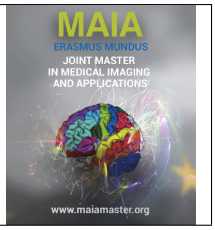
Medical Imaging and Applications

Master Thesis, June 2024



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

Xavier Beltran Urbano*,[1], Sudipto Dolui[1], John A Detre[1]

[1]*Detre Lab, Departments of Neurology and Radiology, University of Pennsylvania, USA*

Abstract

Arterial Spin Labeling (ASL) is a non-invasive magnetic resonance imaging (MRI) technique widely used for measuring cerebral blood flow (CBF). Compared to more conventional approaches, ASL offers several advantages, such as the absence of exogenous tracers and ionizing radiation, lower cost, flexibility of being acquired in routine MRI settings, and is the method of choice to measure CBF in large-scale multisite studies, particularly with repeated acquisitions. However, ASL data can be noisy, and hence quality control (QC) of ASL CBF maps is of particular importance for this modality. Manual QC is time-consuming, laborious, and subjective, highlighting the need for automated solutions. In this study, we proposed three novel deep learning (DL) models designed to provide automatic quality evaluation indices (QEIs) for ASL-derived CBF maps: 7FCN-QEI-Net, Reg-QEI-Net and MSC-QEI-Net. The resulting QEIs are designed to be continuous numbers in the range of 0 and 1. We also trained a deep learning algorithm (BC-Net) to provide a binarized decision about the quality of the CBF map, which indicates if the map should be kept or discarded from group analysis. Additionally, we also considered ensembles of the different networks. These approaches leverage advanced DL techniques to enhance feature representation and achieve superior performance compared to previous state-of-the-art methods.The models were trained on a diverse dataset that included 250 samples from multiple multisite studies. These samples were acquired using different protocols and were rated for quality by three raters, ensuring robustness and generalizability. Additionally, in a separate test set comprising 50 samples, all the deep learning strategies performed better than the current state-of-the-art method. The correlations between the automated QEIs and the average manual ratings were higher than the inter-rater correlations. We also derived and reported QEI thresholds for each method to binarize CBF maps into acceptable and unacceptable categories for each of the non-binarized methods. While the ensemble approaches perform slightly better, the Reg-QEI-Net provided comparable performance and is currently our recommended strategy. The results highlight the potential of DL models in automating and improving the QC process for ASL CBF maps, reducing reliance on manual assessments, minimizing subjectivity, and enhancing reproducibility and consistency across studies.

The code developed for this work is publicly available at: https://github.com/xavibeltranurbano/QEI-Net

Keywords: Arterial Spin Labeled, Deep Learning, CNN, Quality Assessment, FCN, Regression, Reg-QEI-Net, CBF

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 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

*Corresponding author
Email address: xavibeltranurbano00@gmail.com

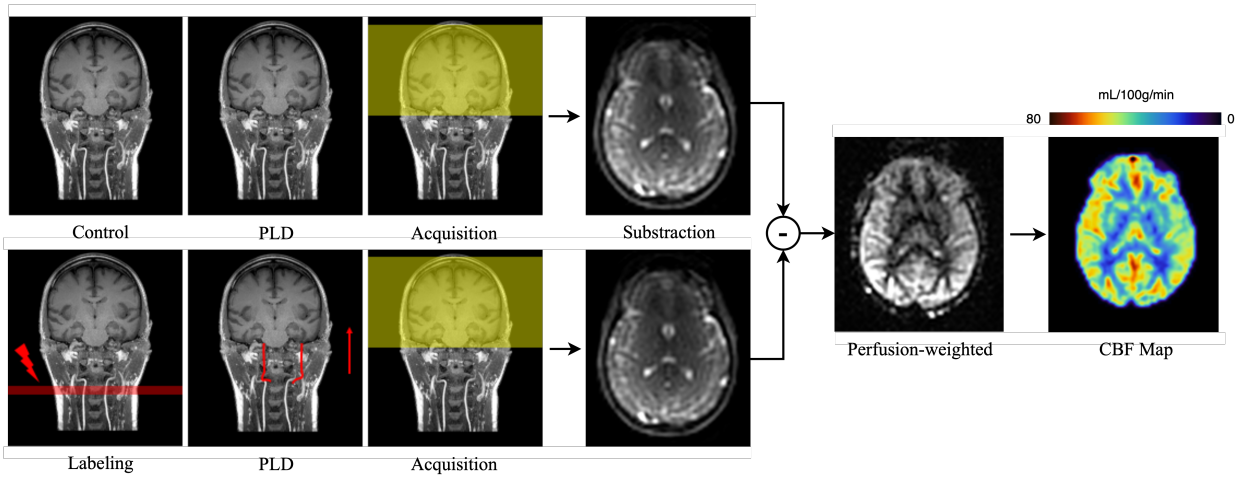


Figure 1: Sequential workflow for ASL CBF map acquisition. This diagram delineates the procedural stages, beginning with the acquisition of control images, followed by the application of labeling and post-labeling delay (PLD). Subsequent subtraction generates the perfusion-weighted images, which are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

serves as a biomarker of functional neurodegeneration due to the strong association of changes in CBF with neural activity (Dolui et al. (2017a)), and therefore can potentially replace glucose metabolism measurements obtained using ^{18}F-Fluorodeoxyglucose Positron Emission Tomography (^{18}F-FDG-PET)(Dolui et al. (2020)). CBF changes have been associated with the incidence and severity of dementia (Dolui 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);Dolui et al. (2024);Fazlollahi et al. (2020)). Moreover, CBF is potentially modifiable therapeutically and hence can be used to monitor 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 (Wolk and Detre (2012)).

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 ^{15}O-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 ^{133}xenon (Lassen et al. (1981)) and stable xenon computed tomography (CT) (Yonas et al. (1991)). Related methods include accumulative radioactive tracers with single-photon emission computed tomography (SPECT) scanning, though agreement of these methods with ^{15}O-PET is suboptimal (Ito et al. (2006)), and methods that use intravascular tracers such as perfusion CT (Koenig et al. (1998)) 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 the administration of an exogenous tracer and exposure to ionizing radiation. Hence, they are at least somewhat invasive and can be difficult to administer to 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 associated costs.

1.2. Arterial Spin Labeled (ASL) perfusion MRI

ASL is a non-invasive magnetic resonance imaging (MRI) technique for measuring tissue perfusion by magnetically labeling arterial blood water as an endogenous tracer (Detre et al. (1992)). Since its inception in 1992 (Detre et al. (1992);Williams et al. (1992)), ASL has been increasingly included in multisite research studies of brain health. Compared to other techniques for measuring cerebral perfusion, ASL offers advantages due to its non-invasive nature and the absence of exogenous radioactive and potentially harmful contrast agents. Furthermore, because MRI does not involve ionizing radiation, this method can be used repeatedly, for example, to assess the effects of drugs or to assess longitudinal changes in cerebral perfusion. Finally, ASL can be acquired as a part of routine MRI,

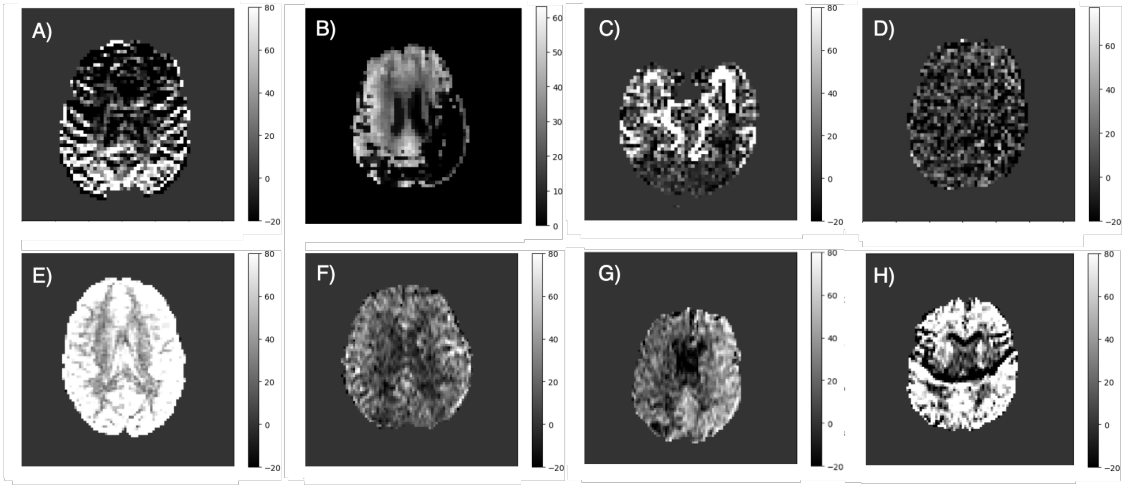


Figure 2: Examples of different sources of artifacts in ASL CBF maps. A) Motion Artifact B) Clipping Artifact C) Transit Artifact D) Low SNR E) High CBF Values F) Low CBF Values G) Probable Label Asymmetry H) Fat Shift Artifact.

which is almost universally acquired in research studies of brain disorders. ASL has been validated against other established modalities for measuring CBF (Ewing et al. (2005);Heijtel et al. (2014);Ye et al. (2000a)). Its use also extends beyond the brain studies and is being applied to other organs, including the kidneys, lungs, heart, placenta, eye, liver, pancreas, and muscle (Taso et al. (2023)). ASL MRI has also been translated to clinical use.

