



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

Xavier Beltran Urbano*, Sudipto Dolui¹, John A Detre¹

¹*Detre Lab, Department of Neurology, University of Pennsylvania, USA*

Abstract

Keywords: Arterial Spin Labeled, Alzheimer Disease, Deep Learning, AI, Regression, QEI-Net, CBF, MRI

1. Introduction

The brain is one of the most highly perfused organs in the body, utilizing approximately 15% of the cardiac output and 20% of the total body oxygen (Jain et al. (2010)). Cerebral blood flow (CBF) is classically defined as the volume of blood flowing through a specific region of the brain tissue per unit time and is expressed in units of milliliters of blood per 100 gram of brain tissue per unit time (unit: ml/100g/min). It is

an important physiological quantity of the cerebrovascular health and provides an important biomarker for the latter. Changes in CBF correlate with various indicators of cerebrovascular disease, including white matter hyperintensities (Bernbaum et al. (2015)) and cerebral microbleeds (Gregg et al. (2015)). Additionally, it also serves as a biomarker of functional neurodegeneration due to its strong association with neural activity (Dolui et al. (2017a)), and therefore can potentially replace glucose metabolism measurements obtained using 18F-Fluorodeoxyglucose Positron Emission Tomography (18F-FDG-PET) (Dolui et al. (2020)). CBF is associated with incidence and severity of dementia (Dolui

*Corresponding author

Email address: xavibeltranurbano00@gmail.com

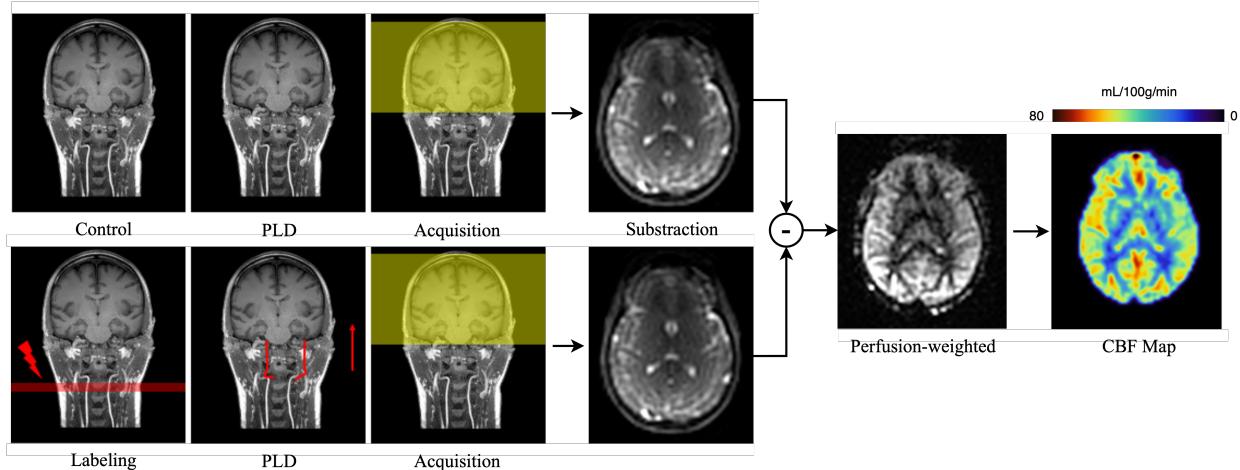


Figure 1: Sequential workflow for ASL CBF Maps Acquisition. This diagram delineates the procedural stages, beginning with the acquisition of control images, followed by the application of labeling and post-labeling delay (PLD). Subsequent subtraction generates the perfusion-weighted images, which are then utilised to create the detailed CBF maps.

et al. (2020); Dolui et al. (2017a); Binnewijzend et al. (2013); Wolk and Detre (2012)) and has been shown to be one of the earliest biomarkers to change in the Alzheimer’s Disease continuum (Iturria-Medina et al. (2016); Fazlollahi et al. (2020)). Moreover, CBF is potentially modifiable and hence is associated with treatment response (De La Torre (2013); Dolui et al. (2022)). Consequently, CBF measurement is considered very important in studies on healthy aging, cerebrovascular and neurodegenerative disease.

1.1. Classical methods of measuring CBF

Classical CBF is measured using a “diffusible” tracer that exchanges from the blood compartment to the tissue compartment, allowing CBF in ml/100g/min to be measured directly. The first CBF measurements in humans were made by Kety and Schmidt (Kety and Schmidt (1945)) by monitoring arteriovenous differences in nitrous oxide. The current “gold-standard” for CBF imaging in humans is ¹⁵O-Positron Emission Tomography (PET) scanning (Zhang et al. (2014); Herscovitch et al. (1983)), which utilizes radioactively labeled water as a perfusion tracer. Other diffusible tracer approaches used to measure CBF in humans include radioactive ¹³³Xenon (Lassen et al. (1981)) and stable xenon computed tomography (CT) (Yonas et al. (1991)). Related methods include accumulative radioactive tracers with SPECT scanning, though agreement of these methods with ¹⁵O-PET is sub-optimal (Ito et al. (2006)), and methods that used intravascular tracers such as perfusion CT [19] and dynamic susceptibility contrast (DSC) MRI (Rempp et al. (1994)). Intravascular tracer methods do not measure CBF directly but allow CBF to be inferred. All these methods require administration of an exogenous tracer and exposure to ionizing radiation. Hence, they are at least somewhat invasive and can be

difficult to administer in clinically vulnerable population groups, including the elderly, infants, and individuals with renal impairments. Moreover, using such methods to track CBF changes in healthy aging and in drug studies can be problematic as these studies require serial measurements with repeated exposure to tracers or ionizing radiation and associate costs.

1.2. ASL MRI Data Acquisition

The acquisition of ASL MRI data involves magnetically labeling inflowing protons of proximal arterial blood water. For brain perfusion, labeling typically occurs in the neck where blood flows through the internal carotid and the vertebral arteries that supply blood to the brain (see Figure 1). After waiting for a brief period (post-labeling delay) to allow the flow of the labeled blood to reach brain tissue, a brain MRI (labeled image) is acquired. Thereafter a “control” brain image is obtained with a sham labeling procedure that does not magnetically label blood water. The difference between the control and label image is proportional to the cerebral perfusion which is converted to absolute CBF quantification using a proton density image and with appropriate models and assumptions (Alsop et al. (2015); Buxton et al. (1998)). The control-label difference is a small percentage of the background signal, which results in a low signal to noise ratio (SNR) in the CBF images. Additionally, subject motion, sub-optimal choice of imaging parameters, and other non-idealities inherent to MRI scanners can lead to severe artifacts (Dolui et al. (2017b)) (see Figure 2), potentially impacting subsequent statistical analysis. This can be mitigated by averaging multiple control-label pairs, using advanced signal processing strategies or using background suppression (BS) techniques which increases the difference image by 3-10 times (Dolui et al.

