2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Automatic Assessment Of Hip Effusion From MRI

Abhilash Rakkunedeth Hareendranathan,Yungchan Jin,Banafshe Felfeliyan,Janet Lenore Ronsky,Bashiar Thejeel, Vanessa Quinn-Laurin,Jacob L. Jaremko

Abstract— **Joint effusion is a hallmark of osteoarthritis (OA) associated with stiffness, and may relate to pain, disability, and long-term outcomes. However, it is difficult to quantify accurately. We propose a new Deep Learning (DL) approach for automatic effusion assessment from Magnetic Resonance Imaging (MRI) using volumetric quantification measures (VQM). We developed a new multiplane ensemble convolutional neural network (CNN) approach for 1) localizing bony anatomy and 2) detecting effusion regions. CNNs were trained on femoral head and effusion regions manually segmented from 3856 images (63 patients). Upon validation on a non-overlapping set of 2040 images (34 patients) DL showed high agreement with ground-truth in terms of Dice score (0.85), sensitivity (0.86) and precision (0.83). Agreement of VQM per-patient was high for DL vs experts in term of Intraclass correlation coefficient (ICC)= 0.88[0.80,0.93]. We expect this technique to reduce inter-observer variability in effusion assessment, reducing expert time and potentially improving the quality of OA care.**

Clinical Relevance— **Our technique for automatic assessment of hip MRI can be used for volumetric measurement of effusion. We expect this to reduce variability in OA biomarker assessment and provide more reliable indicators for disease progression.**

I. INTRODUCTION

Osteoarthritis (OA) has high prevalence across various age groups and is the most common disease affecting hip and knee joints[1]–[4]. The economic burden of OA is significant when considering the diminished quality of life and earning capacity of the affected population. Predicted productivity costs of work loss (PCWL) associated with OA are expected to increase from $12 billion to $17.5 billion Canadian Dollars from 2010 to 2031[2].

OA is increasingly recognized to have an inflammatory-like component which is thought to represent a target for therapy [5]–[7]. Effective clinical management requires accurate quantification of features related to inflammation, particularly synovitis and joint effusion. X-rays generally show only structural damage and are poor at visualizing soft tissue features. Assessment of effusion using ultrasound examination is possible but challenging due to the depth and complex structure of the hip joint [8], [9]. MRI examination is a more reliable modality for detecting and measuring effusion. It is also safer compared to X-ray and CT as it does not involve ionizing radiation. However, manual quantification of hip effusion is tedious and user dependent [10]. Likely as a result, correlations between clinical outcomes and hip effusion are moderate at best [8] .

The clinical utility of effusion measurement is limited by problems with obtaining accurate measurements by conventional methods. For example, quantification using single measurements [8], [11] or semi-quantitative techniques [12], [13] might not accurately measure the volume of joint effusion, especially in cases where structural damage has already occurred. Voxel based volumetric quantification measures(VQM) have been proposed to address these issues. However, in practice, measurement of VQM is cumbersome as it involves complex measurements over multiple slices and requires many hours of training. Hence automatic approaches that measure effusion volume are being explored.

Earlier attempts to automatically quantify knee joint effusion from MRI used multiple thresholding[14]. In general, such threshold-based approaches are easily affected by image quality and the presence of noise artifacts as it relies on pixel intensity values. Another limitation of thresholding is that conditions like bone marrow edema (BME) also cause high intensity regions in the images that might be wrongly interpreted as effusion. In recent years, data driven techniques like Deep Learning (DL) have been used to successfully address these issues in similar MRI post processing applications. DL has been applied in applications like assessment of knee and hip OA, femoral head osteonecrosis and elbow joint effusion from radiographs [15]. Automatic assessment of effusion from MRI is relatively rare. A DL technique using a Siamese network was developed to identify knee pain based on structural features seen in MRI[16]. This approach gave 86% accuracy in correctly detecting subjects with knee pain. Synovitis or effusion was seen in most images identified by the network. Unlike earlier approaches, we propose a new technique that directly detects regions with effusion and measures VQM to quantify the extent of damage. Using a two-stage DL approach, we first localize the femoral head region from an MRI hip image and then identify effusion in the surrounding regions. To the best of our knowledge, this is the first DL based technique for automatic VQM quantification.

