QEI-Net: A Deep learning-based automated quality evaluation index for ASL CBF Maps

Xavier Beltran Urbano[1,2], Manuel Taso[3], Ilya Nasrallah[1], John A. Detre[1,2], Ze Wang[4], and Sudipto Dolui[1]
[1]Department of Radiology, University of Pennsylvania, Philadelphia, PA, United States, [2]Department of Neurology, University of Pennsylvania, Philadelphia, PA, United States, [3]Siemens Medical Solutions, Philadelphia, PA, United States, [4]University of Maryland School of Medicine, Maryland, MD, United States

Synopsis

Motivation: Arterial Spin Labeling (ASL) cerebral blood flow (CBF) maps can be noisy, which can bias statistical results. Quality control (QC) by visual inspection is subjective and time-consuming. Automated objective QC addresses these limitations, but previous methods lacked consistency across datasets.

Goal(s): Develop a deep learning model (QEI-Net) to derive a robust quality evaluation index (QEI) for ASL CBF maps.

Approach: We trained QEI-Net on manually rated multi-protocol ASL datasets and compared QEI-Net with both manual ratings and the previous state-of-the-art method.

Results: QEI-Net strongly correlated with manual ratings and outperformed the reference approach. This provides reliable and reproducible assessments suitable for large-scale studies.

Impact: We propose QEI-Net, a deep learning based automated quality evaluation method for Arterial Spin Labeling (ASL) derived cerebral blood flow images. QEI-Net can enable consistent and reproducible quality assessments and reduce the time burden and subjectivity in studies using ASL.

Introduction

Arterial Spin Labeling (ASL) MRI is widely used to measure cerebral blood flow (CBF)[1,2] for its non-invasiveness, non-radiation, and cost-effectiveness, especially in large-scale research studies. However, ASL data is prone to noise and artifacts, necessitating robust quality control (QC). Traditional manual QC methods are laborious and subject to bias. We previously proposed an automated quality evaluation index[3] (hereafter referred to as QEI_{basic}), which was based on fitting three pre-determined features on a small set of data obtained with non-background suppressed 2D protocols. While QEI_{basic} showed reliable agreement with manual rating on average, there are disagreements in individual cases and a lack of generalization across different imaging protocols. Here, we proposed QEI-Net, a deep learning (DL) model to derive a QEI (hereafter referred to as QEI_{DL}), which was expected to provide more robust results. We also aimed to derive the region of artifacts in the image. Compared to QEI_{basic}, the QEI_{DL} used a wide variety of scanning protocols for training and used the whole image instead of predetermined features.

Method

The QEI_{DL} provides a numerical value between 0 and 1, where a higher value indicates better quality. We utilized N=150 ASL scans acquired with different protocols on Siemens 3T scanner as detailed in Table 1 in Figure 1. Three raters with extensive experience working with ASL data visually rated each dataset on a scale between 0 and 1 following some specific guidelines (see Figure 2A). Thereafter we divided the data randomly for training and validation (N=120) and testing (N=30). To assess the robustness of the method, an independent fourth rater also rated the test set.

The preprocessing pipeline involved deriving CBF from the raw ASL data, registration to MNI space, down-sampling the images to 64x64x64 size, intensity clipping to a range of [-80, 80], and scaling to [-1,1] range. The proposed network consists of four convolutional blocks, each with residual connections (Figure 2B). Max pooling layers with a size of 2 were applied after the first three blocks. The network ends with three fully connected layers followed by a sigmoid-activated output neuron. A dropout rate of 20% was applied after the fully connected layers. Training used the Adam optimizer (initial learning rate 1e-4) with a batch size of 16, and early stopping (patience of 30 epochs) to prevent overfitting. Mean Squared Error (MSE) was used as the loss function of this approach. The model was trained using a 5-fold cross-validation strategy and the five model-predictions were averaged to provide a more robust prediction on the test set. Region of artifacts maps were empirically defined as the GradCAM[10] heatmap from the second convolutional layer of the first convolutional block.

QEI_{DL} was validated and compared with QEI_{basic} by comparing them with the average manual ratings using Pearson correlation coefficient and the squared errors as the performance indices. Next, we binarized the ratings as unacceptable and acceptable using a threshold of 0.25 and computed the area under the receiver operating characteristic (ROC-AUC) curve of this classification. Additionally, we computed the Youden index as a recommended threshold to be used to discard unacceptable quality CBF maps.