1.3. 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 microvasculature and tissue, a brain MRI (labeled image) is acquired. A "control" brain image is also obtained with a sham labeling procedure that does not magnetically label blood water. The difference between the control and label image is proportional to CBF and is converted to absolute CBF quantification using a proton density image 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, suboptimal choice of imaging parameters, and other non-idealities inherent to MRI scanners can lead to severe artifacts (Dolui et al. (2017b);Li et al. (2018)) (see Figure 2). This can be partially mitigated by averaging multiple control-label pairs, using advanced signal processing strategies, and using background suppression (BS) of static brain water. BS 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.

1.4. 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 (see Figure 3). 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) amplifiers are not capable of continuous RF excitation. Pulsed ASL (PASL) instantly labels a thick slab in the neck, and is compatible with body RF excitation, though the method suffers from lower SNR compared to CASL. The current recommended labeling strategy is pseudo-continuous labeling (PCASL), which employs a series of short RF pulses to mimic continuous labeling. ASL type can also vary based on the duration of the post labeling delay (PLD) – a longer post labeling delay can ensure delivery of the labeled blood to the brain tissue, though at the expense of reduced SNR since the magnetic label decays rapidly. A series of ASL images acquired with different labeling and/or PLDs can also be combined to obtain a CBF map, allowing more accurate modeling of regional CBF values (Woods et al. (2024)). Finally, ASL image quality can vary based on the type of image readout. Echo-planar imaging (EPI) was initially the preferred choice because of speed and sensitivity, though it is being slowly replaced by 3D imaging (GRASE or SPIRAL) optimally combined with BS of static brain tissue. Notably, several other variants of ASL exist; for example, velocity selective ASL (VASL) is an emerging method that labels the arterial blood water close to the imaging site

instead of the neck (Qin et al. (2022)).

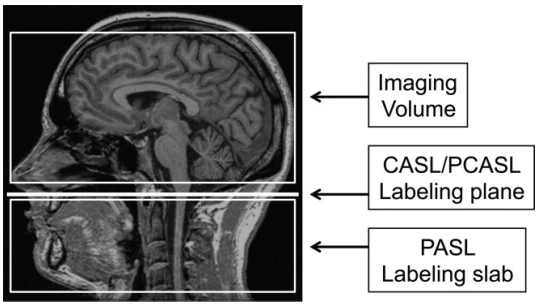


Figure 3: Schematic diagram of imaging and labeling regions for CASL/PCASL and PASL. In CASL/PCASL, labeling occurs as blood flows through a single labeling plane, while in PASL, a slab of tissue, including arterial blood, is labeled (Alsop et al., 2015).

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

In recent years, ASL has gained popularity among perfusion imaging modalities for its use in research settings, 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 the 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 (Friston et al. (1996);Power et al. (2012)). Motion effects are less visible, though still present, in acquisitions using BS of static signal (Ye et al. (2000a);Fernandez-Seara et al. (2005);Maleki et al. (2012)). Artifacts can also result from an 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 artifacts, and clipping of signals, 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 QC a more prominent need in these clinical applications.

Because of potential artifacts in the ASL derived CBF maps, QC 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 QC heavily depends on manual assessment, which is time consuming, laborious, and subjective, and therefore not reproducible and generalizable, especially for large-scale 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.

1.6. Deep Learning

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

1.7. Contribution of this work

In this work, we aimed to tackle the challenge of providing an automatic and robust QC method for ASL-derived CBF maps by leveraging DL. We explored multiple strategies to derive this metric, including both the use of predetermined features and the entire CBF map for automatic feature extraction. We then compared their performances, demonstrating their superiority over previous approaches.

The specific contributions of this work include the development of the following DL-based methods to obtain a QEI of raw CBF maps:

- A feature-based regression model, for which we extracted 7 predetermined features to train a fully connected network (named 7-FCN-QEI-Net).
- A 3D DL-based regression model (named Reg-QEI-Net).
- A 3D multi-stage classification model (named MSC-QEI-Net).
- A 3D binary classification model (named BC-Net).
- Three ensemble methods of the best performing algorithms.

An extensive comparison of these new approaches with the current state-of-the-art method was performed, 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 (Plis et al. (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 improvement of computational resources such as GPUs. Thanks to these advancements, deep learning has emerged as a leading approach in medical imaging research, with segmentation and classification tasks ranking at the forefront of the most explored areas. However, regression tasks, which aim to predict a continuous outcome, have received comparatively less attention due to their perceived complexity. Consequently, several studies, such as that by (Peng et al. (2021)), 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. (Leonardsen et al. (2022)) delve into a comparative analysis of both methodologies, focusing on predicting brain age from structural MRI scans. Employing a 3D Convolutional Neural Network (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, (He et al. (2022)) 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. By leveraging deep neural networks to capture complex and non-linear interactions, this method provides a more nuanced understanding and robust predictions than traditional methods.

2.2. Deep Learning-based approaches for ASL MRI

In recent years, there have also 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 lengthy acquisition periods, inadequate SNR, and low spatial and temporal resolution. In their study, (Kim et al. (2018)) reported significant advancements in the quality of ASL MRI images using CNNs that surpassed those created by traditional averaging techniques. Building on these improvements in imaging techniques, the application of transfer learning has demonstrated potential for augmenting sensitivity, especially in clinical contexts involving AD. For instance, (Zhang et al. (2022)) 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 AD, illustrating how advances in deep learning could be specifically tailored to improve diagnostic processes. The investigation conducted by (Xie et al. (2020)) presented an innovative DL-based ASL MRI denoising algorithm that improved the SNR of CBF images and enabled a 75% reduction in acquisition time while maintaining the integrity of the measurements. Similarly, (Gong et al. (2020)) introduced a DL algorithm for denoising ASL MRI that combines 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.

2.3. Quality index of ASL CBF maps

As previously stated, QC 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. The work in (Fallatah et al. (2018)) introduced a well-characterized dual-component scoring system that evaluates the image quality based on visual contrast and artifact detection and thus reduces the subjectivity of the rating system. 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; however, it still suffers from most of the drawbacks of manual rating.

Parallel to these manual evaluation strategies, there have been efforts to automate quality assessments. For

instance, (Li et al. (2019)) developed ASLMRICloud, an online platform that facilitates the processing of ASL MRI data. Among other features, ASLMRICloud enables the calculation of a quality index by analyzing and averaging the voxelwise temporal standard error (SNR) across the CBF time series obtained from the repeated acquisitions of the multiple control/label pairs. However, this approach cannot assess systematic artifacts that are consistent in the time series, such as those caused by short PLD. Moreover, it cannot be applied to datasets that include only one output volume of the average control-labeled difference image rather than the control-labeled image time series (e.g., product ASL on a GE MRI scanner). Finally, temporal standard error considers the quality of the raw data instead of the final CBF map, which can be of improved quality through the application of signal processing strategies.

The most recently published contribution to the development of an automated QEI for ASL CBF maps was made by (Dolui et al. (2024)). This novel QEI assigns a continuous value between 0 and 1 to each CBF map, with higher values indicating a superior quality of the CBF map. The algorithm used predefined features to train a model against human rating, where the features were chosen to replicate the meticulous visual inspections usually performed by experts during manual QC. The computational features integrated into the 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 a weighted sum of tissue probability maps 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 was used as a feature in the QEI derivation.
- **Spatial Variability:** Although CBF differs among tissue types, unusual spatial variability might suggest the presence of artifacts, such as those from motion or inadequate PLD (see examples 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 mean GM CBF.
- **Negative GM CBF:** Given that physiological CBF should be positive, the QEI incorporates the proportion of GM voxels showing negative CBF values, since those voxels represent non-physiological artifact-affected measures.