II. PROPOSED METHODOLOGY

We trained separate semantic segmentation networks for 1) localization of femoral head and 2) multiplane detection of hip effusion regions. As shown in Fig 1, the localization network uses a single MRI frame located at the center of the image sweep and identifies left and right femoral heads. The location of the femoral head is used to identify rectangular regions in all slices of the input image *I*. We resampled this region into image sequences in three orthogonal planes I_{cor} ,

* Abhilash R Hareendranathan (email: hareendr@ualberta.ca), Yungchan Jin, ,Bashiar Thejeel, Vanessa Quinn-Laurin and Jacob L. Jaremko are with Department of Radiology and Diagnostic Imaging,University of Alberta Canada. Banafshe Felfeliyan and Janet Lenore Ronsky are with University of Calgary, Canada.

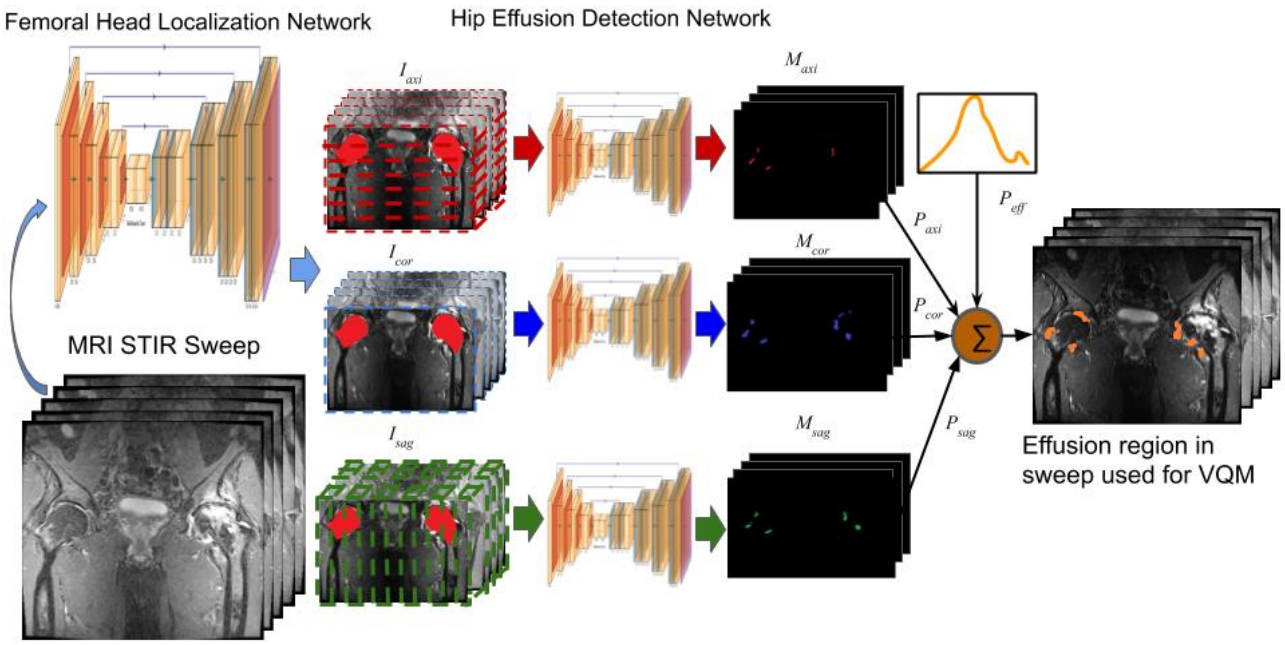

Fig. 1. Overview of the proposed segmentation technique for assessment of VQM. The region identified using the Femoral Head localization network is sliced along the sagittal, axial and coronal planes.

