Observer performance in semi-automated microbleed detection

Hugo J. Kuijf[a], Manon Brundel[b], Jeroen de Bresser[c], Max A. Viergever[a], Geert Jan Biessels[b], Mirjam I. Geerlings[d], and Koen L. Vincken[a]

[a]Image Sciences Institute, University Medical Center Utrecht, Utrecht, the Netherlands;
[b]Department of Neurology, Rudolf Magnus Institute of Neuroscience, University Medical Center Utrecht, Utrecht, the Netherlands;
[c]Department of Radiology, University Medical Center Utrecht, Utrecht, the Netherlands;
[d]Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, the Netherlands

ABSTRACT

Cerebral microbleeds are small bleedings in the human brain, detectable with MRI. Microbleeds are associated with vascular disease and dementia. The number of studies involving microbleed detection is increasing rapidly. Visual rating is the current standard for detection, but is a time-consuming process, especially at high-resolution 7.0 T MR images, has limited reproducibility and is highly observer dependent. Recently, multiple techniques have been published for the semi-automated detection of microbleeds, attempting to overcome these problems.

In the present study, a 7.0 T dual-echo gradient echo MR image was acquired in 18 participants with microbleeds from the SMART study. Two experienced observers identified 54 microbleeds in these participants, using a validated visual rating scale.

The radial symmetry transform (RST) can be used for semi-automated detection of microbleeds in 7.0 T MR images. In the present study, the results of the RST were assessed by two observers and 47 microbleeds were identified: 35 true positives and 12 extra positives (microbleeds that were missed during visual rating). Hence, after scoring a total number of 66 microbleeds could be identified in the 18 participants.

The use of the RST increased the average sensitivity of observers from 59% to 69%. More importantly, inter-observer agreement (ICC and Dice's coefficient) increased from 0.85 and 0.64 to 0.98 and 0.96, respectively. Furthermore, the required rating time was reduced from 30 to 2 minutes per participant. By fine-tuning the RST, sensitivities up to 90% can be achieved, at the cost of extra false positives.

Keywords: Observer performance evaluation, technology assessment, technology impact

1. INTRODUCTION

Interest in cerebral microbleeds has been increasing rapidly in the past years and the number of studies involving microbleed detection is still growing. Microbleeds are small spherical lesions that are visible as hypointensities on gradient echo MR scans. Studies showed that microbleeds are associated with cerebrovascular disease and dementia.[1,2,3] The current standard for microbleed detection is visual rating with validated visual rating scales, such as the Microbleed Anatomical Rating Scale (MARS) or the Brain Observer MicroBleed Scale (BOMBS).[4,5] However, visual rating is known to be time-consuming, has limited reproducibility, and is highly observer dependent. The challenges for visual rating are most evident for high resolution scans obtained at high field strength.[6]

The recent introduction of semi-automated techniques for microbleed detection is likely to overcome these problems. These techniques detect potential microbleed locations and present them to an observer to identify true microbleeds and censor any false positives. The detection of potential microbleed locations in these

Corresponding author:
Hugo J. Kuijf, Image Sciences Institute, University Medical Center Utrecht, Heidelberglaan 100, Room Q0S.459, 3584CX, Utrecht, the Netherlands. E-mail: h.kuijf@umcutrecht.nl. Phone: +31 88 75 58562.