The final QEI was performed by fitting these features separately to human ratings of 101 CBF maps, and subsequently performing a geometric average of the fits corresponding to each feature as follows:

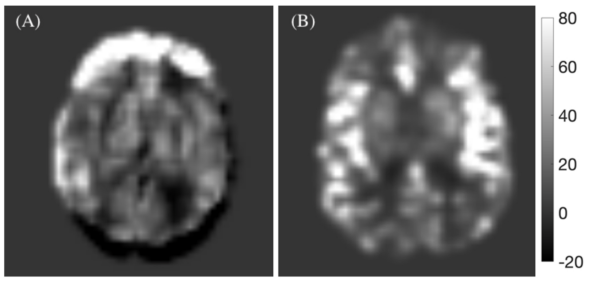


Figure 4: Examples of large spatial variability in ASL derived CBF (A) due to motion or (B) the post-labeling delay (150ms) being significantly shorter than the arterial transit time resulting in labeled signal retained in the arteries instead of the tissue parenchyma while imaging (Dolui et al. (2024)).

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

where

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

This method showed similar agreement to inter-rater reliability, improved statistical analyses, and performed better than the method developed by (Li et al. (2019)). 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 to train the different models. 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. Dr. Detre, the inventor of ASL, has over 30 years of experience, while Dr. Dolui and Dr. Wang each have more than 10 years of experience with this technique. Their extensive experience in ASL CBF quality assurance ensures the dataset's reliability and validity. Additionally, a separate set of 50 CBF maps rated by Dr. Detre and Dr. Dolui was used as the test set to assess the performance of the algorithms on unseen data. All the data used in this project have been acquired using Siemens MRI scanners.

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

Dataset	Protocol	Sample Size
Alzheimer's Disease Neuroimaging Initiative (ADNI) (Wang et al. (2013))	2D PASL	79
Multi-Ethnic Study of Atherosclerosis (MESA) (Austin et al. (2024))	3D BS PCASL	57
Systolic Blood Pressure Intervention Trial (SPRINT) (Dolui et al. (2022))	2D PCASL	49
Coronary Artery Risk Development in Young Adults (CARDIA) (Dolui et al. (2016))	2D PCASL	25
National Alzheimer's Coordinating Center (NACC) (Dolui et al. (2019))	3D BS PCASL	34
Vascular Contributions to Cognitive Impairment and Dementia (VCID) (Sadaghiani et al. (2023))	3D BS PCASL	6

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.

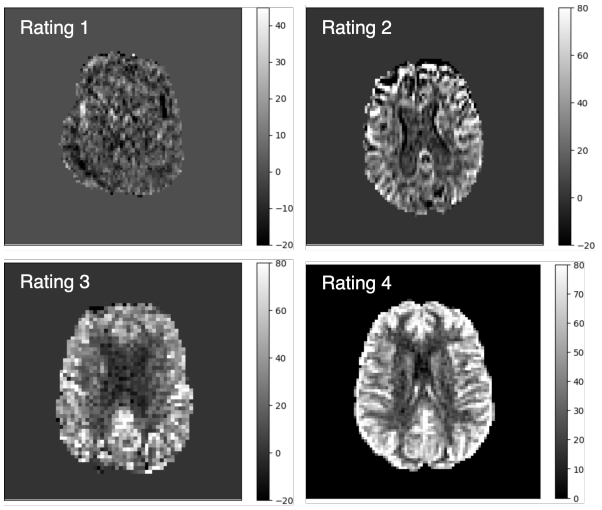


Figure 5: Examples of a distinct case for each rating value.

In the regression-based approaches (mentioned in the introduction and in more detail below), we averaged the ratings to obtain a composite rating score and also to increase the reliability of the measures. Furthermore, we wanted the final QEI to be in the [0,1] range and hence normalized the ratings between 0 and 1. 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 5-fold cross-validation (CV) strategy. Thus, in each fold, 80 percent of the data was used to train the model, and the remaining 20 percent was kept as a validation set. Finally, as previously mentioned, we tested our models using a test set consisting of 50 samples.

3.3. Preprocessing

The CBF maps were derived from ASL data using standard processing strategies (Alsop et al. (2015)). For the purpose of developing the QEI, additional preprocessing was required (see Figure 6). We have followed two different DL strategies, a FCN based on predetermined features and CNNs using the CBF images. For the former approach, two preprocessing steps were applied:

- Generation of binary masks corresponding to GM, WM and CSF to extract CBF signal in the regions.
- Smoothing of the CBF images using a 5 mm isotropic kernel. A similar approach was used by (Dolui et al. (2024)) to extract features from the CBF maps.

For the CNN approaches (Reg-QEI-Net, MSC-QEI-Net, and BC-Net), we used the SimpleITK library to perform an affine transformation, resampling the dimensions and spacing of the images to a uniform size of 64x64x32. This step accounted for variations in image sizes acquired across different studies and protocols. After resampling, the images were intensity-clipped to the range [-10, 80] and subsequently normalized to a range of [0, 1] before being fed into the network.

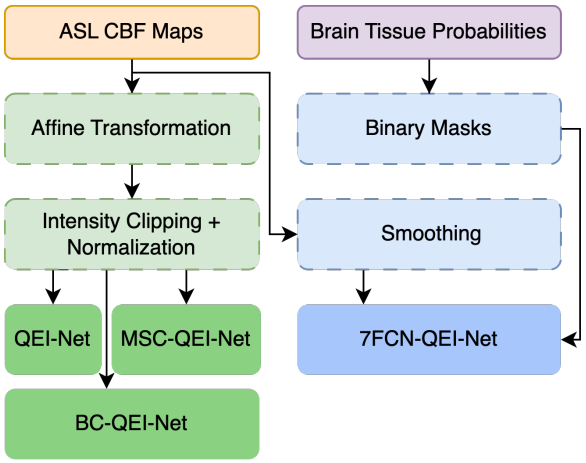


Figure 6: Workflow of the preprocessing pipeline.

3.4. Data Augmentation

Data augmentation techniques are methods used to artificially increase the variability of a dataset by applying various transformations to the original data. These transformations enhance the generalization capabilities of CNN models by exposing them to a wider range of variations. In this work, we used random vertical and horizontal flips, as well as rotations between -5 to 5 degrees.

3.5. Deep Learning models

3.5.1. 7- Feature-based FCN model (7-FCN-QEI-Net)

As previously stated, (Dolui et al. (2024)) introduced a novel algorithm that utilizes three key features commonly employed in manual QC of ASL CBF maps to provide a QEI. While this method achieved high performance and set a new benchmark in the field, its capability is likely constrained by the limited number of features. In our research, we build upon that foundational work by proposing the integration of four additional features. These features 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:** 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 in intensities along the three axes due to possible intensity variation or incorrect application of model equations. 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, they are combined with the features from (Dolui et al. (2024)) and used as input for an FCN architecture (named **7-FCN-QEI-Net**) comprising of seven fully connected layers (FCL) with [64,256,512,256,64,16,1] neurons in each layer, respectively. In the last layer of this network, a sigmoid activation function is used to predict a continuous value constrained between [0,1]. Finally, squared error (SE, defined in section 3.7 below) was designated as the principal metric for this project, and thus, Mean Squared Error (MSE) was used as the loss function for the training of this model. An example of this network is presented in Figure 7.

3.5.2. Deep learning-based regression model (Reg-QEI-Net)

Next, instead of the manual feature extraction used in the 7FCN-QEI-Net, we opted for data-driven approaches using CNNs where the CBF maps were used as input. These methods do not require a segmented image of different brain tissues, making them effective even when a structural image necessary for accurate segmentation is unavailable. This technique involves a sophisticated deep-learning based regression model, which we have named **Reg-QEI-Net**.

