compounded by the limited size of datasets where ground truth (human expert annotation) is available. To address this problem, data augmentation methods are often used [5]. Synthetic Minority Oversampling Technique SMOTE [5], is one strategy to balance the classes by generating new pseudo TP (true positives) instances from the existing minority TP cases. In this paper, we investigate data augmentation by generating synthetic CMBs. We compare this new approach to SMOTE and the other widely used data augmentation method using rotation and translation. We describe the method to create realistic synthetic CMB and demonstrate that performance of a neural network classifier can be improved by transferring the learning from the synthetic data to real clinical data. We do not aim at presenting a new classifier, but at demonstrating that using synthetic lesions can improve the performance of supervised learning. We use in this paper a simple feed forward Artificial Neural Network as a case example.

2 Method

2.1 Dataset

Data are from the Australian Imaging Biomarkers & Lifestyle (AIBL). We used 64 scans from 39 participants (including some repeat scans at 18 months interval). There were 24 patients clinically diagnosed with Alzheimer's disease (AD), 34 subjects with mild cognitive impairment (MCI) and 6 healthy controls (HC), with an average age of 79.21, 77.74 and 79.98, respectively. The scans comprised 27 females and 37 males. All subjects underwent an anatomical T1-weighted (T1 W) and a SWI acquisitions on a 3T Siemens TRIO scanner. SWI was reconstructed online using the scanner system (software VB17). The 3D SWI parameters were: 0.93×0.93 mm^2 in-plane resolution

and 1.75 mm slice thickness, repetition time/echo time of 27/20 ms, and flip angle 20°. T1 W images were acquired using a standard 3D magnetization-prepared rapid gradient echo (MPRAGE) sequence with in-plane resolution 1.0 × 1.0 mm^2, slice thickness 1.2 mm, repetition-time/echo-time/TI = 2.300/2.98/900, flip angle 9°, field of view 240 × 256, and 160 slices. Scans were reviewed by two clinical experts and marked using the MARS criteria as definite and possible.

2.2 Preprocessing

For all the scans, N4 bias field correction technique [6], skull-stripping masking to exclude non-brain tissues [4], and histogram matching were applied before any processing. All SWI images were normalized between [0 1]. Brain tissues were segmented from the T1W into gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF) using posterior probability classification after fitting a mixture of Gaussian distribution to the histogram using the expectation-minimization method. A final region of interest (ROI) mask was built by merging the WM and GM voxels.

2.3 Features of Actual CMBs

From the analysis of the SWI datasets, CMBs were characterized using four parameters: size, shape, intensity, and location. In our dataset, the CMBs were uniformly located in the WM, GM tissues, and therefore location of synthetic lesions was uniformly distributed in the corresponding mask.

CMB Intensity: It was observed that large CMB intensity had a minimum value of 0, while the intensity of smaller size lesions became closer to tissue intensity. Figure 1 top left panel shows the minimum value of expert identified CMB. We assumed that this effect was entirely due to partial volume effect and that the minimum intensity value of a CMB is 0. This is consistent with SWI processing where the square of the high frequency filtered phase multiplies the signal magnitude, resulting in 0 intensity.

CMB Size: 11 × 11 × 11 patches with definite CMB at their center were extracted from the 64 SWI images, resulting in 144 patches making up the TP class. For each patch, a mixture of distributions was fitted using the expectation minimization algorithm: a Gaussian to model brain tissues and a uniform distribution to model outliers due to partial volume effect, and low intensity pixels from vessels and CMBs. A maximum a posterior classification was performed to create a mask of the CMB in the patch (cleaned using connected components). The mean of the Gaussian was assumed to be the tissue mean intensity while an intensity of zero was assumed to be the intensity of the CMB. Fractional CMB content was estimated for each pixel of the CMB mask by using a standard partial volume model. Eventually, fractions of CMB for each pixel were summed up to estimate the volume of each CMB as shown in Fig. 1.

2.4 Generation of Synthetic CMBs

It is our goal to create realistic synthetic CMB (sCMB), which should thus have the same characteristics as the real ones defined in the previous section.

A 3D Gaussian function was generated on a $110 \times 110 \times 110$ patch (ratio matching the SWI resolution ratio). Its location was the center of the middle pixel of the patch with added uniform random noise in all three axis within $[\pm 15, \pm 15, \pm 15]$ voxels. In order to simulate variation around a spherical shape, given a spherical volume, two of the 3D Gaussian standard deviations were multiplied by a random number between 0.5 and 0.9, while the third one was defined so that the overall volume was conserved. The shape was adjusted using the 3D Gaussian standard deviations to match the actual SWI resolution ratio. Random variation of the three standard deviation values of the 3D Gaussian allowed random variation of shape around a mean ellipsoid while keeping the same overall volume. The volume of the sCMB was randomly sampled from a smooth approximation of the size distribution of real CMBs (shown in Fig. 1 top left panel). Uniform random rotation was also added on the three axis within $\pm 30^{\circ}$.