(a) First echo   (b) Second echo
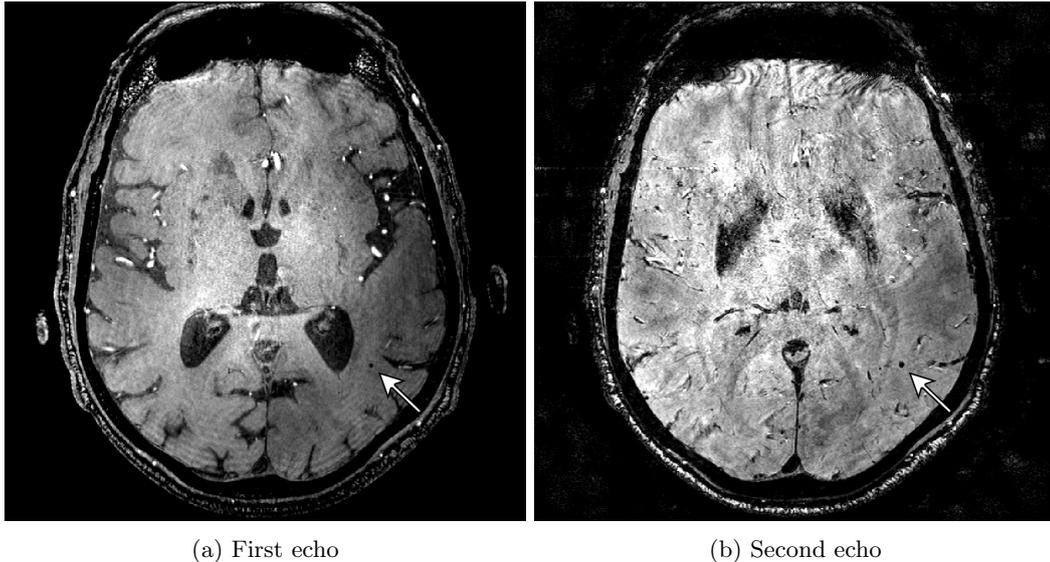
Figure 1: A typical slice of a 7.0 T dual-echo gradient echo MR scan, showing the first and second echo. The arrows annotate the same microbleed in both figures.

methods is performed with different image processing techniques, including a unified segmentation-normalization approach,[7] a combination of statistical thresholding and a support vector machine supervised learning classifier on susceptibility weighted images,[8] and the radial symmetry transform (RST).[9] The sensitivity of these techniques ranges from 50%[7] to 82%[8] and the time required to censor false positives ranges from 2[9] to 15[7,8] minutes per scan.

Of these techniques, the RST has been applied on high resolution 7.0 T dual-echo gradient echo MRI scans that have a high susceptibility for microbleed detection. However, the visual detection of microbleeds on these scans is difficult owing to the high number of slices, the smaller size of microbleeds, and the fact that other small structures are also visualized at high field strength.

In this study, we assess the performance of the RST as a tool for semi-automated microbleed detection on 7.0 T scans, and the performance of observers with and without support by the RST.

2. METHODS AND MATERIALS

2.1 Participants

For this study, 18 participants (mean age: 60 years, sd: 13 years; 15 men and 3 women) with microbleeds were included from the Second Manifestations of ARTerial disease (SMART) study.[10] The objectives of the SMART study are to determine the prevalence of vascular risk factors and concomitant arterial disease and to study the incidence of future cardiovascular events and its predictors in patients newly referred to our hospital with atherosclerotic disease. The SMART study and the 7.0 T imaging were approved by the Medical Ethics Committee. Written informed consent was given by all participants. Of the 18 participants, 6 were included with diabetes mellitus, 4 with cerebrovascular disease, 2 with coronary artery disease, 2 with peripheral artery disease, 1 with hypertension, and 3 with other vascular risk factors or disease.