I_{sag} and I_{axi} and trained separate CNNs for each plane to obtain effusion image masks M_{cor} , M_{sag} and M_{axi}. As a saliency measure of each mask we compared the similarity of predicted pixel distribution against the overall effusion pixel intensity distribution P_{eff}. We use the discrete Bhattacharyya Coefficient (BC) for similarity which measures the approximate extent of overlap between the distributions. In general, having small values for bin size (X) tends to overestimate the similarity and very high values underestimate the similarity. We defined the number of bins X= 26. Hence the BC for each orthogonal image sequence can be written as:

$$BC_{cor} = \sum_{x \epsilon X} \sqrt{P_{eff}^{x} P_{cor}^{x}} \quad (1)$$

$$BC_{sag} = \sum_{x \epsilon X} \sqrt{P_{eff}^{x} P_{sag}^{x}} \quad (2)$$

$$BC_{axi} = \sum_{x \epsilon X} \sqrt{P_{eff}^{x} P_{axi}^{x}} \quad (3)$$

We combine the effusion masks based on the values of the respective Bhattacharyya Coefficient as

$$M = \frac{M_{cor} BC_{cor} + M_{sag} BC_{sag} + M_{axi} BC_{axi}}{BC_{cor} + BC_{sag} + BC_{axi}} \quad (4)$$

A. MRI Image Dataset

We retrospectively analyzed MRI images from 97 patients who met the American College of Rheumatology clinical classification criteria for reporting OA[20]. Participants were recruited from patients presenting at radiology clinics for hip joint intra-articular steroid injection. Routine clinical, functional, and physical examination was performed along with MRI scanning at baseline and eight weeks after the injection [3],[21]. Images of both hips were acquired using a 1.5T Siemens Symphony scanner with coronal short-tau inversion recovery (STIR) sequences at slice thickness 3 mm, FOV 349x349, matrix 384x384, flip angle 150 and TR/TE/TI 5,710/53/160 ms).

B. Manual Labeling

Effusion regions in all images were labeled by 2 expert radiologists using an interactive software developed inhouse. The region of overlap in 3856 images (from 63 patients) was used as ground truth labels to train the effusion detection network. In a small subset of 150 images, a medical student also identified the boundaries of the left and right femoral head which was used to train the localization network.

C. Validation

We compared effusions regions detected by DL (P) to ground truth (GT) in terms of dice score, sensitivity and precision which are defined below:

$Dice\ Score = |P \cap GT| / |P \cup GT|,$

$Sensitivity = |P \cap GT| / |GT|$

$Precision = |P \cap GT| / |P|$ where $|\ |$ is the area

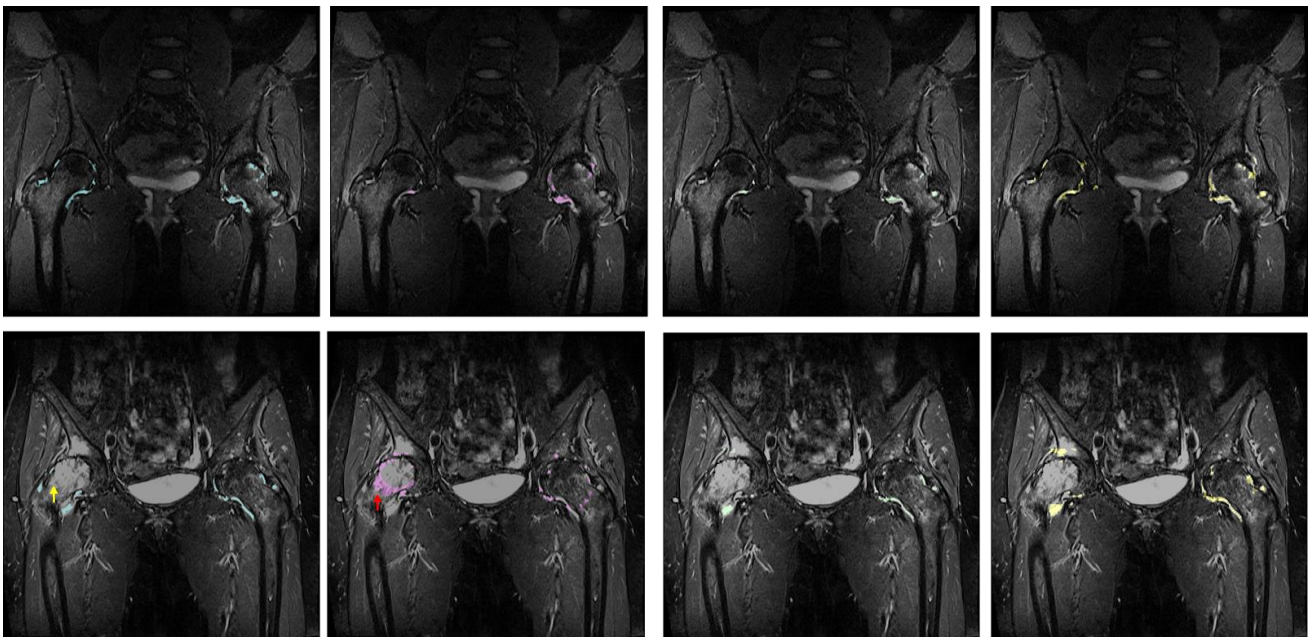

Fig. 2. Effusion regions detected on slices of STIR Coronal MRI. Column1: Ground truth masks segmented by experts, Column 2: Effusion regions detected from a thresholding approach after localizing the femoral head, Column 3: Effusion regions detected by the single CNN approach. Column 4: Effusion regions detected by the multiplane CNN after combining the masks

As a baseline, we also implemented 1) an Otsu thresholding-based approach [22] for detecting effusion within the region detected by the localization network 2) an effusion network using a single CNN. Using a paired student's t-test we compared the dice score of our multiplane ensemble and single plane approaches vs thresholding. Agreement of VQM with corresponding measurements from experts was assessed using ICC3k (two-way mixed average score) first between the experts (refer Table 2, row 4) and then with the DL based approaches (refer Table 2, row 1-2) and thresholding approach (refer Table 2, row 3) as a third reader.

III. Results

We validated our DL approach on 2040 MRI images (derived from 34 patients) that were not included in the training. Examples of effusion regions detected by our technique along with the corresponding ground truth segmentations are shown in Fig 2. Otsu thresholding (column 2 in Fig 2) within the region detected by the localization network wrongly categorized high pixel intensity (such as bone marrow edema) regions as effusion. The single CNN approach (column 3 in Fig 2) under segmented the effusion region when compared to the multiplane ensemble approach (column 4 in Fig 2). Masks detected using both DL models (single and multiplane) were closer to ground truth and correctly excluded bone edema (at the right femoral neck in Fig 2), unlike the thresholding approach.

The overlapping region between manual segmentations by experts was considered as ground truth. As summarized in Table 1, our DL based technique gave higher values for dice score, sensitivity, and precision. The difference in dice score of DL vs thresholding was statistically significant (p-value < 0.001).We also calculated the effusion volume based on the regions detected from DL. Each patient was scanned twice, and the effusion values were computed separately for each hip. Table 2 summarizes agreement of these volumes with corresponding measurements first between the experts (refer Table 2, row 4) and then with the DL based approaches and thresholding approach (refer Table 2, row 3) in terms of ICC.

TABLE I. Comparison of Dice Score, Sensitivity and Precision of the Multiplane DL, Single Plane DL and Thresholding.

Technique	Dice score	Sensitivity	Precision
Multiplane CNN	0.85	0.86	0.83
Single plane CNN	0.83	0.85	0.81
Thresholding	0.72	0.78	0.65

TABLE II. ICC values between readers and with automatic technique (multiplane DL, single plane DL or thresholding) as a third reader. The 95% confidence intervals are shown in brackets.

Technique	ICC patient	ICC right hip	ICC left hip
Multiplane CNN	0.88 [0.80,0.93]	0.87 [0.75,0.93]	0.89 [0.80,0.91]
Single plane CNN	0.86 [0.76, 0.92]	0.86 [0.75, 0.91]	0.87 [0.77, 0.92]

Threshold	0.80 [0.61,0.91]	0.75 [0.55, 0.86]	0.83 [0.71, 0.91]
Manual	0.97 [0.93, 0.98]	0.97 [0.95, 0.99]	0.96 [0.92, 0.98]

IV. DISCUSSION

We proposed a fully automated technique for assessment of hip effusion volume from MRI. Manual segmentation by two expert radiologists were used to establish the ground truth. Using a multiplane ensemble approach, we aimed to improve the generalization and remove spurious artifacts appearing in single planes. Our approach was highly accurate (Dice score = 0.85) with 0.86 sensitivity and 0.83 precision in detecting effusions. This was significantly higher than the accuracy of our thresholding-based approach which is similar to earlier works[14]. A direct comparison to earlier approaches might not be possible due to the lack of open datasets that evaluate MRI hip effusion. Both single plane and multiplane ensemble DL approaches differentiated and excluded regions with BME that were misclassified by the thresholding approach. Compared to the single plane approach, the multiplane CNN gave higher overlap with ground truth and was more sensitive in identifying effusion regions. As a third reader, our method showed high agreement with the expert human readers ICC = 0.88. Our agreement was less than the agreement between experts (> 0.90) which was expected as both experts in this study were radiologists with many years of experience in reading MRI scans representing an idealized scenario. We expect to see higher variability when measurements are performed by non-experts for example, trainees or clinicians with lesser experience.