Drawing inspiration from the 3D VGG architecture delineated by (Simonyan and Zisserman (2014)), we have incorporated several tailored modifications. The presented network, illustrated in Figure 7, is structured into four convolutional blocks, each augmented with residual connections to mitigate the vanishing gradient problem (see Figure 8). After the first three blocks, max pooling layers with a pooling size of 2 are employed for downsampling each channel. The network concludes with a series of three FCL, culminating in a final neuron activated by a sigmoid function. For better weight initialization, we utilized Glorot's initialization method (Glorot and Bengio (2010)), which ensures the variance of activations remains consistent across every layer, preventing the gradient from exploding or vanishing. The Adam optimization algorithm was used with an initial learning rate of 0.0001. Moreover, a batch size of 32 samples and a learning rate decay strategy were applied, with a decay factor of 0.1 and a patience threshold of 15 epochs. Although the training was initially set to run for 400 epochs, an early stopping mechanism with a patience parameter of 60 epochs was implemented to prevent overfitting. Additionally, a dropout rate of 20% was applied after the fully connected layers to further prevent overfitting. Finally, similar to the 7FCN-QEI-Net approach, MSE was used as the loss function for training this model.

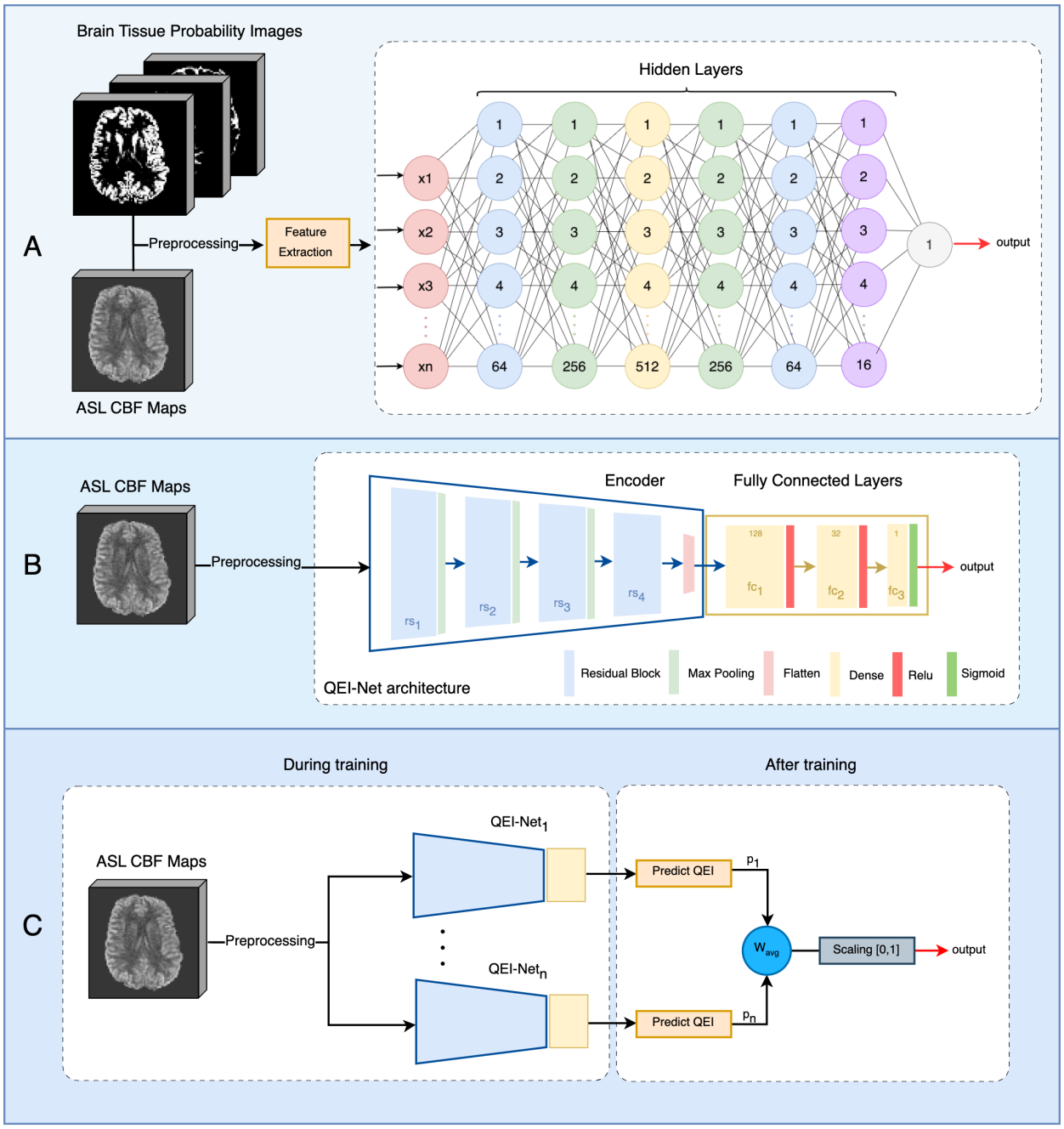


Figure 7: Schematic of the different deep learning pipelines implemented in this work. A) Feature-Based approach (FCN-QEI-Net) B) Regression approach (Reg-QEI-Net) C) Multi-Stage Classification approach (MSC-QEI-Net).

3.5.3. 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-

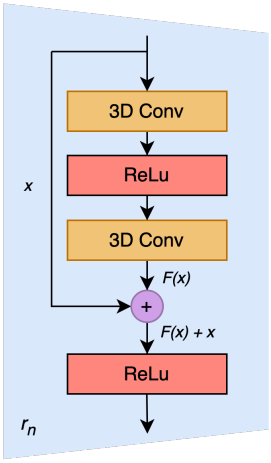


Figure 8: Schematic of the Residual Block used in this study. In the diagram, r_n indicates the sequence number of the block, reflecting their multiple uses throughout the model.

QEI-Net comprises a series of multi-label classification networks, each corresponding to an individual rater's assessments within the dataset. The network used to perform this classification is based on the one presented in Section 3.5.2 with some minor changes. In this architecture, since we want to perform multi-label classification instead of regression, the last FCL contains 4 neurons, corresponding to each of the labels of the classification. In line with this modification, the softmax activation function, which is widely used for multi-label classification tasks, was utilized as the activation function of this layer. For both optimization and training, we applied similar hyperparameters to those previously used in the Reg-QEI-Net model. For the loss function, however, we opted for Focal Categorical Crossentropy loss, a prevalent choice in multiclass classification tasks with imbalanced data.

After training the network, we compute the weighted average of the prediction by following the formula delineated in Equation 2.

$$\text{Weighted Average Prediction} = \sum_{i=1}^{n} (p_i \cdot i) \qquad (2)$$

Where:

- n is the number of classes.
- p_i is the prediction score for the i-th class.
- i is the class label, ranging from 1 to n.

Then, by aggregating the outputs of these networks and subsequently normalizing them, the system synthesizes a continuous value within the [0,1] range, representing the QEI of the image.

3.5.4. A 3D Binary Classification Network (BC-Net)

One of the main objectives of this project is to develop a robust method for discarding unacceptable CBF maps, which can be framed as a binary classification problem instead of assigning a continuous number defining the quality. Therefore, we also implemented a 3D binary classification approach named **BC-Net**. To do so, we have first binarized the expert ratings by following these criteria:

- **Unacceptable Quality (0):** if any of the raters gave a rating of 1 to the image.
- **Acceptable Quality (1):** otherwise.

Furthermore, we used the same parameters and architecture as the Reg-QEI-Net methodology described in Section 3.5.2. However, some minor adjustments were made to optimize the network. The main difference lies in the ground truth used to train the network. For Reg-QEI-Net, we used continuous values within the range [0,1], whereas for BC-QEI-Net, we used binary decision values explained above. For this reason, we utilized a binay cross-entropy loss function and a sigmoid activation function in its final FCL. The output of the BC-Net falls within the range of 0 to 1, representing the probability that a given sample is of acceptable quality.

3.5.5. Additional Experiments

Various combinations of the previous methods (Reg-QEI-Net, 7FCN-QEI-Net, and MSC-QEI-Net), that could potentially result in a better model, were also studied. BC-Net was not used in the combination since it represents a binary decision while other outputs a QEI value. The different combination methods are as follows:

- **Ensemble 1:** This is the simplest ensemble method, which consists of averaging the predictions from each of the networks.
- **Ensemble 2:** In this method, we calculate the weighted average of the predictions. To calculate the weights of each method, we have trained a function that optimizes the weights assigned to the different models to minimize the MSE between the ratings and the predictions.
- **Ensemble 3:** This method utilizes stacking, an ensemble technique that combines the predictions of multiple base models to enhance predictive performance. In this approach, the predictions from the QEI models serve as input features for a meta-model, which was trained using a 5-Fold CV with a linear regression algorithm that learns to make the final prediction by leveraging the strengths and mitigating the weaknesses of the individual models.