2.2 MRI acquisition

MRI acquisition was performed as described previously by Conijn et al.,[11] on a 7.0 T whole-body system (Philips Healthcare, Cleveland, OH). A dual-echo gradient echo sequence was acquired (first TE: 2.5 ms, second TE: 15.0 ms, TR: 20.0 ms) and the images were reconstructed to $0.35 \times 0.35 \times 0.30$ mm^3. A typical transversal slice of a high-resolution 7.0 T scan is shown in Figure 1.
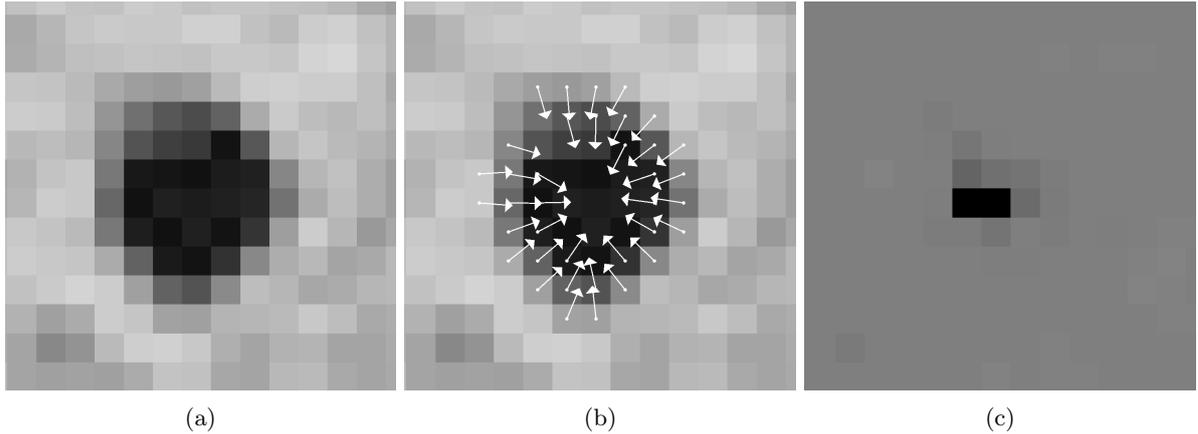
Figure 2: a) Enlargement of the microbleed shown in Figure 1b. b) Gradient information is computed as part of the RST. c) Final output of the RST, highlighting the center voxel of the microbleed. Thresholding on this final output results in potential microbleed locations.

2.3 Visual rating of microbleeds

All scans were independently and visually scored by two observers (observers 1 and 2), both neuroradiologists with more than 20 years of experience, as part of a previous study.[12] Rating was performed according to the MARS, using minimal intensity projection post-processing of both echo times of the scan. The observers were blinded to all other clinical information. The lesions on which the two observers disagreed were discussed in a consensus meeting to obtain a final score. In case of discordance, an additional observer was consulted.

2.4 Radial symmetry transform

The RST is an image processing technique that is originally used to highlight circular objects in 2D images. It has numerous medical and non-medical applications, including the detection of nuclei in H&E stained breast cancer biopsy images, detection of the sternum for costal cartilage segmentation, and the detection of eyes in pictures of human faces.[13, 14, 15] A 3D version of the RST was implemented to detect potential microbleed locations in 7.0 T dual-echo gradient echo MR images.[9] The procedure is illustrated in Figure 2.

Full details of the procedure are described by Loy *et al.* and Kuijf *et al.*[9, 15] In short, the procedure that is computed for both echoes of a dual-echo 7.0 T scan is: 1) compute 3D gradient information at a certain scale, 2) for every voxel, "count" the number of neighboring voxels within some radius, where the gradient direction of that voxel "points" towards the current voxel (see Figures 2a and 2b), and 3) combine the output of the first and second echo.

The center voxel of any spherical object in the scan will receive a high value and thresholding the output will result in potential microbleed locations. Two parameters are involved: threshold θ_1 on the first echo and threshold θ_2 on the second echo.

2.5 Experiments

Optimal settings for θ_1 and θ_2 were defined in a previous study.[16] Two human observers (observers 3 and 4, not involved in the original visual rating) each censored all potential microbleed locations detected by the RST using a dedicated tool with a user-friendly interface developed in MeVisLab (MeVis Medical Solutions AG, Bremen, Germany[17]). A meeting was held to obtain consensus for the cases on which the two observers disagreed and an additional observer was consulted in case of discordance.