(2019);Maleki et al. (2012);Ye et al. (2000b)). Nevertheless, a noticeable amount of artifact might remain in the resulting CBF image.

1.3. ASL Labeling Methods

Ever since its establishment in 1992, several ASL protocols have been devised and used which primarily differ in labeling and signal readout strategies. The classical method invented in 1992, which was referred to as Continuous ASL (CASL) (Detre et al. (1992)), continuously saturates or inverts arterial blood water at the neck for several seconds. However, modern human MRI scanners utilizing whole-body radiofrequency (RF) coils are not capable of continuous RF excitation. Pulsed ASL (PASL) instantaneously labels a thick slab in the neck, and is compatible with body RF excitation, though the method suffers from lower signal to noise ratio (SNR) compared to CASL. The current recommended labeling strategy (see Figure 3) is pseudo-continuous labeling (PCASL), which uses a series of short RF pulses to mimic continuous labeling. ASL type can also vary based on the duration of the post labeling delay – a longer post labeling delay can ensure delivery of the labeled blood to the brain tissue, though at the expense of reduced SNR. A series of ASL images acquired with different labeling and/or PLDs can also be combined to obtain a CBF map. Finally, ASL can vary based on the type of image acquisition. Echo-planar imaging (EPI) was initially be the preferred choice because of speed and sensitivity, though it is being slowly replaced by 3D imaging (GRASE or SPIRAL) optimally combined with background suppression of static brain tissue. Notably, several other variants of ASL exists; for example, velocity selective ASL (VASL) labels the arterial blood water close to the imaging site instead of the neck.

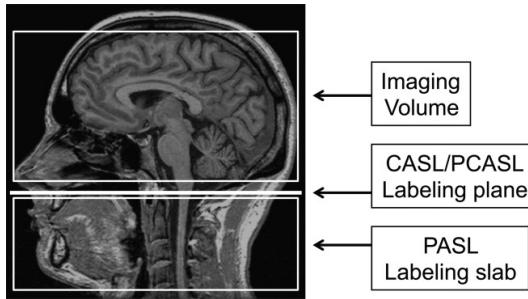


Figure 3: Schematic diagram of imaging and labeling regions for CASL/PCASL and PASL. In CASL/PCASL, labeling occurs as blood flows through a single labeling plane, while in PASL, a slab of tissue, including arterial blood, is labeled (Alsop et al., 2015).

1.4. Artifacts in ASL MRI and the need of an automated quality evaluation index (QEI)

In recent years, ASL has gained popularity among perfusion imaging modalities for its use in research set-

tings, largely due to its potential as a biomarker of cerebrovascular health and brain function, and its ability to be acquired in routine MRI settings. Despite recent advancements in improving the quality of ASL images, the resulting CBF maps can still be contaminated by artifacts. The most significant source of artifact is physiological noise due to motion, particularly in non-compliant subjects or in patients who have difficulty staying still during the scan. Because the control/label difference represents only a small percentage of the background signal, any variability in background signal due to motion can dominate the difference signal leading to large errors that are often not removed during averaging. Retrospective motion correction techniques are generally used to account for bulk motion, but such techniques cannot correct for variation in intensities occurring during the image readout (Kety and Schmidt (1945); Power et al. (2012)). Motion effects are less visible, though still present, in acquisitions using background suppression (BS) of static signal (Fernandez-Seara et al. (2005); Maleki et al. (2012); Ye et al. (2000a)). Artifacts can also result from incorrect or suboptimal choice of acquisition parameters. For example, an insufficiently long PLD results in labeled blood remaining in large arteries rather than in the microvasculature or parenchyma, an effect known as transit time artifact, that affects both BS and non-BS acquisitions. Other problems inherent to MR imaging such as thermal noise, chemical shift artifact, and clipping of signal can produce errors and artifacts in the resulting CBF maps. For clinical research, another concern is that the number of corrupted ASL CBF images may increase with disease severity as previously found in the AD continuum (Moonen et al. (2020)), making quality assurance (QA) a more prominent need in these clinical applications.

Because of potential artifacts in the ASL derived CBF maps, QA is critical for clinical research of ASL MRI to exclude CBF maps of poor quality that can reduce sensitivity to biological effects of interest. Current QA heavily depends on manual assessment, which is time consuming, laborious, and subjective, and therefore not reproducible and generalizable especially for large size multisite studies. Therefore, there is a critical need for a robust and reliable automated quality evaluation index (QEI) that can objectively assess the quality of ASL CBF scans. This QEI could also potentially facilitate real-time feedback during scanning, allowing for immediate adjustments and thereby improving the overall quality of the acquired images.

1.5. Deep Learning

Deep Learning (DL), a subtype of machine learning, provides astonishing performance compared to other state-of-art computational methods across various approaches (Bengio et al. (2013);Deng and Dong

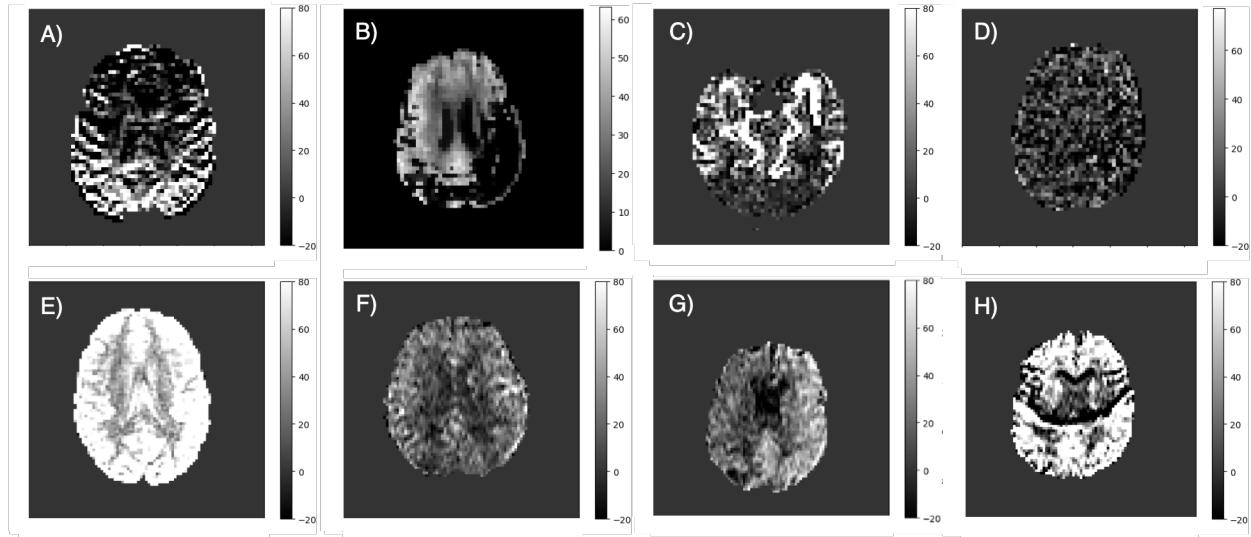


Figure 2: Examples of different sources of artifacts in ASL MRI. A) Motion Artifact B) Clipping Artifact C) Transit Artifact D) Low SNR E) High CBF Values F) Low CBF Values G) Probable Label Asymmetry H) Fat Shift Artifact.