To limit the number of ensembles, only the best-performing models (Reg-QEI-Net and 7FCN-QEI-Net, see Section 4) were used. After training Ensemble 2 on the validation data, the resulting weights assigned to Reg-QEI-Net and 7FCN-QEI-Net were 0.663 and 0.337, respectively. These weights were then used to compute the weighted average of the predictions on the test data. Similarly, after training the linear regression models on the validation set, these models were subsequently applied to the test set.

3.6. Gradient-weighted Class Activation Mapping (Grad-CAM) and Heatmap Generation

The QEI developed from the above approaches provides a summary metric for assessing the overall quality of the entire image. However, when the quality is not perfect, the QEI only indicates the presence of the artifacts in the image, without providing information about the location of the artifact. This is important information in region of interest (ROI) analysis as the mean CBF in the corresponding ROI can be contaminated by artifacts, although the overall CBF map might pass the QEI threshold, and that can subsequently bias the analysis. To visualize where the networks are focusing their attention, or in other words, which region of the image is contributing most to the QEI, we have implemented Gradient-weighted Class Activation Mapping (Grad-CAM) Selvaraju et al. (2017). Grad-CAM leverages the gradients flowing into a chosen convolutional layer to generate a localization map, or heatmap, which highlights the important regions in the input image. This technique provides a visual explanation for the model's predictions by identifying the areas in the brain images that contribute the most to the network's decision-making process. For our implementation, we have utilized the Reg-QEI-Net model to generate the heatmap. Among all the convolutional layers of the network, we utilized the 5th 3D convolutional layer, which is located in the third residual block. This decision was made because this intermediate layer provides a balance between low-level feature extraction and high-level semantic information, making it ideal for generating detailed and informative heatmaps.

3.7. Algorithm Evaluation Metrics

To assess the performance of the algorithms, we computed the SE between the average manual ratings and the automated QEI for each CBF map, as defined below.

$$\mathrm{SE}_i = (\hat{r}_i - \overline{r}_{\mathrm{norm},i})^2 \tag{3}$$

with:

- $\overline{r}_{\mathrm{norm}}$: Normalized average rating of the experts.
- $\hat{r}_i$: Predicted rating.

In addition to that, we also reported the Pearson's correlation (PC) coefficient between the automated QEI and the average human rating and compared that to the correlation between the raters. Finally, dividing the data as unacceptable and acceptable as described in Section 3.5.4, we computed the receiver operating characteristic (ROC) curve and the area under the curve (AUC). To establish a QEI threshold, we have calculated the Youden Index (YI), as introduced by (Ruopp et al. (2008)). The YI is a statistical measure that aims to maximize both sensitivity and specificity. By computing the euclidean distance between all points of the ROC curve and the ideal point located at the coordinates [0,1], the YI identifies the best operating point in the curve. Thereafter, we computed sensitivity and specificity based on that threshold.

3.8. Computational resources

The models were implemented using Python version 3.10.12 and TensorFlow version 2.16.1. The experiments were conducted on Google Cloud Platform (GCP) 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.

4. Results

4.1. Algorithm Evaluation Metrics

Table 2 shows the mean, standard deviation, median, and IQR of the SE of the validation set (obtained from the 5-fold CV strategy), while Table 3 shows the same for the test set. Figure 10(a) and Figure 10(b) present the violin plots for the same. Table 2 and Table 3 also list the PC coefficients with the average expert ratings. Notably, the PC coefficient for the 250 samples used for training is 0.85 between Dolui and Detre, 0.84 between Dolui and Wang, and 0.80 between Detre and Wang. Furthermore, the correlation coefficient between Dolui and Detre was 0.77 for the test data set. In each case, the agreement between the raters was lower than the agreement between the average rating and the automated methods.

Additionally, Table 2 and Table 3 show the AUC, sensitivity, and specificity as detailed in Section 3.7. Note that the YI was based on the validation set and hence has not been presented in the table related to the test set. Figure 11(a) and Figure 11(b) show the ROC for the validation and the test sets. As expected, the performance of the test set was slightly worse than the validation set based on all the metrics. Although all the algorithms provided comparable performance, with the

Table 2: Comparison of the current state-of-the-art in the field of QEI of ASL CBF Maps (Dolui et al. (2024)) with the different QEI methods presented in this study using the validation data set.

Method	MSE ± std SE	Median of SE (IQR)	PC Coefficient	AUC	Sensitivity	Specificity	YI
Dolui et al. 2024 QEI	0.02160 ± 0.03184	**0.00416 (0.01416)**	0.943	0.948	0.904	0.922	0.457
7FCN-QEI-Net	0.01646 ± 0.02986	0.01044 (0.02562)	0.903	0.950	**0.911**	0.922	0.325
Reg-QEI-Net	0.01251 ± 0.02213	0.00611 (0.01556)	0.923	0.958	0.815	**0.965**	0.461
MSC-QEI-Net	0.02123 ± 0.02579	0.01348 (0.02287)	0.921	0.941	0.822	0.930	0.419
BC-Net	-	-	-	0.940	0.889	0.852	0.614
Ensemble 1	0.01144 ± 0.02008	0.00505 (0.01124)	0.947	0.963	0.889	0.930	0.348
Ensemble 2	**0.01112 ± 0.01917**	0.00432 (0.01078)	**0.949**	**0.964**	0.896	0.913	0.327
Ensemble 3	0.01184 ± 0.02109	0.00439 (0.01134)	0.945	0.961	0.896	0.904	0.335

Table 3: Comparison of the current state-of-the-art in the field of QEI of ASL CBF Maps (Dolui et al. (2024)) with the different QEI methods presented in this study using the test data set.

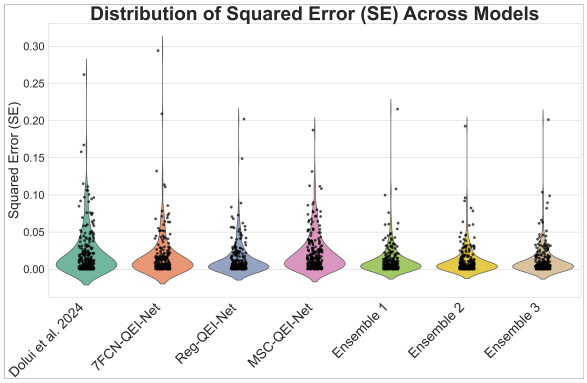
Method	MSE ± std SE	Median of SE (IQR)	PC Coefficient	AUC	Sensitivity	Specificity
Dolui et al. 2024 QEI	0.04730 ± 0.05045	0.02945 (0.05103)	0.808	0.896	0.865	0.583
7FCN-QEI-Net	0.02552 ± 0.03811	0.01256 (0.02680)	0.844	0.915	0.757	0.571
Reg-QEI-Net	0.02308 ± 0.02758	0.01464 (0.02414)	**0.905**	**0.950**	0.892	0.765
MSC-QEI-Net	0.02776 ± 0.03141	0.02179 (0.03967)	0.877	0.909	0.838	0.625
BC-Net	-	-	-	0.946	0.880	0.705
Ensemble 1	**0.01795 ± 0.02002**	**0.00904 (0.02551)**	0.897	0.946	0.892	0.750
Ensemble 2	0.01822 ± 0.01854	0.01126 (0.02616)	**0.905**	0.946	**0.919**	0.786
Ensemble 3	0.01814 ± 0.01864	0.01153 (0.02659)	**0.905**	0.946	**0.919**	**0.800**

ensembles performing slightly better than the individual algorithms, Reg-QEI-Net delivered the best performance among the individual approaches in most metrics, and its results were also comparable to those of the ensembles.

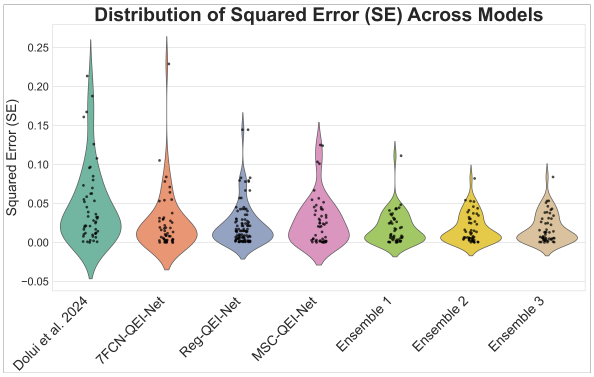
Figure 9 shows examples of the prediction using each method in 4 samples from the test set, one per rating category, in which all raters agreed with the same ratings. Each image also shows the QEI obtained using different methods, with the first entry showing the manual rating scaled in the [0,1] range. The best methods in each case, as determined by a QEI value closest to the manual rating, are shown in green. Finally, in Figure 13, we show the heatmap of the Reg-QEI-Net model, the best performer amongst the individual approaches, corresponding to various samples, each demonstrating different sources of artifacts.