The Gaussian intensity profile was thresholded at half maximum to create a mask simulating a high resolution sCMB (0 within the mask and 1 outside). The patch was then down sampled to $11 \times 11 \times 11$ to simulate partial volume effect (values ranging between 0 and 1 to simulate PVE). The resulting patches were then multiplied with patches extracted from the SWI images without any CMB, whose selection is explained in the next section. As a results, a synthetic TP patch (sTP) could be created for which the lesion size and appearance could be computed in the same way as actual CMB patches as explained in the previous section and displayed in Fig. 1 right panels.

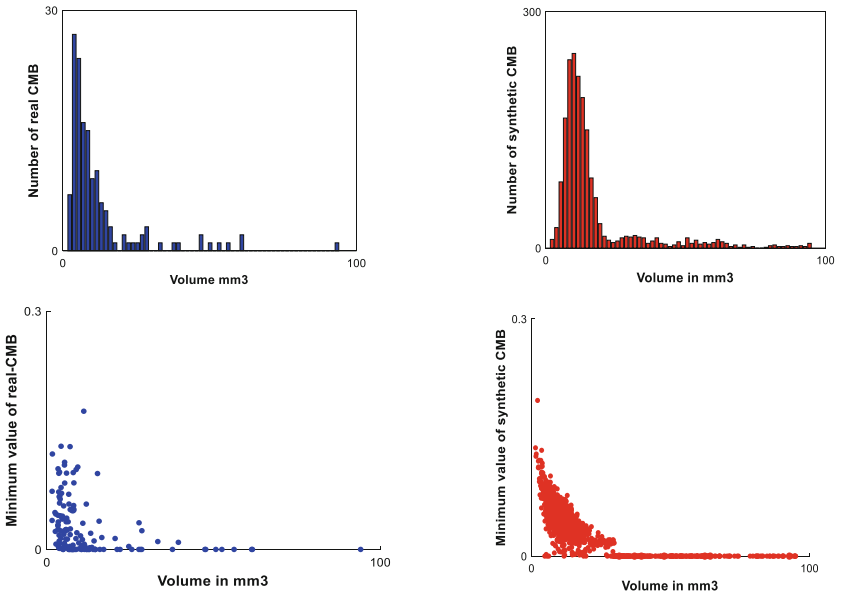

Fig. 1. The size distribution (top panels) and minimum intensity (bottom panels) for the real CMB in blue (left panels) and the generated synthetic CMB in red (right panels). The minimum intensities are comparable and entirely due to partial volume effect in the synthetic case. (Color figure online)

Table 1. Models considered in this study with approximate number of patches in each class.

Model	# of TP	# of sTP	# of TN	Comment
M1	70	0	140	Train on actual only
M2	70	0	1400	M1 but highly unbalanced
M3	70	1200	1400	Train on actual + synthetic
M4	0	1200	1400	Train only on Synthetic
M5	0	70	140	Train on synthetic with same size as M1
M6	70	1200	1400	Train on augmented data
M7	70	600 + 600	1400	Train on synthetic + augmented data
M8	70	1200	1400	Train on augmented data by SMOTE [5]

2.5 Patch Extraction

From the SWI dataset, patches were extracted and assigned to two categories: (1) Patches with real CMB defined as definite by the two experts (TP), (2) Patches with no actual CMB (neither definite nor possible). The second type was further divided into those patches containing either blood vessels or some CSF (seen as dark and possibly confounding CMB), and patches containing mostly GM or WM tissues. They were labeled as true negative (TN) with and without blood vessels. Vessels were detected using the Radial Symmetry Transform RST [7]. In order to create a patch with a synthetic CMB, the $11 \times 11 \times 11$ mask described in the previous section was multiplied with a patch from the TN class, creating a synthetic TP patch (sTP). We selected the TN patches so that the proportion of sTP with and without vessels was 50%/50%.

2.6 Experimental Method

We used a 2-fold with repeated 25 random sub-sampling validation: for each of the 25 sampling, 19 random subjects (out of 39 subjects) were set aside as the testing data, while the remaining were used for training. This resulted in a total number of TP (actual CMB patches) to be around 70 for testing and 70 for training (actual number depends on the subjects selected). The description of the training strategy for the 8 models can be found in Table 1. The concatenation of the results for the 25 sampling for each model was used to smooth the ROC curves.