(2014);Lecun et al. (2015);Litjens et al. (2017)), including medical imaging. Initially introduced for image classification in computer vision (Krizhevsky et al. (2012)), DL is now extensively employed to tackle complex problems that analytical methods or traditional machine learning cannot solve. DL networks are motivated by the neuronal visual processing pathway, where a visual observation is hierarchically processed along multiple layers of neurons and eventually abstracted to different top-level features. Multi-layer artificial neural networks were proposed decades ago to mimic this complex learning process, but their use only became practical with the advent of powerful graphical processing units (GPUs) capable of massively parallel computing (Bengio et al. (2013);Deng and Dong (2014);Le-cun et al. (2015);Litjens et al. (2017)). Deep networks are commonly trained with references; this supervised learning is equivalent to nonlinear data fitting. While traditional data fitting is based on a weighted sum of well characterized base functions, DL is based on the weighted sum of the output of a hierarchical network consisting of multiple layers of computing units (artificial neurons).

1.6. Contribution of this work

In this work, we aimed to tackle the challenge of providing an automatic and robust quality assessment method for ASL derived CBF maps by leveraging DL. We provide multiple strategies to derive such metric, both using predetermined features, and the whole CBF map for automatic feature extraction and compare their performances and establish their superiority over previous approaches.

The specific contributions of this work are summarized as follows:

- A feature-based fully connected network (FCN), in which we utilized 7 features commonly used in manual QA of ASL CBF Maps (7-FCN-QEI-Net).
- Development of a 3D DL-based regression model (named QEI-Net) designed to calculate the QEI of raw ASL CBF maps.
- Development of a 3D multi-stage classification model (named MSC-QEI-Net) for assessing the QEI of raw ASL CBF maps.
- An exhaustive comparison of these new approaches with the current state-of-the-art method, providing insights into their relative performances and improvements.

2. State of the art

2.1. DL-based regression approaches for neuroimaging

Since deep learning models first made their mark on neuroimaging in 2014 (?), there has been an exponential increase in research within the field. This remarkable growth can be attributed to two main factors: the increasing availability of data and the emergence of innovative deep learning models , with segmentation and classification tasks ranking at the forefront of the most explored areas. Regression tasks, however, which aim to predict a continuous outcome, have received comparatively less attention due to their perceived complexity. Consequently, several studies, such as that by (?), have opted to recast the initial regression challenge into a classification problem by discretizing the continuum of values into distinct bins, treated as independent classes during training.

(?) delve into a comparative analysis of both methodologies, focusing on predicting brain age from structural MRI scans. Employing a Simple Fully Convolutional 3D CNN architecture with six convolutional blocks, the study experimented with both approaches by merely altering the last Dense layer and meticulously fine-tuning the hyperparameters for each approach. Although the outcomes on the test set were comparably effective for both approaches, the regression method demonstrated markedly superior generalization capabilities on an unseen dataset, thereby underscoring its enhanced potential for broader applicability.

In line with these findings, recent studies highlight the increasing sophistication of deep regression models tailored for neuroimaging data. For instance, (?) introduced deep relation learning, which utilizes a novel approach by considering multiple relational aspects between neuroimaging inputs to enhance regression performance in age estimation tasks. This method leverages deep neural networks to capture complex, nonlinear interactions, providing a more nuanced understanding and robust predictions than traditional methods.

2.2. Deep Learning-based approaches for ASL MRI

In recent years, there have been notable advancements in the utilization of DL for ASL MRI, resulting in considerable improvements when dealing with certain intrinsic difficulties associated with this image modality, including its notoriously lengthy acquisition periods and inadequate SNR. The investigation conducted by (?) was pioneering on that. They presented an innovative DL-based ASL MRI denoising algorithm, with which he improved the SNR of CBF images but also enabled a 75% reduction in acquisition time while maintaining the integrity of the measurements. Similarly, (?) introduced a deep learning algorithm for denoising ASL MRI that combines convolutional neural networks (CNNs) and mutual information from multiple tissue contrasts in ASL acquisition. This approach demonstrated superior performance over traditional and standard deep learning-based denoising methods by significantly enhancing image quality.

A further objective has been to improve the precision and robustness of ASL perfusion imaging. In their study, (?) reported significant advancements in the utilization of CNNs for perfusion imaging. By using CNNs, they found that the quality of perfusion images obtained through CNNs surpassed those created by traditional averaging techniques. Building on these enhancements in imaging techniques, the application of transfer learning has demonstrated potential for augmenting sensitivity, especially in clinical contexts involving Alzheimer's disease. For instance, (?) research highlighted the efficacy of applying transfer learning from healthy subjects to ASL perfusion MRI models. This approach significantly increased the sensitivity of

detection methods for Alzheimer's disease, illustrating how advances in deep learning could be specifically tailored to improve diagnostic processes.

2.3. Quality index of ASL CBF maps

As previously stated, manual quality control of ASL CBF maps through visual inspection is a labor-intensive process that requires significant expertise. This method is also prone to user bias and subjectivity, particularly when applied to large sample sizes. To address these challenges, recent developments in quality assessment for ASL CBF maps have introduced automated tools and visual quality control systems. A notable development in this domain is the one proposed by (Fallatah et al. (2018)). They introduced a dual-component scoring system that evaluates the image quality based on visual contrast and artifact detection. This system, validated across multiple raters, has demonstrated high reproducibility and the ability to effectively discriminate between high- and low-quality clinical scans, offering a reliable threshold for clinical acceptability.

Parallel to these manual evaluation strategies, there has been significant progress in the automation of quality assessments. For instance, (?) developed ASLM-RICloud, an online platform that facilitates the processing of ASL MRI data. Among other features, ASLM-RICloud enables the calculation of a QEI by analyzing factors such as signal-to-noise ratio, contrast-to-noise-ratio, and the detection of artifacts, including motion and distortion.