Microbleeds detected by the RST that were not in the original visual rating were added to the ground truth if both observers confirmed them as true microbleeds in the consensus meeting. This improved the ground truth and gave a fair comparison of the sensitivities of human observers versus the RST.

3. RESULTS

In total, 54 microbleeds were found to be present in the 18 participants after the consensus meeting of the visual ratings. Of those microbleeds, 25 were detected by both observers, 9 only by observer 1, 19 only by observer 2, and 1 microbleed was initially not detected by one of the observers, but was detected during the consensus meeting. Both observers identified additional lesions that were rejected as microbleeds during the consensus meeting: 4 were found by observer 1 and 18 by observer 2. Based on these numbers, observer 1 has a sensitivity of 63%, observer 2 has a sensitivity of 81%, the intra-class correlation coefficient (ICC) is 0.85, and Dice's coefficient is 0.64.

The RST identified 353 (mean±sd: 20±12) potential microbleed locations: 35 true positives (microbleeds present in the visual rating), 309 false positives, and 12 extra positives that were not present in the visual rating, but were marked by the RST and confirmed as true microbleeds by observers 3 and 4. Of the 47 detected microbleeds, 43 were identified by both observers, 3 only by observer 3, and 1 only by observer 4. Both observers identified one location as microbleed that was rejected in the consensus meeting.

In total, 66 microbleeds (54 original visual rating, 12 additional after RST) were detected and the final results are shown in Table 1. ICC of observers 3 and 4 after using the RST is 0.98 and Dice's coefficient is 0.96.

Visual rating required on average 30 minutes per participant per observer. RST supported visual rating required much briefer rating times. Observer 3 required 43 minutes to censor all detected microbleeds (2.5±1.5 minutes per participant; 7.9±3.3 seconds per detected location), observer 4 required 28 minutes (1.5±1 minutes per participant; 5.1±2.2 seconds per detected location).

As explained in the Methods section, the RST depends on two parameters: θ_1 and θ_2, which were set at -20 and -100, respectively. Both parameters have negative values, since the RST returns negative values for dark spots, as can be seen in Figure 2c. Varying these parameters influences the results, as shown in Figure 3.

By allowing more false positive locations to be detected, which need to be censored manually, the sensitivity of the RST can achieve 90%. It is expected that additional extra positives will be found for combinations of θ_1 and θ_2 that allow extra locations to be detected.

4. DISCUSSION

The present study is the first that examined the performance of observers using tools for semi-automated microbleed detection. Using the RST, the inter-observer agreement improved, as can be seen from the increment in ICC (from 0.85 to 0.98) and Dice's coefficient (from 0.64 to 0.96). This will probably reduce the time required for consensus meetings, and improve rating reliability and reproducibility.

It must be noted that a bias might be present in the constructed ground truth. After the initial visual rating, the two observers (1 and 2) discussed the lesions they disagreed on in a consensus meeting with a third observer and finally 54 microbleeds were identified. The RST detected an additional 12 microbleeds, identified by observers 3 and 4, that were added to the ground truth rating, resulting in a total of 66 microbleeds. The inclusion of microbleeds detected by the RST causes a bias in the results with regards to the sensitivity of the method. However, any visual rating of microbleeds on 7.0 T images will not result in a ground truth rating that

Table 1: Sensitivity of the visual rating (observers 1 and 2), the RST, and observers 3 and 4 when censoring the results of the RST. A total of 66 microbleeds was present in the images.