4.2. QEI across studies

Given that the dataset used in this study includes data from six different multisite studies, we have analyzed the performance of the presented approaches across these sources. Figure 12 shows the distribution of the QEI for each method across the different studies for both the validation and the test set. Note that the test set does not encompass all the studies. The figure also shows the color-coded manual ratings for each method. As expected, the VCID with its advanced protocol had the best QEI, while the ADNI ASL, with a relatively poor protocol and also acquired in older healthy participants and patients who are more susceptible to move, performed worst.



(a)



(b)

Figure 10: Violin plot illustrating the distribution of SE across all the methods compared in this study for (a) validation and (b) test set.

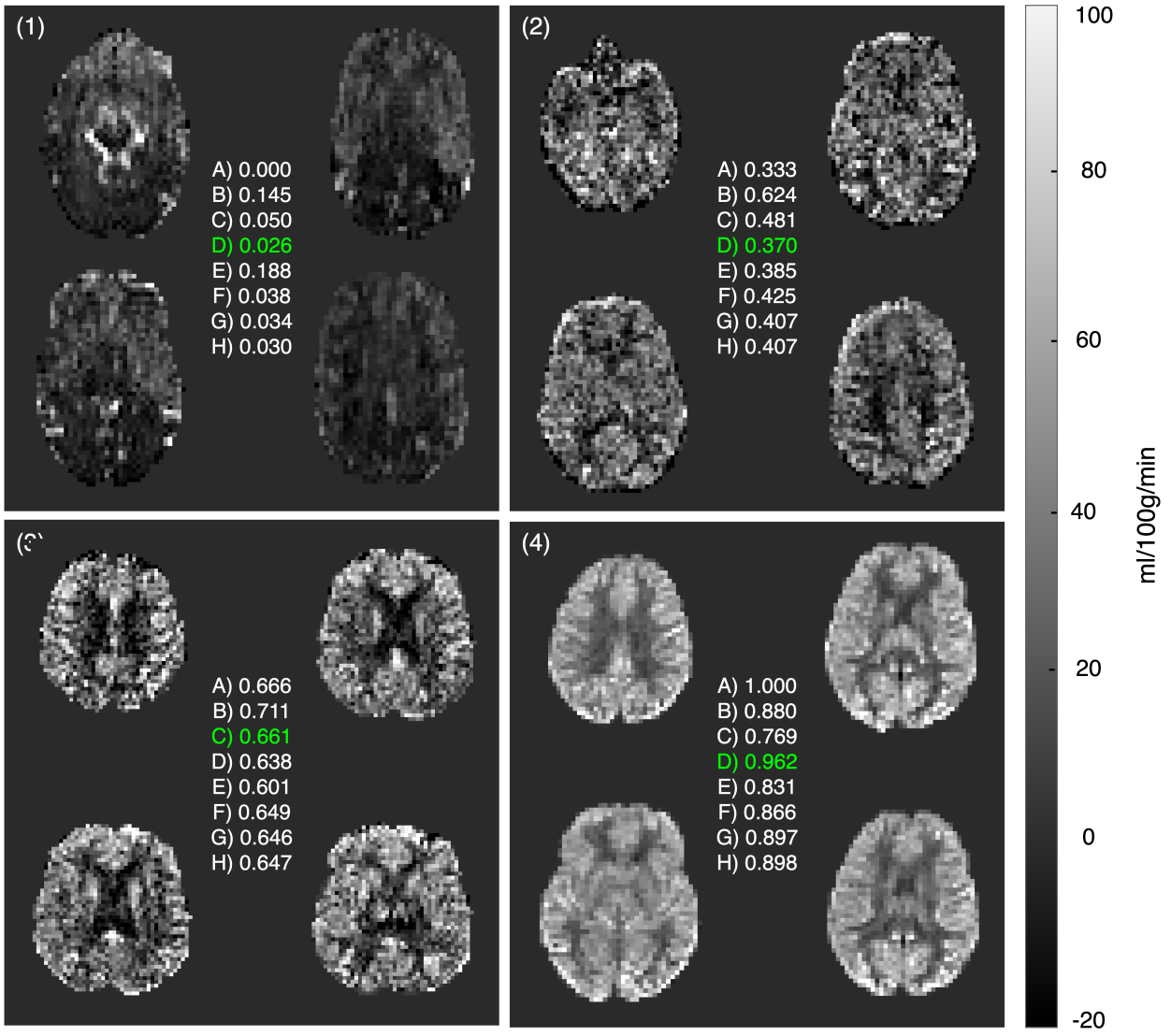


Figure 9: : Example of ASL CBF Maps with (1) Unacceptable quality (Rating 1) (2) Poor quality (Rating 2) (3) Average Quality (Rating 3) and (4) Excellent Quality (Rating 4) from the test set. Each example includes the QEI prediction for each of the presented approaches. A,B,C,D,E,F,G,H correspond to the average ratings of the raters, Dolui et al. (2024), 7FCN-QEI-Net, Reg-QEI-Net, MSC-QEI-Net, Ensemble 1, Ensemble 2, and Ensemble 3, respectively.

5. Discussion

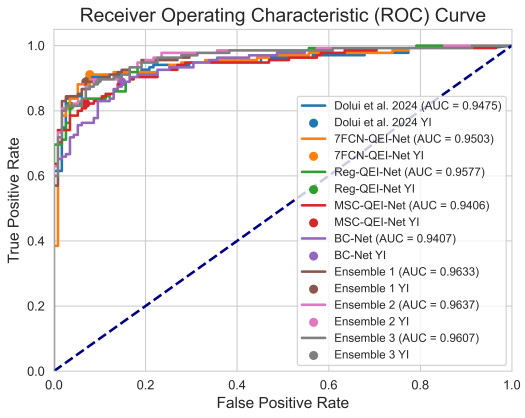
In this work, we developed several automated QEIs of ASL CBF maps by leveraging DL techniques. We improved the current state-of-the-art method (Dolui et al. (2024)) by introducing four new features and using them to train an FCN. While this method already surpassed the performance of (Dolui et al. (2024)), its limitations in the number of features and lack of automation prompted the exploration of other possibilities. To automate the feature extraction process, we developed multiple CNN approaches. These models outperformed the previous results, demonstrating the superiority of CNNs in finding better feature representations. Note that these methods only used the CBF map as input and did not require a structural image, unlike the 7-FCN-Net method, which extracts features from different tissue types. We also considered ensembles of some of the individual approaches, however, Reg-QEI-Net provided results comparable to the ensembled approaches, and is therefore our recommendation to be used clinically or in research.

5.1. Quality assessment methods

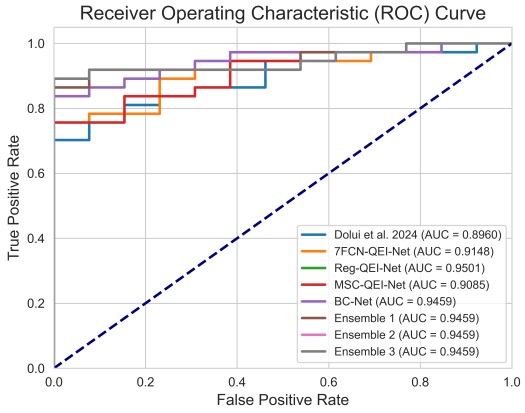
Table 2 and Table 3 present detailed comparisons of all the proposed approaches against the current state-of-the-art method (Dolui et al. (2024)) for the validation and the test sets, respectively. These results show that all the DL methods agree with the manual ratings. Specifically, the automated measures correlated better with the average ratings than the inter-rater correlation.