Another significant contribution to the field was made by (Dolui et al. (2024)), who introduced a sophisticated automated QEI for ASL CBF maps. This novel QEI is designed to replicate the meticulous visual inspections usually performed by experts during manual quality control. As a result of this algorithm, a continuous value between 0 and 1, with higher values indicating superior quality of CBF maps, is provided. The computational features integrated into this QEI methodology involve:

- **Structural Similarity:** The QEI considers the similarity between the brain structure and CBF maps, acknowledging the natural correlation between structure and function. This feature is calculated by constructing a structural pseudo-CBF (spCBF) map, utilizing tissue probability maps weighted to reflect the higher CBF in gray matter (GM) compared to white matter (WM). The Pearson Correlation between the spCBF map and the original CBF map quantifies the structural similarity, contributing to the overall QEI.
- **Spatial Variability:** Recognizing that CBF differs among tissue types, it's important to note that unusual spatial variability might suggest the presence

of artifacts, such as those from motion or inadequate PLD (see an example in Figure 4). Therefore, to accurately reflect these variations, QEI integrates a dispersion index (DI) for CBF values across GM, WM, and cerebrospinal fluid (CSF) masks, normalized by the average GM CBF.

- **Negative GM CBF:** Given that physiological CBF should be positive, the QEI assesses the proportion of GM voxels showing negative CBF values, with an understanding that these represent non-physiological artifact-affected measures.

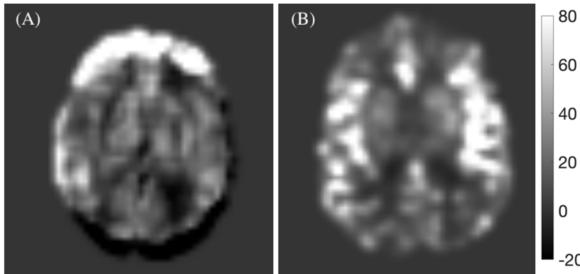


Figure 4: Examples of large spatial variability in ASL derived CBF (A) due to motion or (B) the post-labeling delay (150ms) being significantly shorter than the arterial transit time resulting in labeled signal retained in the arteries instead of the tissue parenchyma while imaging (Dolui et al. (2024)).

The calculation of the final QEI is performed using a formula that geometrically averages the above features, applying exponential functions to restrict each component's influence to the range [0,1]. The formula is as follows:

$$QEI = \sqrt[3]{(1 - e^{-3p_{ss}^{2.4}})e^{-(0.1DI^{0.9} + 2.8p_nG_{MCB}^{0.5})}} \quad (1)$$

where

- p_{ss} is the structural similarity.
- DI is the spatial variability.
- p_nG_{MCB} is the proportion of negative voxels in GM CBF Maps.

This method was evaluated on 48 samples and demonstrated significantly superior performance compared to the approach described by (?). Consequently, it is recognized as the state-of-the-art in automatic QEI for ASL CBF Maps. Therefore, we have used this study as a benchmark to compare the various approaches presented in this work.

3. Material and methods

3.1. Datasets

In this study, a dataset comprising 250 samples was utilized. The samples were collected from several large,

multisite studies that utilized diverse ASL acquisition protocols, as detailed in Table 1. The ratings of the ASL CBF data were meticulously assessed by three expert raters: John A. Detre, Sudipto Dolui, and Ze Wang. Each of these raters brings decades of experience in ASL CBF quality assurance, ensuring the dataset's reliability and validity.

To ensure consistency in the evaluation process across different raters, specific guidelines were established and followed (see Figure 5). These guidelines are defined below:

- **Unacceptable (rating 1):** CBF map is severely degraded by artifacts and is uninterpretable.
- **Poor (rating 2):** CBF map has one or more major artifacts, but can still potentially yield useful information.
- **Average (rating 3):** Acceptable quality CBF map with minor artifacts that do not significantly reduce information value
- **Excellent (rating 4):** High quality CBF map without artifacts.

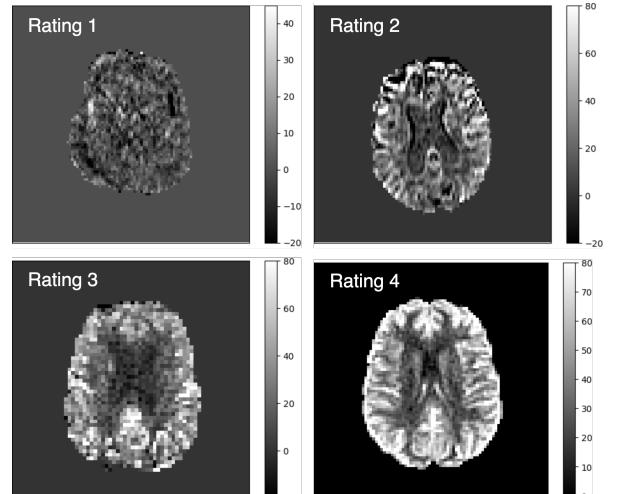


Figure 5: Examples of a distinct case for each rating value

Due to the requirement for a continuous value for some methods presented in this section, the average rating from each rater was calculated and subsequently scaled to fit within a [0,1] range. The formula used is as follows:

$$\bar{r}_{\text{norm}} = \frac{\frac{r_1+r_2+r_3}{3} - 1}{3} \quad (2)$$

with:

- \bar{r}_{norm} : Normalized average rating.
- r_1, r_2, r_3 : Ratings given by three different raters.

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

Dataset	Protocol	Sample Size
Alzheimer's Disease Neuroimaging Initiative (ADNI) (Petersen et al. (2010))	2D PASL	79
Multi-Ethnic Study of Atherosclerosis (MESA) (?)	3D BS PCASL	57
Systolic Blood Pressure Intervention Trial (SPRINT) (?)	2D PCASL	49
Coronary Artery Risk Development in Young Adults (CARDIA) (?)	2D PCASL	25
National Alzheimer's Coordinating Center (NACC) (?)	3D BS PCASL	23
National Institute for Aging – Alzheimer's Association (NIA-AA) (?)	2D PCASL	9
Vascular Contributions to Cognitive Impairment and Dementia (VCID) (?)	3D BS PCASL	6
Private dataset from the University of Pennsylvania	3D BS PCASL	2

To facilitate the rating process, a specialized tool was developed, as outlined in **Appendix A**.

3.2. Dataset Partitioning

To validate the proposed approaches, we employed a stratified 5-fold cross-validation (CV) strategy. Each rating group was independently partitioned, maintaining a consistent class distribution across both training and validation sets.

3.3. Metrics

To assess the effectiveness and analyze the various approaches discussed in this work, we utilized the Mean Squared Error (MSE) metric, as defined in Equation Equation 3.

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (\hat{r}_i - \bar{r}_{\text{norm},i})^2 \quad (3)$$

with:

- \bar{r}_{norm} : Normalized average rating of the experts.
- \hat{r}_i : Predicted rating.
- n : Number of samples.

3.4. Preprocessing

To preserve the integrity of the raw images while addressing inconsistencies in size and format across the dataset, minimal preprocessing was required (see Figure 6). For the FCN-QEI-Net approach, two preprocessing steps were applied: 1) Binarization of the brain tissue probability maps using a 0.9 threshold 2) Smoothing of the image using a 5 mm isotropic kernel. To ensure a consistent comparison with (Dolui et al. (2024)), this smoothing process was carried out using the SPM tool, following the methodology employed in Dolui's study.