	#TP	#FP	#FN	Sensitivity
Observer 1	34	4	32	51%
Observer 2	44	18	22	67%
Observers 1+2	53	0	13	80%
RST	47	309	1	71%
Observer 3	46	1	20	70%
Observer 4	44	1	22	67%

(a)

(b)

(c)
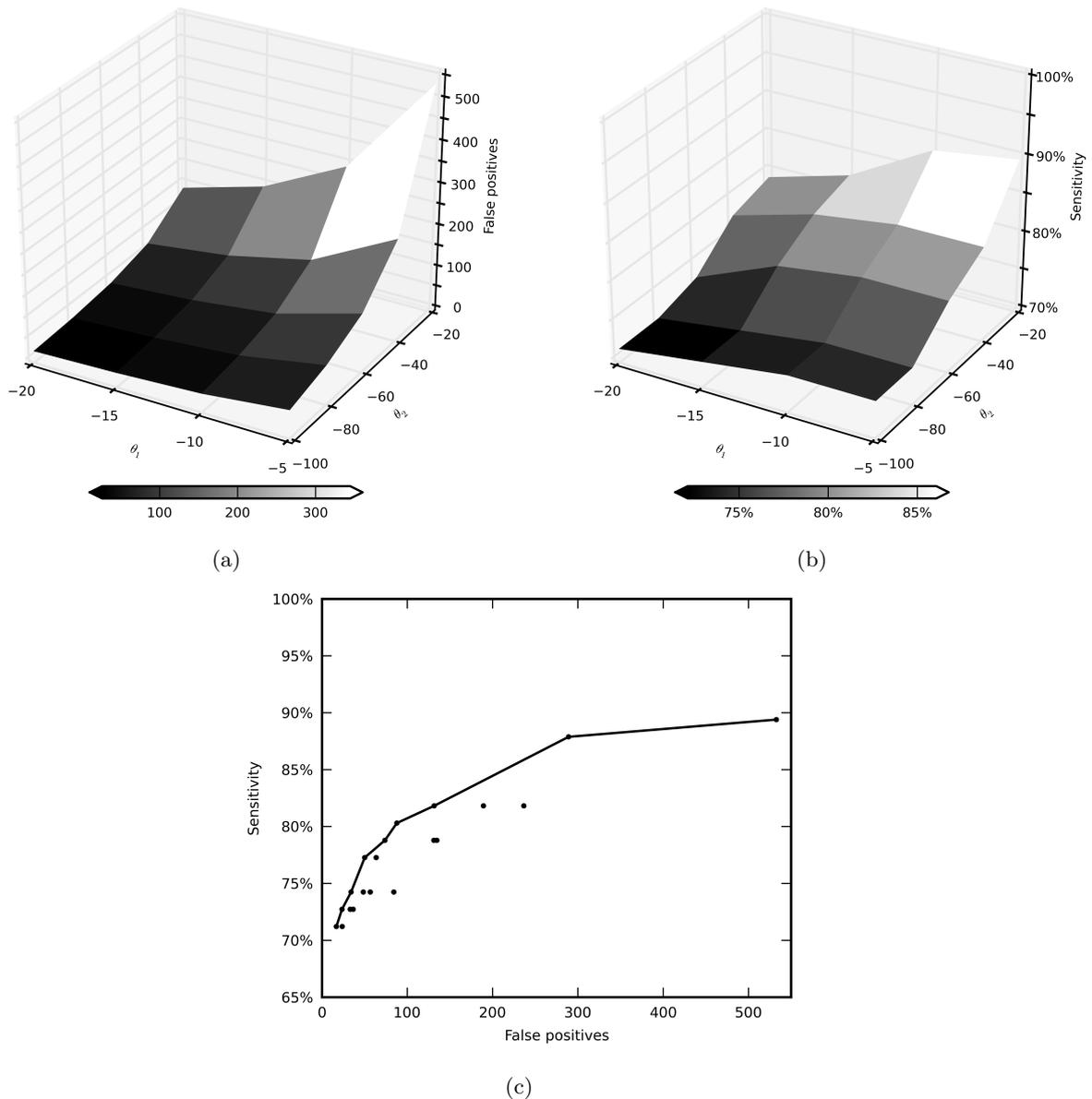
Figure 3: a) Average number of false positive locations per participant detected as a function of θ_1 and θ_2. b) Sensitivity of the RST as a function of θ_1 and θ_2. c) FROC curve, where each point corresponds to a combination of θ_1 and θ_2 and the solid line indicates the optimal combinations.

can be considered the gold standard, but rather as a "silver" standard, owing to the difficulties involved with visual rating.[6,12] The inclusion of extra microbleeds into the ground truth will improve the ground truth and give a fair comparison of the sensitivities of human observers and automated methods.