While all the raters are highly experienced researchers at the forefront of ASL MRI, their agreement is not perfect, highlighting the inherent difficulty and subjectivity of this task. Although not tested explicitly as a part of this study, the intra-rater agreement is also not expected to be perfect, and the agreement can be lower with raters new to the field who have limited experience with ASL CBF maps. The automated rating, being an objective measure, has the advantage of perfect reproducibility, thus increasing scientific rigor and reliability. All the DL approaches outperformed the current state-of-the-art approach (Dolui et al. (2024)). The 7FCN-QEI-Net model incorporates more features and uses a better machine learning approach to fit to the training data than (Dolui et al. (2024)), which uses a relatively naïve approach to fit each feature separately and combine them subsequently. As mentioned before, the improvements are even more pronounced with the CNN-based approaches, as showcased in Table 2 and Table 3. The MSC-QEI-Net approach performed slightly worse than Reg-QEI-Net. While that can be simply due to the nature of the problem, which inputs and outputs continuous variables, other aspects could have affected the performance of the algorithm. For example, we are currently using a categorical focal cross-entropy loss function for model training, which helps in dealing with imbalanced datasets. It might be beneficial to implement a customized weighted categorical cross-entropy loss function, where predictions are weighted according to each rater's class distribution. This approach might better address the underrepresented classes and improve overall performance.

Following the implementation of CNN-based models, we developed ensemble approaches. The goal of combining these models is based on their fundamentally different natures. For example, one model consists of a FCN, while the others are CNNs designed for completely different tasks. As a result, their performance and feature vectors vary due to their individual strengths and weaknesses. This is illustrated in Figure 12, where the networks exhibit varying levels of difficulty in predicting different rating values. For instance, in the ADNI dataset, the 7FCN-QEI-Net and Reg-QEI-Net were less successful at predicting samples rated as 2 than samples with other ratings. In contrast, MSC-QEI-Net did not encounter significant issues with samples from this rating group. Instead, this network performed less effectively when predicting samples rated as 1 and 3. The ensemble methods aim to address this by combining predictions from all models, creating a single, more robust, and more accurate final prediction. Although theoretically sound, we only found a minor improvement in performance with this approach. However, we expect further improvement when we train our models with a wider variety of ASL data from different scanners in our future work (more details in the Future Work section below).



(a)



(b)

Figure 11: ROC Curve of the different approaches compared in this study corresponding to the (a) validation and the (b) test set.

5.2. Identifying unacceptable quality CBF Maps

Once the QEI has been obtained from the presented methods, to exclude unacceptable CBF maps, we have presented recommendations for cut off values based on the YI, which optimizes both sensitivity and specificity. However, in research studies, the preference for higher sensitivity or higher specificity may vary depending on the specific task and the type of ASL data that is used. For example, a research study dealing with poor ASL data can drastically reduce its sample size using the optimal cutoff value. Therefore, it may be beneficial for such a study to lower the cutoff value to preserve enough data for analysis. On the other hand, a study dealing with state-of-the-art ASL data, or having a very large sample size, can use a higher QEI threshold to preserve only the ASL data with the best quality. Since the QEI produces a continuous number between 0 and 1, this provides the researcher flexibility to choose a threshold depending on the ASL data.

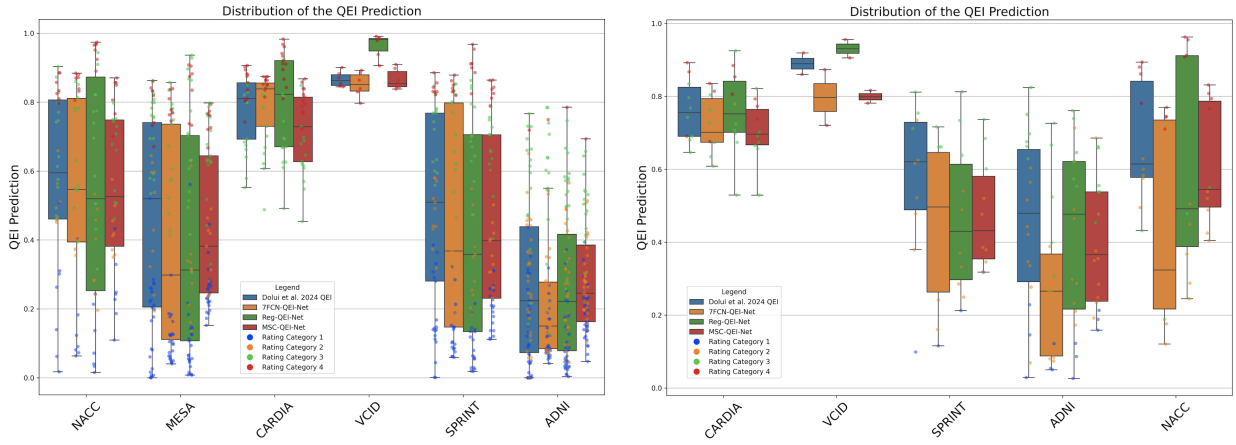


Figure 12: The QEI values across studies for both the (a) validation and the (b) test set.

5.3. Interpreting the heatmaps: artifact detection

The QEI presented in this study represents an estimate of the overall CBF map quality. A mediocre QEI value indicates that there are artifacts in the image, but it does not specify their location. The CNN-based QEI models do not provide a direct explanation for providing a low or high QEI, as the features are automatically extracted. A heatmap generated by one of their convolutional layers, however, can provide such information. This heatmap could be used for region of interest analysis. For example, for CBF maps with mediocre QEI values, the heatmaps can be used to create regions of unreliable CBF maps that can be discarded from statistical analysis. As shown in Figure 13, the higher intensities in the heatmap in artifactual CBF maps coincide with the region of artifacts. In samples free of artifacts, the network typically focuses on the GM and WM areas, where CBF is most relevant and significant. In the presented artifact-free case, the network has focused on the mentioned regions but has also shown a special interest in the right occipital lobe, identifying a potential source of artifact. This sample was originally rated as a 4 (free of artifacts) by two raters and as a 3 by the third rater. After discussing this case with the two raters who rated it as a 4, they agreed that the image might include a small amount of transit artifacts in the highlighted area. Due to their extensive expertise in ASL, the two raters knew that the protocol used for this sample was a single PLD. This protocol minimizes the transit artifact but does not eliminate it. Therefore, they concluded that this image was of very high quality (rating 4) considering the protocol used in the acquisition. The network QEI for this sample was 0.9057. This demonstrates the high correlation between the network's assessments and those of the raters, while also showcasing the network's potential ability to detect even the smallest artifacts.

5.4. Limitations

This study has several limitations. First, the ASL data that was used for this study was all acquired with Siemens scanners. Hence, although the study utilizes different ASL methods, there can potentially be further variability due to differences across MRI vendor platforms that were not captured by the models and need to be studied in the future. Second, the models were trained with a very limited sample size. This study is the beginning of a 5-years project funded by the National Institutes of Health (NIH) and eventually the models will be trained with a much larger sample size, including data obtained with other scanning platforms. Third, this study did not cover all possible artifacts or disease types because of limited availability; the dataset will be expanded in the future phase of the project. Fourth, we had 3 raters who rated the images on 4 scales, which led to a limited range of unique numbers when averaged. Some raters expressed that, for certain images, they were unsure between two rating levels and would have preferred more options rather than being forced to choose one that they did not fully agree with. Therefore, it would be beneficial for this task to extend the current rating levels to a wider range, which would provide more options to the raters and a richer representation of the network's ground truth, thereby improving the model's ability to learn and perform accurately. Finally, we had only 3 raters who rated the images. Incorporating additional raters to rate the images can generalize the QEIs, as different raters might have different sensitivities to different types of artifacts.