On the other hand, for the CNN approaches (QEI-Net and MSC-QEI-Net) we employed Affine Registration using the SimpleITK library to resample the dimensions and spacing of the images to a uniform size of 64x64x32. Following this step, before feeding the data to the network, the images were normalized into a range of [0,1].

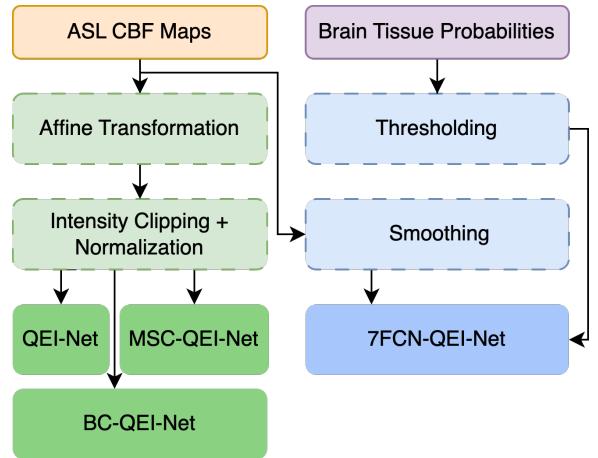


Figure 6: Workflow of the preprocessing pipeline.

3.5. 3-Feature-based FCN (3FCN-QEI-Net)

As previously stated, (Dolui et al. (2024)) introduced a novel algorithm that utilizes three key features commonly employed in manual quality assessment of ASL CBF maps to provide a QEI. While this method achieved high performance and set a new benchmark in the field, it also presented opportunities for further enhancement. In our research, we build upon Dolui's foundational work by incorporating these three features into a FCN comprising six fully connected layers with [16,64,128,64,16,4,1] neurons in each layer, respectively. In the last layer, a Sigmoid activation function is used to predict a continuous value constrained between [0,1]. An example of this network is presented in Figure 7.

3.6. 7- Feature-based FCN (7-FCN-QEI-Net)

In the previous section, we introduced the 3FCN-QEI-Net, which utilizes three features to assess the quality of ASL CBF maps. While this model exhibits considerable potential compared to Dolui's approach, its capability is likely constrained by the limited number of features. To enhance the model's learning ability and provide a more robust assessment, we propose the integration of four additional features. These are as follows

- **SNR:** For this feature, we have computed the Spatial SNR as the ratio of the GM CBF to the standard deviation of the signal in CSF CBF.
- **Summary Statistics:** Several statistics are calculated from the GM and WM of CBF Maps. They consist of the mean, the inverse of the standard deviation and 5th and 95th percentiles of kurtosis.
- **Shannon Entropy:** In order to measure the ghosting and blurring induced by head motion, we have computed the Shannon Entropy. The inverse of this measure is used as a feature for our model.
- **Spatial Gradients:** In ASL CBF Maps, there can be differences along the three axes. The variance of the inverse of CBF Map gradients along each spatial dimension is then used as a feature for our model.

After computing these features, the FCN architecture utilized for the training is the same as the one presented in section 3.5.

3.7. Deep learning-based regression model (QEI-Net)

Building on the methodologies outlined in the previous section, we have developed an alternative approach to providing a QEI for ASL CBF Maps. This new method abandons the manual feature extraction used in the FCN-QEI-Net, opting instead for a more robust solution using CNNs. This technique involves a sophisticated deep-learning based regression model, which we have named QEI-Net.

Drawing inspiration from the 3D VGG architecture delineated by (Simonyan and Zisserman (2014)), we have incorporated several tailored modifications. The presented network (see Figure 7) is structured into four convolutional blocks, each augmented with residual connections to mitigate the vanishing gradient problem (Figure 8). Subsequent to each block, we employed max pooling layers for downsampling with a pooling size of 2 for each channel. At the end of the network, a series of three Fully Connected layers converge to a final neuron activated by a Sigmoid activation function. Finally, for better weight initialization, Xavier’s initialization (?) was utilized. This method consists of initializing the weights such that the variance of the activations is the same across every layer, helping to prevent the gradient from exploding or vanishing. For optimization, we utilized the Adam algorithm, starting with an initial learning rate of 0.0001. We integrated a learning rate decay strategy, predicated on a factor of 0.1 and a patience threshold of 15 epochs. Moreover, the training design was initially set to endure for 400 epochs; however, an early stopping mechanism with a patience parameter of 60 epochs was implemented to prevent overfitting. Finally, since MSE was designated as the principal metric of this project, we have utilized this metric as a loss function for the training of this model.

3.8. A 3D Multi-Stage Classification Model (MSC-QEI-Net)

As delineated in Section 2.1, current advancements in deep learning-based regression models typically reformulate the regression problem as a classification task. This is achieved by discretizing the prediction range into distinct intervals, each representing a unique label. While this technique has been shown to enhance the efficacy of regression methods, it does have a substantial drawback: the precision is dependent on the number of intervals (bins) that are defined. An increased number of bins can yield higher precision but it also intensifies the data imbalance among the bins.

To address these challenges, we propose a multi-stage classification methodology named MSC-QEI-Net. This novel framework diverges from the aforementioned methods, which are focused on converting a regression task into a classification one by dividing the output into bins. Instead, MSC-QEI-Net comprises a series of multi-label classification networks, each corresponding to an individual rater’s assessments within the dataset. By aggregating the outputs of these networks and subsequently normalizing them as per Equation 3, the system synthesizes a continuous value within the [0,1] range, representing the QEI of the image.