Results

The correlation between the ratings of raters 1 and 2 was 0.88, between 1 and 3 was 0.80 and between 2 and 3 was 0.79 ($p<0.0001$ in each case). Table 2 in Figure 1 presents a comparison between QEI_{basic} and QEI_{DL}. The correlation between QEI_{DL} with the average manual ratings was 0.92, which was comparable to the inter-rater agreement and was significantly higher ($p=0.02$) than the correlation between QEI_{basic} and the average rating ($R=0.81$). QEI_{DL} correlated strongly with the fourth rater ($R=0.93$) whose rating was not used to train the model. Figure 3A shows scatter plots between the individual ratings and the QEI_{basic} and QEI_{DL} predictions. The squared errors obtained with QEI_{basic} were significantly higher ($p=0.002$) than QEI_{DL} (Figure 3B). Both methods provided comparable ROC-AUC values (Figure 3C). Figure 4 shows examples of artifactual images and the corresponding heatmaps indicating the region of artifacts. Figure 5 shows examples where both QEI_{basic} and QEI_{DL} provided similar values and agreed with the manual ratings (A and B), as well as cases where QEI_{DL} provided better agreement with the manual ratings than QEI_{basic} (C and D).

Discussion and Conclusion

QEI-Net demonstrated superior performance compared to the current state-of-the-art automated method[3] for predicting the quality of ASL CBF maps, achieving lower SE metrics and better correlation with expert ratings. Future work will include training the model with larger datasets that will include ASL from wider type of protocols and scanner vendors.

Acknowledgements

NIH grant R21AG080518, R01AG081693 and R01EB031080.

References

1. Detre JA, Leigh JS, Williams DS, Koretsky AP. Perfusion imaging. Magnetic Resonance in Medicine 1992;23(1):37-45.
2. Alsop DC, Detre JA, Golay X, et al. Recommended implementation of arterial spin-labeled perfusion MRI for clinical applications: A consensus of the ISMRM perfusion study group and the European consortium for ASL in dementia. Magnetic Resonance in Medicine 2015;73(1):102-116.
3. Dolui S, Wang Z, Wolf RL, et al. Automated Quality Evaluation Index for Arterial Spin Labeling Derived Cerebral Blood Flow Maps. Journal of magnetic resonance imaging : JMRI 2024.
4. Wang Z, Das SR, Xie SX, et al. Arterial spin labeled MRI in prodromal Alzheimer's disease: A multi-site study. NeuroImage Clinical 2013;2:630-636.
5. Austin TR, Nasrallah IM, Erus G, et al. Association of Brain Volumes and White Matter Injury With Race, Ethnicity, and Cardiovascular Risk Factors: The Multi-Ethnic Study of Atherosclerosis. Journal of the American Heart Association 2022;11(7).
6. Dolui S, Detre JA, Gaussoin SA, et al. Association of Intensive vs Standard Blood Pressure Control With Cerebral Blood Flow: Secondary Analysis of the SPRINT MIND Randomized Clinical Trial. JAMA neurology 2022;79(4):380-389.
7. Dolui S, Wang Z, Wang DJJ, et al. Comparison of non-invasive MRI measurements of cerebral blood flow in a large multisite cohort. Journal of Cerebral Blood Flow and Metabolism 2016;36(7):1244-1256.
8. Dolui S, Tisdall D, Vidorreta M, et al. Characterizing a perfusion-based periventricular small vessel region of interest. NeuroImage Clinical 2019;23:101897.
9. Sadaghiani S, Tackett W, Tisdall MD, Detre JA, Dolui S. Reliability of Periventricular White Matter Cerebral Blood Flow using Different ASL protocols. Proceedings of the International Society of Magnetic Resonance in Medicine. London; 2022.
10. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. IEEE International Conference on Computer Vision (ICCV). Venice, Italy; 2017. p. 618-626.

Figures

Table 1: Information of the different datasets used in this work.

Dataset	Protocol	Sample Size
Alzheimer's Disease Neuroimaging Initiative (ADNI)[4]	2D PASL	43
Multi-Ethnic Study of Atherosclerosis (MESA)[5]	3D BS PCASL	28
Systolic Blood Pressure Intervention Trial (SPRINT)[6]	2D PCASL	31
Coronary Artery Risk Development in Young Adults (CARDIA)[7]	2D PCASL	19
National Alzheimer's Coordinating Center (NACC)[8]	3D BS PCASL	24
Vascular Contributions to Cognitive Impairment and Dementia (VCID)[9]	3D BS PCASL	5
		150

Abbreviations: PASL: Pulsed ASL, PCASL: Pseudo-Continuous labeling ASL, BS: Background Suppression.

Table 2: Comparison of QEI_{basic} (Dolui et al [3]) and QEI_{DL} (proposed) on the test set.