The increase in inter-observer agreement is explained by the fact that the RST directs the attention of all observers to the same potential microbleed locations. This reduces the effect of the individual detection strategy that each observer applies during visual detection. Although normal visual detection is performed with standardized visual rating scales, which also improve the inter-observer agreement, individual differences between observers are still present. Semi-automated techniques reduce – but do not fully eliminate – these individual observer differences.

Some differences between observers are still present with the use of semi-automated techniques, although to a minimal extent. These differences are caused by the subjectivity that is involved in the rating of microbleeds. There is a standardized definition of true microbleeds, but censoring so-called microbleed mimics is a difficult and subjective task. Therefore it is unlikely that an inter-observer agreement of 1 is ever achieved, even with the use of semi-automated techniques. The final decision if a potential microbleed location is a true microbleed is still the responsibility of the medical expert.

Censoring false positives required on average 2 minutes per participant by an experienced observer, as compared to 30 minutes for full visual rating of microbleeds. This is a significant reduction in time that allows observers to rate more participants in a given amount of time and enables the inclusion of larger patient cohorts in research studies. Furthermore, the improved inter-observer agreement shortens the time required for consensus meetings, since fewer potential microbleed locations have to be discussed.

The sensitivity of the RST is limited, but higher than that of individual observers who are not supported by the RST. In situations where only a single observer is assessing all scans, applying the RST would be useful to achieve a higher sensitivity. By altering parameters of the RST, the sensitivity can be increased up to 90% at the cost of more false positive locations, and hence increasing the rating time. As can be seen in Figure 3c, an increase of 10% in sensitivity will result in five times as many false positives. This requires about ten minutes to censor the false positives, which would still be significantly less than the time required for a full visual rating on 7.0 T images.

Future research will focus on reducing the number of false positives, allowing individual observers to achieve a sensitivity that cannot be reached without using semi-automated techniques. As inter-observer agreement increases when using the RST, the need for censoring by multiple observers might then become superfluous.

5. CONCLUSIONS

The use of semi-automated techniques for microbleed detection improves rater performance, results in a higher inter-rater agreement, and significantly reduces the required rating time.

REFERENCES

[1] Greenberg, S. M., Vernooij, M. W., Cordonnier, C., Viswanathan, A., Salman, R. A.-S., Warach, S., Launer, L. J., Buchem, M. A. V., and Breteler, M. M., "Cerebral microbleeds: a guide to detection and interpretation," *The Lancet Neurology* **8**(2), 165 – 174 (2009).

[2] Vernooij, M. W., van der Lugt, A., Ikram, M. A., Wielopolski, P. A., Niessen, W. J., Hofman, A., Krestin, G. P., and Breteler, M., "Prevalence and risk factors of cerebral microbleeds: The rotterdam scan study," *Neurology* **70**(14), 1208–1214 (2008).

[3] Brundel, M., Heringa, S. M., de Bresser, J., Koek, H. L., Zwanenburg, J. J., Kapelle, L. J., Luijten, P. R., and Biessels, G. J., "High prevalence of cerebral microbleeds at 7tesla mri in patients with early alzheimers disease," *Journal of Alzheimers Disease* **31**, 259–263 (2012).

[4] Gregoire, S. M., Chaudhary, U. J., Brown, M. M., Yousry, T. A., Kallis, C., Jager, H. R., and Werring, D. J., "The microbleed anatomical rating scale (mars): Reliability of a tool to map brain microbleeds," *Neurology* **73**(21), 1759–1766 (2009).