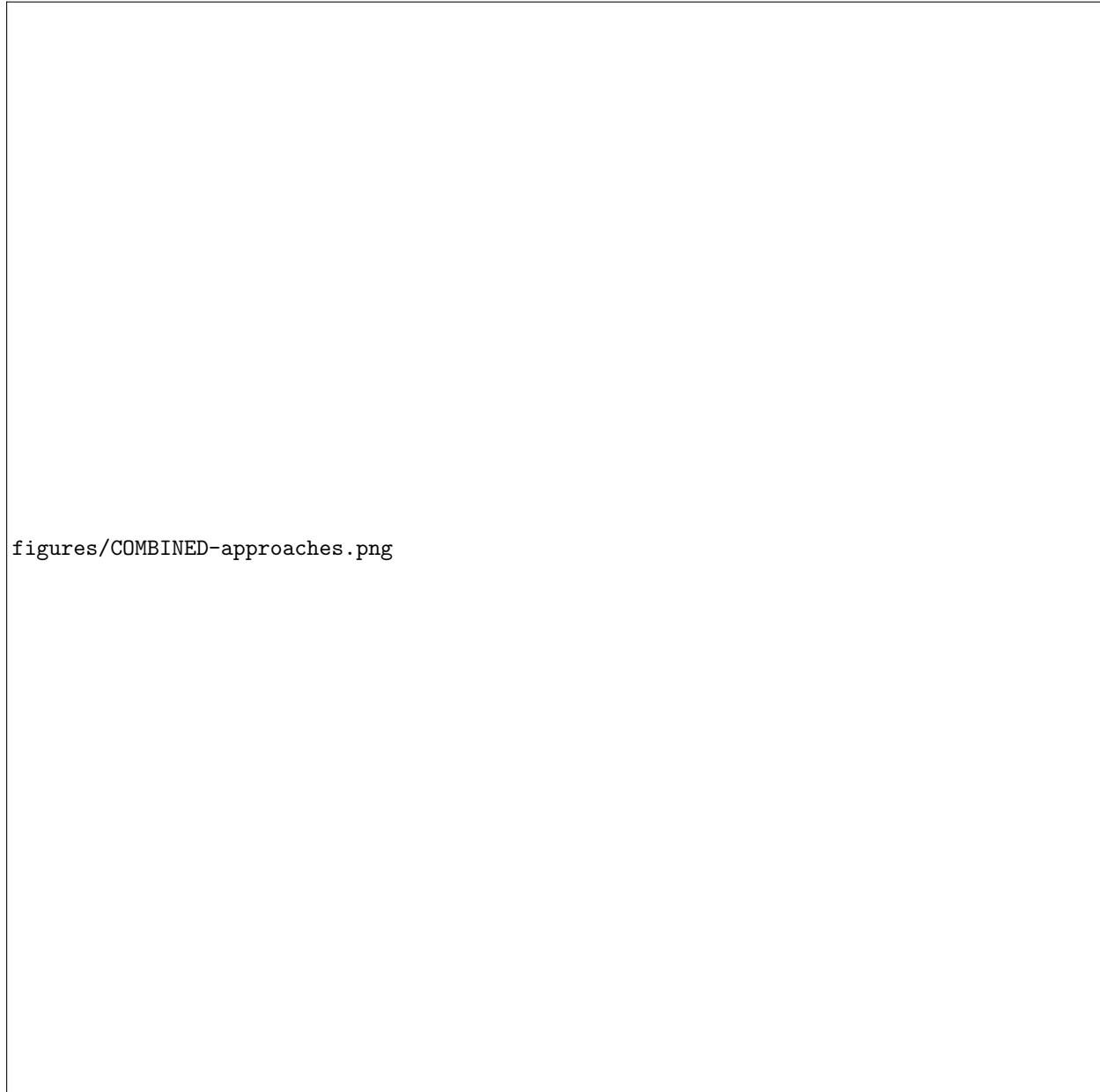
The network used to perform this classification is based on the one presented in section 3.5, with some minor changes. In this architecture, only one fully connected layer has been used as the output layer. Moreover, since we want to perform multi-label classification instead of regression, this layer contains 4 neurons, corresponding to each of the labels of the classification. Finally, SoftMax activation function, which is widely used for multi-label classification tasks, is utilized as the activation function of this layer. For both optimization and training, we applied the same parameters that were previously utilized in the QEI-Net model. However, for the loss function we opted for Categorical Crossentropy loss, a choice prevalent in multiclass classification tasks.

3.9. Computational resources

The models were implemented using Python version 3.10.12 and TensorFlow version 2.16.1. The experiments were conducted on a server using a 64-bit GNU/Linux operating system (Ubuntu 22.04.04). The server was equipped with two Intel Xeon CPUs (2.30GHz), 8 GB of RAM, and a Tesla T4 GPU with 16 GB of memory, utilizing CUDA 12.4 for the experiments.

3.10. Results

This section presents the results of the various approaches discussed in this work. In Table 2 we can observe the different performances of all the models by



figures/COMBINED-approaches.png

Figure 7: Schematic of the different deep learning pipelines implemented in this work. A) Feature-Based approach (FCN-QEI-Net) B) Regression approach (QEI-Net) C) Multi-Stage Classification approach (MSC-QEI-Net)

computing the MSE and the computational time. Additionally, in Table 3 we can observe the sensitivity, specificity, area under the curve (AUC) and the YI chosen for each of the developed approaches.

4. Discussion

5. Conclusions

Acknowledgments

References

- Alsop, D., Detre, J., Golay, X., et al., 2015. 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. *Magn Reson Med* 73, 102–116. doi:10.1002/mrm.25197.
- Bengio, Y., Courville, A., Vincent, P., 2013. Representation learning: A review and new perspectives. *IEEE Trans Pattern Anal Mach Intell* 35, 1798–1828.

Table 2: Performance measure on the dataset presented in section 3.1, of all the approaches compared in this study.

Approach	MSE*	Computational time [min]
(Dolui et al. (2024))'s approach	0.0261	
3FCN-QEI-Net	0.0299	
7FCN-QEI-Net	0.0228	
QEI-Net	0.0159	
MSC-QEI-Net	0.0256	

*MSE**: This metric corresponds to the average of the MSE for each of the 5 validation folds.

Table 3: Performance measures on the dataset presented in section 3.1, of all the approaches compared in this study.

Approach	Sensitivity	Specificity	AUC	Youden Index
(Dolui et al. (2024))'s approach	0.8834	0.8621	0.9270	0.3186
3FCN-QEI-Net	0.8850	0.8773	0.9195	0.3352
7FCN-QEI-Net	0.8282	0.9080	0.9242	0.3042
QEI-Net	0.8527	0.9425	0.9562	0.3355
MSC-QEI-Net	0.7361	0.9540	0.9206	0.5
BC-QEI-Net	0.9018	0.8506	0.9514	0.7524

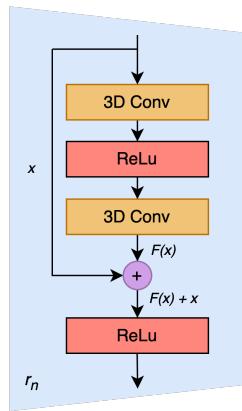


Figure 8: Schematic of the Residual Block used in this study. In the diagram, r_{sn} indicates the sequence number of the block, reflecting their multiple uses throughout the model.

Bernbaum, M., Menon, B., Fick, G., Smith, E., Goyal, M., Frayne, R., Coutts, S., 2015. Reduced blood flow in normal white matter predicts development of leukoaraiosis. *Journal of Cerebral Blood Flow and Metabolism* 35, 1610–1615. doi:10.1038/jcbfm.2015.92.

Binnewijzend, M., Kuijer, J., Benedictus, M., van der Flier, W., Wink, A., Wattjes, M., van Berckel, B., Scheltens, P., Barkhof, F., 2013. Cerebral blood flow measured with 3d pseudocontinuous arterial spin-labeling mr imaging in alzheimer disease and mild cognitive impairment: a marker for disease severity. *Radiology* 267, 221–230. doi:10.1148/radiol.12120928.