2.7 Classifier

To test our hypothesis that synthetic lesion would improve the performance of a classifier, we used a single-hidden layer feed-forward neural-network [8], trained with scaled conjugate gradient [9]. This ANN was implemented with Matlab. The activation function in the hidden layer was Leaky ReLU [10] with 55 neurons. The activation function of the output unit was the logistic sigmoid (LOSI) [11]. The number of epochs was 2000, the mean square error was used as the cost function.

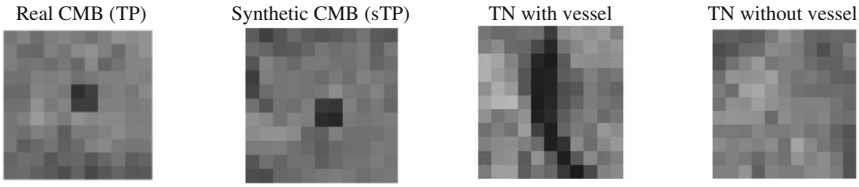

Fig. 2. Typical examples of axial section for the different classes. Note that the sCMB is not centered (second left panel), similar to the real one (left panel), because we added random variation in location and shape.

3 Results and Discussion

3.1 Synthetic CMB Generation

Synthetic CMB could be created with no discernable differences from actual lesions. The size and intensity were similar between actual and synthetic lesions as shown in Fig. 1. The location of synthetic samples was uniform in the GM and WM tissues as observed for real ones. Figure 2 shows examples of axial sections from the $11 \times 11 \times 11$ patches.

3.2 Results of CMB Classification Using ANN

Results for the ANN classification on the testing datasets is reported in Table 2. We show the area under ROC (AUR), mean sensitivity, specificity and accuracy along with the standard deviation over the 25 sampling 2-fold cross validation design. Model 2 would not converge to any meaningful results due to the unbalanced classes and is omitted here. Figure 3 shows the ROC curve for all the models. The results using SMOTE (model 8) were the lowest of the three data augmentation methods. Using synthetic data provided good results, and the best model was the one training on synthetic data alone. Merging synthetic with the TP class had also good results close to training on synthetic data alone. Using standard geometric data augmentation improved performance but not as much strategy involving using synthetic data.

Table 2. Performance of the models on the test datasets. Values are averages ± standard deviation over the 25 draws from averaging the outputs. Bold shows the best performance models.

Model	Accuracy	Sensitivity	Specificity	AUR
M1	0.88 ± 0.022	0.87 ± 0.028	0.88 ± 0.030	0.93 ± 0.020
M3	**0.91 ± 0.021**	**0.90 ± 0.035**	**0.91 ± 0.026**	**0.95 ± 0.018**
M4	0.90 ± 0.021	**0.90 ± 0.029**	**0.91 ± 0.030**	**0.95 ± 0.018**
M5	0.83 ± 0.033	0.85 ± 0.037	0.82 ± 0.050	0.89 ± 0.024
M6	0.88 ± 0.014	0.87 ± 0.035	0.88 ± 0.024	0.93 ± 0.012
M7	0.90 ± 0.019	0.89 ± 0.034	**0.91 ± 0.026**	**0.95 ± 0.013**
M8	0.85 ± 0.020	0.87 ± 0.032	0.84 ± 0.031	0.91 ± 0.016