Our method saves expert time significantly when used in standalone mode. On average, our technique takes < 1 sec per scan (~ 20 MRI images) when executed on V100 GPU. While performing manual segmentation using our interactive tool the radiologist spent ~3 minutes per MRI scan. The AI technique could also be used as an aid for image analysis in which the radiologist would have the option to add or remove specific information from the AI result.

One study has limitations. Firstly, the ground truth segmentation is based on human assessment. Although very high agreement (ICC > 0.95) between human experts indicates limited inter-observer variability for VQM, this method is tedious and not applicable for general use in the hundreds of thousands of hip MRI performed annually. Secondly our approach has been trained on data from a single MRI scanner. MRI intensities values are dependent on the type of scanner and MRI sequence[23-25]. Ideally the bin number and size would have to be selected based on the intensity ranges present in each sequence. Hence more extensive validation using images collected from multiple centers would be required to assess the generalizability of our technique. As an extension of our study, we plan to adapt our DL technique for similar image-based assessment of knee MRI scans. To further increase AI accuracy to that of human experts, on the AI front the UNet model used could be replaced with other semantic segmentation models based on the complexity and size of the MRI dataset. As future work we also plan to validate our technique on a larger hip MRI dataset and grade the extent of damage-based effusion volume.

V. CONCLUSION

We have proposed and evaluated a Deep Learning framework for volumetric quantification of hip effusion from MRI. Our approach is fully automatic (hence non-subjective), fast and highly accurate in detecting hip effusion from MRI slices. We expect this technique to enhance the use of volumetric measures for assessment and treatment of osteoarthritis.

VI. ACKNOWLEDGMENT

Jacob Jaremko is supported by the AHS Chair in Diagnostic Imaging and his academic time is made available by Medical Imaging Consultants (MIC), Edmonton, Canada. We thank NSERC for the research funding that supported Dr Janet Ronsky's research. We acknowledge the support of Compute Canada in providing us with high power GPUs for deep learning models.

REFERENCES

[1] B. Sharif et al., "Projecting the direct cost burden of osteoarthritis in Canada using a microsimulation model," Osteoarthritis Cartilage, vol. 23, no. 10, pp. 1654–1663, Oct. 2015.

[2] B. Sharif, R. Garner, D. Hennessy, C. Sanmartin, W. M. Flanagan, and D. A. Marshall, "Productivity costs of work loss associated with osteoarthritis in Canada from 2010 to 2031," Osteoarthritis Cartilage, vol. 25, no. 2, pp. 249–258, Feb. 2017.

[3] A. Maetzel, L. C. Li, J. Pencharz, G. Tomlinson, C. Bombardier, and Community Hypertension and Arthritis Project Study Team, "The economic burden associated with osteoarthritis, rheumatoid arthritis, and hypertension: a comparative study," Ann. Rheum. Dis., vol. 63, no. 4, pp. 395–401, Apr. 2004.

[4] L. Murphy and C. G. Helmick, "The impact of osteoarthritis in the United States: a population-health perspective," Am. J. Nurs., vol. 112, no. 3 Suppl 1, pp. S13–9, Mar. 2012.

[5] D. Loeuille et al., "Macroscopic and microscopic features of synovial membrane inflammation in the osteoarthritic knee: correlating magnetic resonance imaging findings with disease severity," Arthritis Rheum., vol. 52, no. 11, pp. 3492–3501, Nov. 2005.

[6] F. Fernandez-Madrid, R. L. Karvonen, R. A. Teitge, P. R. Miller, T. An, and W. G. Negendank, "Synovial thickening detected by MR imaging in osteoarthritis of the knee confirmed by biopsy as synovitis," Magn. Reson. Imaging, vol. 13, no. 2, pp. 177–183, 1995.