Buxton, R., Frank, L., Wong, E., Siewert, B., Warach, S., Edelman, R., 1998. A general kinetic model for quantitative perfusion imaging with arterial spin labeling. *Magnetic Resonance in Medicine* 40, 383–396. doi:10.1002/mrm.1910400308.

De La Torre, J., 2013. Vascular risk factors: A ticking time bomb to alzheimer's disease. *American Journal of Alzheimer's*

Disease and other Dementias 28, 551–559. doi:10.1177/1533317513494457.

Deng, L., Dong, Y., 2014. Deep learning: Methods and applications. *Found Trends® Signal Process* 7, 197–387.

Detre, J., Leigh, J., Williams, D., Koretsky, A., 1992. Perfusion imaging. *Magn Reson Med* 23, 37–45. doi:10.1002/mrm.1910230106.

Dolui, S., Detre, J., Gaussoin, S., Herrick, J., Wang, D., Tamura, M., Cho, M., Haley, W., Launer, L., Punzi, H., Rastogi, A., Still, C., Weiner, D., Wright, J.J., Williamson, J., Wright, C., Bryan, R., Bress, A., Pajewski, N., Nasrallah, I., 2022. Association of intensive vs standard blood pressure control with cerebral blood flow: Secondary analysis of the sprint mind randomized clinical trial. *JAMA neurology* 79, 380–389. doi:10.1001/jamaneurol.2022.0074.

Dolui, S., Li, Z., Nasrallah, I., Detre, J., Wolk, D., 2020. Arterial spin labeling versus $(18)\text{f-fdg-pet}$ to identify mild cognitive impairment. *NeuroImage Clinical* 25, 102146. doi:10.1016/j.niclin.2019.102146.

Dolui, S., Tisdall, D., Vidorreta, M., et al., 2019. Characterizing a perfusion-based periventricular small vessel region of interest. *NeuroImage Clinical* 23, 101897.

Dolui, S., Vidorreta, M., Wang, Z., Nasrallah, I., Alavi, A., Wolk, D., Detre, J., 2017a. Comparison of pasl, pcasl, and background-suppressed 3d pcasl in mild cognitive impairment. *Human Brain Mapping* 38, 5260–5273. doi:10.1002/hbm.23732.

Dolui, S., Wang, Z., Shinohara, R., Wolk, D., Detre, J., I, A.D.N., 2017b. Structural correlation-based outlier rejection (score) algorithm for arterial spin labeling time series. *Journal of Magnetic Resonance Imaging* 45, 1786–1797. doi:10.1002/jmri.25436.

Dolui, S., Wang, Z., Wolf, R., Nabavizadeh, A., Xie, L., Tosun, D., Nasrallah, I., Wolk, D., Detre, J., I, A.D.N., 2024. Automated quality evaluation index for arterial spin labeling derived cerebral blood flow maps. *Journal of magnetic resonance imaging: JMRI* doi:10.1002/jmri.29308.

Fallatah, S., Pizzini, F., Gómez-Anson, B., Magerkurth, J., Vita, E., Bisdas, S., Jäger, H., Mutzaerts, H., Golay, X., 2018. A visual quality control scale for clinical arterial spin labeling images. *European Radiology Experimental* 2, 1–8.

- Fazlollahi, A., Calamante, F., Liang, X., Bourgeat, P., Raniga, P., Dore, V., Fripp, J., Ames, D., Masters, C., Rowe, C., Connelly, A., Villemagne, V., Salvado, O., B. A.I., G. L.R., 2020. Increased cerebral blood flow with increased amyloid burden in the preclinical phase of alzheimer's disease. *Journal of Magnetic Resonance Imaging* 51, 505–513. doi:10.1002/jmri.26810.
- Fernandez-Seara, M., Wang, Z., Wang, J., Rao, H., Guenther, M., Feinberg, D., Detre, J., 2005. Continuous arterial spin labeling perfusion measurements using single shot 3d grase at 3 t. *Magn Reson Med* 54, 1241–1247.
- Gregg, N., Kim, A., Gurol, M., et al., 2015. Incidental cerebral microbleeds and cerebral blood flow in elderly individuals. *JAMA Neurology* 72, 1021–1028. doi:10.1001/jamaneurol.2015.1359.
- Herscovitch, P., Markham, J., Raichle, M., 1983. Brain blood flow measured with intravenous h2(15)o. i. theory and error analysis. *Journal of nuclear medicine: official publication, Society of Nuclear Medicine* 24, 782–789.
- Ito, H., Inoue, K., Goto, R., Kinomura, S., Taki, Y., Okada, K., Sato, K., Sato, T., Kanno, I., Fukuda, H., 2006. Database of normal human cerebral blood flow measured by spect: I. comparison between i-123-imp, tc-99m-hmpao, and tc-99m-ecd as referred with o-15 labeled water pet and voxel-based morphometry. *Annals of nuclear medicine* 20, 131–138.
- Iturria-Medina, Y., Sotero, R., Toussaint, P., Mateos-Perez, J., Evans, A., I. A.D.N., 2016. Early role of vascular dysregulation on late-onset alzheimer's disease based on multifactorial data-driven analysis. *Nat Commun* 7, 11934. doi:10.1038/ncomms11934.
- Jain, V., Langham, M., Wehrli, F., 2010. Mri estimation of global brain oxygen consumption rate. *Journal of cerebral blood flow and metabolism: official journal of the International Society of Cerebral Blood Flow and Metabolism* 30, 1598–1607. doi:10.1038/jcbfm.2010.49.
- Kety, S., Schmidt, C., 1945. The determination of cerebral blood flow in man by the use of nitrous oxide in low concentrations. *Am J Physiol* 143, 53–66.
- Krizhevsky, A., Sutskever, I., Hinton, G., 2012. Imagenet classification with deep convolutional neural networks. *Adv Neural Inf Process Syst*.
- Lassen, N., Henriksen, L., Paulson, O., 1981. Regional cerebral blood flow in stroke by 133xenon inhalation and emission tomography. *Stroke; a journal of cerebral circulation* 12, 284–288.
- Lecun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. *Nature* 521, 436–444.
- Litjens, G., Kooi, T., Bejnordi, B., Setio, A., Ciompi, F., Ghafoorian, M., van der Laak, J., van Ginneken, B., Sánchez, C., 2017. A survey on deep learning in medical image analysis. *Med Image Anal* 42, 60–88.
- Maleki, N., Dai, W., Alsop, D., 2012. Optimization of background suppression for arterial spin labeling perfusion imaging. *Magnetic Resonance Materials in Physics, Biology and Medicine* 25, 127–133.
- Moonen, J., Nasrallah, I., Detre, J., Dolui, S., Erus, G., Davatzikos, C., Meirelles, O., Bryan, R., Launer, L., 2020. Race and sex differences in midlife changes in cerebral volume and perfusion. *Alzheimer's Dement* 16, 1–2.
- Petersen, E., Mouridsen, K., Golay, X., 2010. The quasar reproducibility study, part ii: Results from a multi-center arterial spin labeling test-retest study. *NeuroImage* 49, 104–113.
- Power, J., Barnes, K., Snyder, A., Schlaggar, B., Petersen, S., 2012. Spurious but systematic correlations in functional connectivity mri networks arise from subject motion. *Neuroimage* 59, 2142–2154.
- Rempf, K., Brix, G., Wenz, F., Becker, C., Guckel, F., Lorenz, W., 1994. Quantification of regional cerebral blood flow and volume with dynamic susceptibility contrast-enhanced mr imaging. *Radiology* 193, 637–641. doi:10.1148/radiology.193.3.7972800.
- Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* doi:10.48550/arXiv.1409.1556.
- Wolk, D., Detre, J., 2012. Arterial spin labeling mri: an emerging biomarker for alzheimer's disease and other neurodegenerative conditions. *Current opinion in neurology* 25, 421–428. doi:10.1097/WCO.0b013e328354ff0a.
- Ye, F., Berman, K., Ellmore, T., Esposito, G., van Horn, J., Yang, Y., Duyn, J., Smith, A., Frank, J., Weinberger, D., McLaughlin, A., 2000a. H(2)(15)o pet validation of steady-state arterial spin tagging cerebral blood flow measurements in humans. *Magnetic Resonance in Medicine* 44, 450–456.
- Ye, F., Frank, J., Weinberger, D., McLaughlin, A., 2000b. Noise reduction in 3d perfusion imaging by attenuating the static signal in arterial spin tagging (assist). *Magn Reson Med* 44, 92–100.
- Yonas, H., Darby, J., Marks, E., Durham, S., Maxwell, C., 1991. Cbf measured by xe-ct: approach to analysis and normal values. *Journal of cerebral blood flow and metabolism: official journal of the International Society of Cerebral Blood Flow and Metabolism* 11, 716–725. doi:10.1038/jcbfm.1991.128.
- Zhang, K., Herzog, H., Mauler, J., Filss, C., Okell, T., Kops, E., Tellmann, L., Fischer, T., Brocke, B., Sturm, W., Coenen, H., Shah, N., 2014. Comparison of cerebral blood flow acquired by simultaneous [15o]water positron emission tomography and arterial spin labeling magnetic resonance imaging. *Journal of cerebral blood flow and metabolism: official journal of the International Society of Cerebral Blood Flow and Metabolism* 34, 1373–1380. doi:10.1038/jcbfm.2014.92.