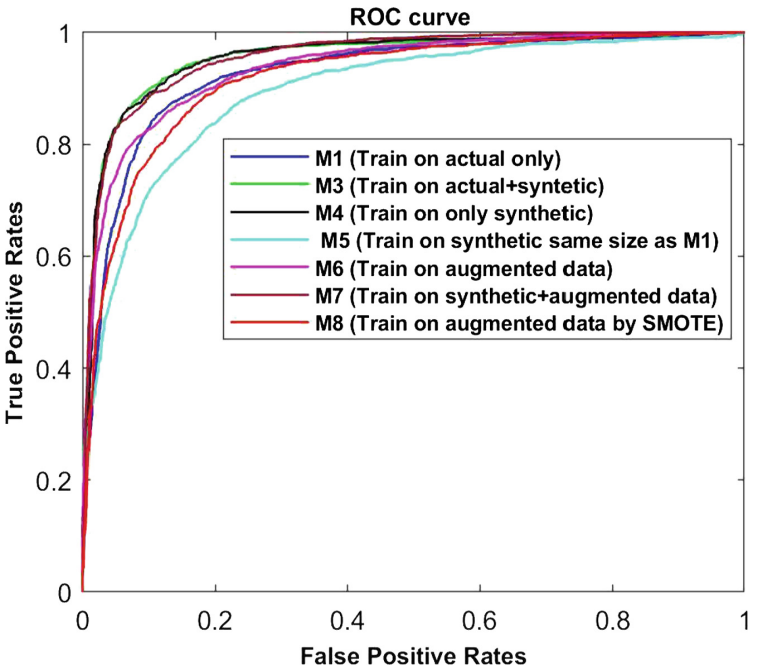

Fig. 3. ROC curve for the different models using a 2-fold with repeated 25 random subsampling validation design.

3.3 Discussion

Our results suggest that augmenting data with synthetic lesions could help increasing supervised classifier performance to detect microbleeds in MRI SWI. The best results were obtained by training on synthetic data alone even though the performance were measured on real data.

Our proposed approach addresses one critical issue for lesion detection task: when the prevalence of a lesion type is low, a very small number of TP are available for training supervised classifier limiting the complexity of the possible models to consider. Another consequence is the class unbalanced between the positive and negative. Indeed, the naïve approach of training the ANN on very unbalanced data (1:10) failed to produce any meaningful results. Adding sCMB allowed to balance the classes and increased the number of examples resulting in much improved performance.

Our proposed approach is very dependent on how realistic the synthetic lesions are created. We described a simple method based on a Gaussian profile that is added to the magnitude SWI images (magnitude multiplied by filtered phase). A better characterization of actual microbleeds using high resolution MRI, such as 7T, could help improve the generation of simulated lesions. In addition, adding the synthetic lesions on the complex image, before SWI processing, might improve the realism of the synthetic data.

Despite those limitations, augmenting medical data using synthetic lesions could allow to use complex classifiers, such as deep learning network, with large number of parameters since arbitrary large training datasets could be created along with the perfect ground truth necessary for supervised learning. Of course that approach is limited by how close the synthetic ground truth is from the actual and elusive ground truth. This is tantamount to converting human expertise (a priori knowledge about what lesions look like) into data that is then transfer through learning into the weights of a classifier. More research on different applications would inform whether our proposed approach could allow automated methods to push performance beyond that of experts.

References

1. Martinez-Ramirez, S., Greenberg, S.M., Viswanathan, A.: Cerebral microbleeds: overview and implications in cognitive impairment. Alzheimers Res. Ther. **6**(3), 33 (2014)
2. Haacke, E.M., Mittal, S., Wu, Z., Neelavalli, J., Cheng, Y.-C.N.: Susceptibility-weighted imaging: technical aspects and clinical applications, part 1. Am. J. Neuroradiol. **30**(1), 19–30 (2009)
3. The Microbleed Anatomical Rating Scale (MARS) | Neurology. http://n.neurology.org/content/73/21/1759. Accessed 30 Apr 2018
4. Fazlollahi, A., et al.: Computer-aided detection of cerebral microbleeds in susceptibility-weighted imaging. Comput. Med. Imaging Graph. **46**(Part 3), 269–276 (2015)
5. He, H., Garcia, E.A.: Learning from imbalanced data. IEEE Trans. Knowl. Data Eng. **21**(9), 1263–1284 (2009)
6. Tustison, N.J., et al.: N4ITK: improved N3 bias correction. IEEE Trans. Med. Imaging **29** (6), 1310–1320 (2010)
7. Loy, G., Zelinsky, A.: Fast radial symmetry for detecting points of interest. IEEE Trans. Pattern Anal. Mach. Intell. **25**(8), 959–973 (2003)
8. Bin, L., Xuewen, R.: Review and performance analysis of single hidden layer sequential learning algorithms of feed-forward neural networks. In: 2013 25th Chinese Control and Decision Conference, CCDC 2013, pp. 2170–2175 (2013)
9. Livieris, I.E., Pintelas, P.: A new conjugate gradient algorithm for training neural networks based on a modified secant equation. Appl. Math. Comput. **221**, 491–502 (2013)
10. Aghdam, H.H., Heravi, E.J., Puig, D.: Recognizing traffic signs using a practical deep neural network. Robot 2015: Second Iberian Robotics Conference. AISC, vol. 417, pp. 399–410. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-27146-0_31
11. Zadeh, M.R., Amin, S., Khalili, D., Singh, V.P.: Daily outflow prediction by multi layer perceptron with logistic sigmoid and tangent sigmoid activation functions. Water Resour. Manag. **24**(11), 2673–2688 (2010)