[7] I. Atukorala et al., "Synovitis in knee osteoarthritis: a precursor of disease?," Ann. Rheum. Dis., vol. 75, no. 2, pp. 390–395, Feb. 2016.

[8] K. J. D. Steer et al., "Can effusion-synovitis measured on ultrasound or MRI predict response to intra-articular steroid injection in hip osteoarthritis?," Skeletal Radiol., vol. 48, no. 2, pp. 227–237, Feb. 2019.

[9] S. N. Sudula, "Imaging the hip joint in osteoarthritis: A place for ultrasound?," Ultrasound, vol. 24, no. 2, pp. 111–118, May 2016.

[10] P. F. Weiss et al., "Feasibility and Reliability of the Spondyloarthritis Research Consortium of Canada Sacroiliac Joint Structural Score in Children," The Journal of Rheumatology, vol. 45, no. 10. pp. 1411–1417, 2018, doi: 10.3899/jrheum.171329.

[11] S. G. Moss et al., "Hip joint fluid: detection and distribution at MR imaging and US with cadaveric correlation," Radiology, vol. 208, no. 1, pp. 43–48, Jul. 1998.

[12] F. W. Roemer et al., “Hip Osteoarthritis MRI Scoring System (HOAMS): reliability and associations with radiographic and clinical findings,” Osteoarthritis Cartilage, vol. 19, no. 8, pp. 946–962, Aug. 2011.
[13] N. Deseyne et al., “Hip Inflammation MRI Scoring System (HIMRISS) to predict response to hyaluronic acid injection in hip osteoarthritis,” Joint Bone Spine, vol. 85, no. 4, pp. 475–480, Jul. 2018.
[14] W. Li et al., “Fully automated system for the quantification of human osteoarthritic knee joint effusion volume using magnetic resonance imaging,” Arthritis Res. Ther., vol. 12, no. 5, p. R173, Sep. 2010.
[15] R. Kijowski, F. Liu, F. Caliva, and V. Pedoia, “Deep learning for lesion detection, progression, and prediction of musculoskeletal disease,” J. Magn. Reson. Imaging, Nov. 2019, doi: 10.1002/jmri.27001.
[16] G. H. Chang, D. T. Felson, S. Qiu, A. Guermazi, T. D. Capellini, and V. B. Kolachalama, “Assessment of knee pain from MR imaging using a convolutional Siamese network,” Eur. Radiol., vol. 30, no. 6, pp. 3538–3548, Jun. 2020.
[17] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” in Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, 2015, pp. 234–241.
[18] S. Ruder, “An overview of gradient descent optimization algorithms,” arXiv [cs.LG], Sep. 15, 2016.
[19] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” arXiv [cs.LG], Dec. 22, 2014.
[20] R. Altman et al., “The American College of Rheumatology criteria for the classification and reporting of osteoarthritis of the hip,” Arthritis Rheum., vol. 34, no. 5, pp. 505–514, May 1991.
[21] V. Quinn-Laurin, B. Thejeel, N. A. Chauvin, T. G. Brandon, P. F. Weiss, and J. L. Jaremko, “Normal hip joint fluid volumes in healthy children of different ages, based on MRI volumetric quantitative measurement,” Pediatr. Radiol., vol. 50, no. 11, pp. 1587–1593, Oct. 2020.
[22] N. Otsu, “A threshold selection method from gray-level histograms,” IEEE Trans. Syst. Man Cybern., vol. 9, no. 1, pp. 62–66, 1979.
[23] L. G. Nyúl, J. K. Udupa, and X. Zhang, “New variants of a method of MRI scale standardization,” IEEE Trans. Med. Imaging, vol. 19, no. 2, pp. 143–150, Feb. 2000.
[24] A. Madabhushi, J. K. Udupa, and G. Moonis, “Comparing MR image intensity standardization against tissue characterizability of magnetization transfer ratio imaging,” J. Magn. Reson. Imaging, vol. 24, no. 3, pp. 667–675, Sep. 2006.
[25] N. Robitaille, A. Mouiha, B. Crépeault, F. Valdivia, S. Duchesne, and The Alzheimer’s Disease Neuroimaging Initiative, “Tissue-based MRI intensity standardization: application to multicentric datasets,” Int. J. Biomed. Imaging, vol. 2012, p. 347120, May 2012.