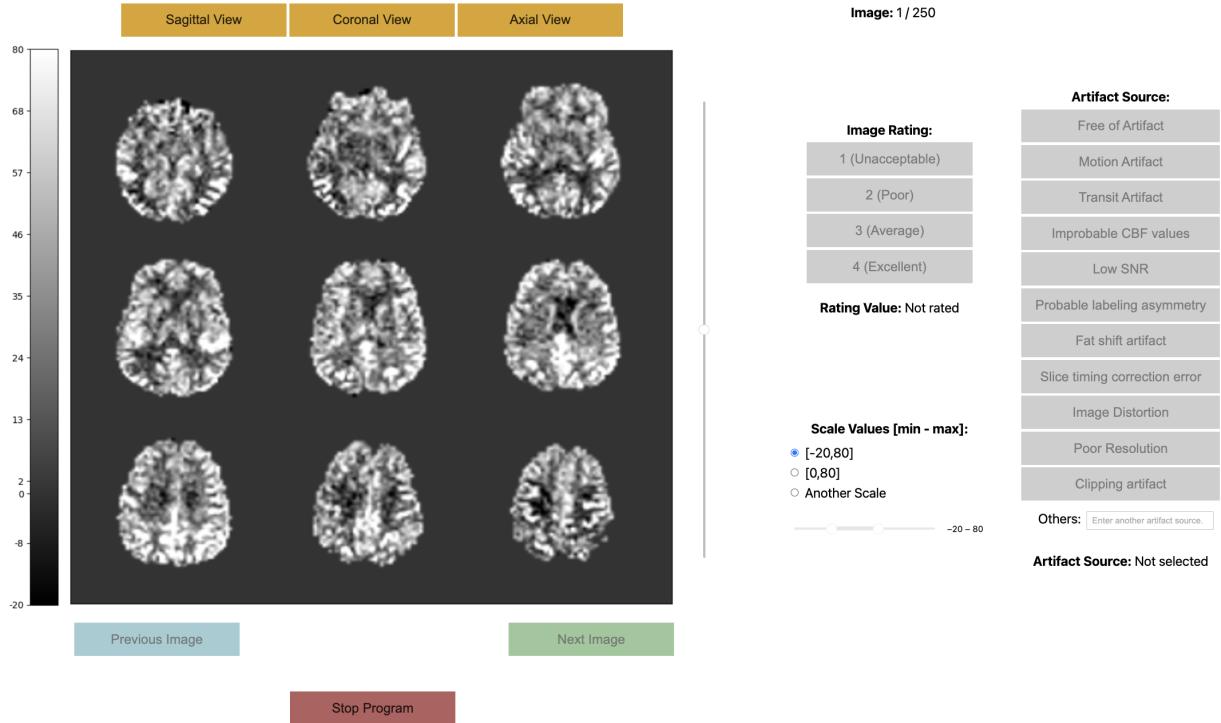


Figure 9: Example of the ASL CBF rating tool.

Appendix A: ASL CBF Rating Tool

Following the depiction of the ASL CBF Rating Tool, we detail the functionality and features of the web-based tool developed to simplify the rating task. This tool is a Python notebook designed to be used with Google Colaboratory, thereby eliminating the need to install software or sensitive data on the user's computer. When the script is initiated, it automatically downloads the dataset from Dropbox to the user's Google Drive. It also generates an Excel file where the ratings are stored.

Some of the tool's characteristics are:

- **Pause and Resume Capability:** The tool allows for pausing and resuming at any point. It automatically checks the Excel file to determine the last image rating, ensuring a seamless continuation of the task.
- **Artifact Documentation:** As part of an upcoming study on classifying imaging artifacts, raters are required to identify and document the sources of any artifacts observed.
- **Intensity Clipping:** To modify image contrast, intensity clipping is employed with default parameters set to [-20, 80].
- **Comprehensive Visualization:** The tool provides multiple views (axial, sagittal, and coronal) of each image. To enable the user to rate the image, all image views must be observed.

- **3D Navigation:** A sliding bar is included to navigate through all slices of the 3D images.

Once all images have been rated, the Excel file is automatically downloaded to the user's computer. This tool is licensed freely and is accessible via the following link: <https://github.com/xavibeltranurbano/ASLCBFRatingTool>