Approach	PC	MSE ± SE std	AUC	Youden Index
QEI_{basic}	0.81	0.034 ± 0.039	0.85	0.60
QEI_{DL}	**0.92**	**0.009 ± 0.011**	**0.88**	0.32

Abbreviations: PC: Pearson Correlation between the prediction and the ratings, MSE: Mean Squared Error, SE: Squared Error, AUC: Area Under the Curve.

Figure 1: Table 1 lists the different datasets utilized in this work. Table 2 compares the QEI_{basic} (Dolui et al.[3]) and QEI_{DL} (proposed) approaches on the test set.

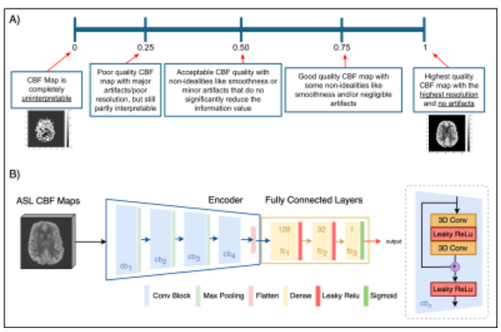

Figure 2: A) Rating strategy guidelines provided to the expert raters to annotate the dataset. B) Overview of the QEI-Net architecture which consists of different convolutional blocks followed by three fully connected layers. The block in the right shows the detailed schematic of each convolutional block, which utilizes residual connections. $cb_{1\text{-}4}$ denotes the corresponding convolutional block numbers.

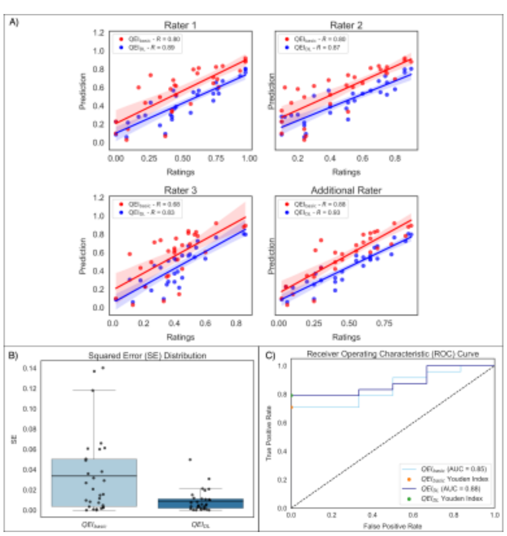

Figure 3A: Correlation between the ratings from individual expert raters and QEI_{basic} (Dolui et al.[3]) and QEI_{DL} (proposed) predictions on the test set. Figure 3B and 3C: Performance comparison between QEI_{basic} and QEI_{DL} on the test set. Figure 3B: Squared Error (SE) distribution for both methods. Figure 3C: Receiver Operating Characteristic (ROC) curves for both models, including their corresponding Youden Index points.

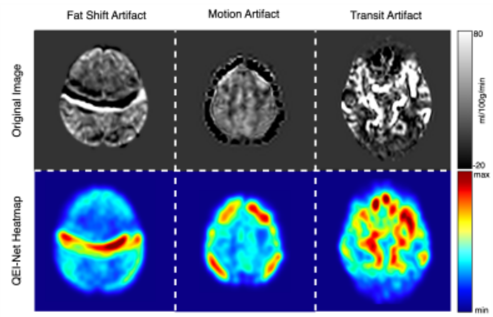

Figure 4: ASL CBF maps contaminated with different types of artifacts and the corresponding heatmaps generated from the QEI-Net showing the location of the artifacts.

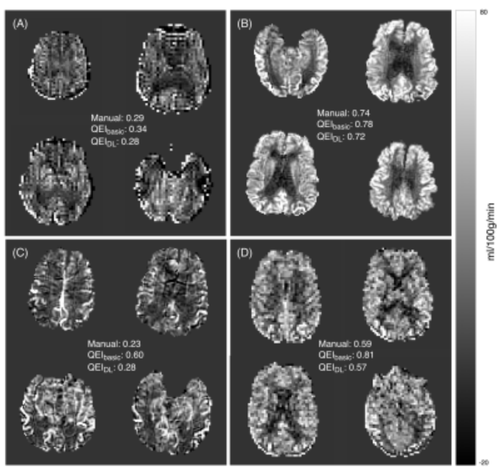

Figure 5: Examples of CBF maps with the corresponding manual ratings and the automated quality evaluation indices derived using QEI_{basic} (Dolui et al.[3]) and QEI_{DL} (proposed). (A) and (B) show examples where both the automated methods agreed with manual ratings while (C) and (D) show examples where the QEI_{DL} provided better agreement than QEI_{basic}.