[5] Cordonnier, C., Potter, G. M., Jackson, C. A., Doubal, F., Keir, S., Sudlow, C. L., Wardlaw, J. M., and Salman, R. A.-S., "Improving interrater agreement about brain microbleeds: development of the brain observer microbleed scale (bombs)," *Stroke* **40**(1), 94–99 (2009).

[6] de Bresser, J., Brundel, M., Conijn, M. M., van Dillen, J. J., Geerlings, M. I., Viergever, M. A., Luijten, P. R., and Biessels, G. J., "Visual cerebral microbleed detection on 7t mr imaging: Reliability and effects of image processing," *American Journal of Neuroradiology* **in press** (2012).

[7] Barnes, S. R., Haacke, E. M., Ayaz, M., Boikov, A. S., Kirsch, W., and Kido, D., "Semiautomated detection of cerebral microbleeds in magnetic resonance images," *Magnetic Resonance Imaging* **29**(6), 844 – 852 (2011).

 [8] Seghier, M. L., Kolanko, M. A., Leff, A. P., Jger, H. R., Gregoire, S. M., and Werring, D. J., "Microbleed detection using automated segmentation (midas): A new method applicable to standard clinical mr images," *PLoS ONE* **6**, e17547 (03 2011).

 [9] Kuijf, H. J., de Bresser, J., Geerlings, M. I., Conijn, M. M., Viergever, M. A., Biessels, G. J., and Vincken, K. L., "Efficient detection of cerebral microbleeds on 7.0t mr images using the radial symmetry transform," *NeuroImage* **59**(3), 2266 – 2273 (2012).

[10] Simons, P., Algra, A., van de Laak, M., Grobbee, D., and van der Graaf, Y., "Second manifestations of arterial disease (smart) study: Rationale and design," *European Journal of Epidemiology* **15**, 773–781 (1999).

[11] Conijn, M. M. A., Geerlings, M. I., Luijten, P. R., Zwanenburg, J. J., Visser, F., Biessels, G. J., and Hendrikse, J., "Visualization of cerebral microbleeds with dual-echo t2*-weighted magnetic resonance imaging at 7.0 t," *J. Magn. Reson. Imaging* **32**(1), 52–59 (2010).

[12] Conijn, M. M. A., Geerlings, M. I., Biessels, G.-J., Takahara, T., Witkamp, T. D., Zwanenburg, J. J. M., Luijten, P. R., and Hendrikse, J., "Cerebral microbleeds on mr imaging: Comparison between 1.5 and 7t," *AJNR Am J Neuroradiol* **32**(6), 1043–1049 (2011).

[13] Veta, M., Kornegoor, R., Huisman, A., Verschuur-Maes, A., Viergever, M., Pluim, J., and van Diest, P., "Prognostic value of automatically extracted nuclear morphometric features in whole slide images of male breast cancer," *Modern Pathology* **25**(12), 1559–1565 (2012).

[14] Noorda, Y. H., Bartels, L. W., and Pluim, J. P., "Segmentation of the cartilage in the rib cage in 3d mri," in [*Abdominal Imaging. Computational and Clinical Applications*], Yoshida, H., Hawkes, D., and Vannier, M. W., eds., *Lecture Notes in Computer Science* **7601**, 229–237, Springer Berlin Heidelberg (2012).

[15] Loy, G. and Zelinsky, A., "Fast radial symmetry for detecting points of interest," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **25**(8), 959–973 (2003).

[16] Kuijf, H. J., de Bresser, J., Biessels, G.-J., Viergever, M. A., and Vincken, K. L., "Detecting cerebral microbleeds in 7.0 t mr images using the radial symmetry transform," in [*Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on*], 758 –761 (2011).

[17] Ritter, F., Boskamp, T., Homeyer, A., Laue, H., Schwier, M., Link, F., and Peitgen, H.-O., "Medical image analysis: A visual approach," *IEEE Pulse* **2**, 60–70 (Nov.-Dec. 2011).