5.5. Future Work

Despite achieving state-of-the-art results, there is potential for further improvement by addressing the limitations mentioned above. First, we will aim to train the models by using a much larger dataset encompassing

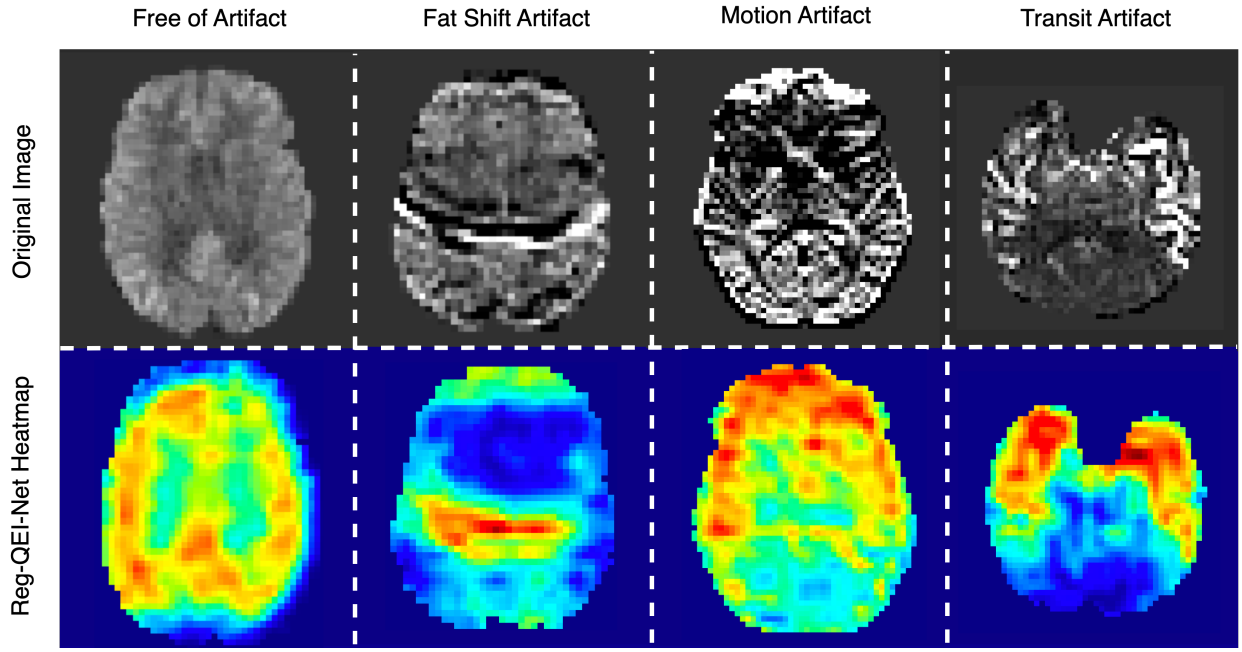


Figure 13: Example of Reg-QEI-Net heatmap visualizations applied to various samples with different sources of artifacts.

a wider variety of ASL protocols, more scanning platforms (e.g., GE and Philips), a wider type of artifact, data from patients with different diseases, and images rated by a greater number of raters. By doing so, the diversity and relevance of the training data can be increased, leading to improved network performance and robustness. Second, we will apply the QEI to actual research studies to assess improvements in statistical tests of group differences. Third, we will use heatmaps to identify regions of unreliable CBF maps and apply that to ROI analysis to assess if that improves statistical results. Lastly, although the current QEI-Net approach demonstrates high performance in assessing the quality of the CBF map, it does not give information about the source of artifacts. To address this, a CNN model aimed at classifying different sources of artifacts could be implemented, that can be used in studies to modify or correct errors in data acquisition protocols. For this approach, the heatmaps obtained from the QEI-Net architecture could serve as ROI extractors, enhancing the network's ability to focus on more meaningful areas of the brain. This improvement would not only increase the interpretability of the results but also provide valuable insights into the types of artifacts affecting the quality of the maps, ultimately contributing to better diagnostic outcomes and model transparency.

6. Conclusions

In this study, we designed, optimized, and validated multiple automated QEIs for ASL-derived CBF maps using DL techniques. The methods perform comparably to manual quality assessments and can rapidly provide an objective quality evaluation that can be used in research studies. These methods can also be incorporated into clinical and research scanners and provide real-time feedback to the scanner technicians that can be used to repeat the scans while the patient or study participant is still in the scanner. The automated QEI is expected to facilitate scientific rigor and reproducibility in research studies.

Acknowledgments

I am deeply grateful to my supervisors, Dr. Sudipto Dolui and Dr. John A. Detre, for their trust, insightful feedback, and the freedom they provided throughout this project. I extend my heartfelt thanks to my family and friends for their unwavering emotional support, despite the physical distances between us. I am also sincerely appreciative of the MAIA master consortium for granting me this life-changing opportunity. Finally, this endeavor would not have been possible without the financial support of the National Institutes of Health (NIH) Grant R21AG080518.

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.

Austin, T.R., Nasrallah, I.M., Erus, G., Desiderio, L.M., Chen, L.Y., Greenland, P., Harding, B.N., Hughes, T.M., Jensen, P.N., Longstreth Jr, W., Post, W.S., Shea, S.J., Sitlani, C.M., Davatzikos, C., Habes, M., Bryan, R.N., Heckbert, S.R., 2024. Association of brain volumes and white matter injury with race, ethnicity, and cardiovascular risk factors: The multi-ethnic study of atherosclerosis .

Bengio, Y., Courville, A., Vincent, P., 2013. Representation learning: A review and new perspectives. IEEE Trans Pattern Anal Mach Intell 35, 1798–1828.

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)f-fdg-pet to identify mild cognitive impairment. NeuroImage Clinical 25, 102146. doi:10.1016/j.nicl.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., Wang, D.J., et al., 2016. Comparison of non-invasive mri measurements of cerebral blood flow in a large multisite cohort. Journal of Cerebral Blood Flow & Metabolism 36, 1244–1256. doi:10.1177/0271678X16646124.

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.

Ewing, J., Cao, Y., Knight, R., Fenstermacher, J., 2005. Arterial spin labeling: validity testing and comparison studies. Journal of Magnetic Resonance Imaging 22, 737–740. doi:10.1002/jmri.20451.

Fallatah, S., Pizzini, F., Gómez-Anson, B., Magerkurth, J., Vita, E., Bisdas, S., Jäger, H., Mutsaerts, 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.

Friston, K., Williams, S., Howard, R., Frackowiak, R., Turner, R., 1996. Movement-related effects in fmri time-series. Magn Reson Med 35, 346–355.

Glorot, X., Bengio, Y., 2010. Understanding the difficulty of training deep feedforward neural networks, in: Teh, Y.W., Titterington, M. (Eds.), Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, PMLR, Chia Laguna Resort, Sardinia, Italy. pp. 249–256. URL: https://proceedings.mlr.press/v9/glorot10a.html.

Gong, E., Guo, J., Liu, J., Fan, A., Pauly, J., Zaharchuk, G., 2020. Deep learning and multi-contrast based denoising for low-snr arterial spin labeling (asl) mri, in: Medical Imaging 2020: Image Processing, SPIE. p. 113131K. doi:10.1117/12.2549765.

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.

He, S., Feng, Y., Grant, P., Ou, Y., 2022. Deep relation learning for regression and its application to brain age estimation. IEEE Transactions on Medical Imaging 41, 2304–2317. doi:10.1109/TMI.2022.3161739.

Heijtel, D., Mutsaerts, H., Bakker, E., Schober, P., Stevens, M., Petersen, E., van Berckel, B., Majoie, C., Booij, J., van Osch, M., Vanbavel, E., Boellaard, R., Lammertsma, A., Nederveen, A., 2014. Accuracy and precision of pseudo-continuous arterial spin labeling perfusion during baseline and hypercapnia: a head-to-head comparison with (1)(5)o h(2)o positron emission tomography. NeuroImage 92, 182–192. doi:10.1016/j.neuroimage.2014.02.011.

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.

Kim, K.H., Choi, S.H., Park, S.H., 2018. Improving arterial spin labeling by using deep learning. Radiology 287, 658–666. doi:10.1148/radiol.2017171154.

Koenig, M., Klotz, E., Luka, B., Venderink, D., Spittler, J., Heuser, L., 1998. Perfusion ct of the brain: diagnostic approach for early detection of ischemic stroke. Radiology 209, 85–93. doi:10.1148/radiology.209.1.9769817.

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.

Leonardsen, E., Peng, H., Kaufmann, T., Agartz, I., Andreassen, O., Celius, E., Espeseth, T., Harbo, H., Høgestøl, E., Lange, A., Marquand, A., Vidal-Piñeiro, D., Roe, J., Selbæk, G., Sørensen, Y., Smith, S., Westlye, L., Wolfers, T., Wang, Y., 2022. Deep neural networks learn general and clinically relevant representations of the ageing brain. NeuroImage 256, 119210. doi:10.1016/j.neuroimage.2022.119210.

Li, Y., Dolui, S., Xie, D., Wang, Z., Initiative, A., 2018. Priors-guided slice-wise adaptive outlier cleaning for arterial spin labeling perfusion mri. Journal of Neuroscience Methods 307, 248–253. doi:10.1016/j.jneumeth.2018.06.007.

Li, Y., Liu, P., Li, Y., et al., 2019. Asl-mricloud: An online tool for the processing of asl mri data. NMR in Biomedicine 32, e4051. doi:10.1002/nbm.4051.

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.

Peng, H., Gong, W., Beckmann, C., Vedaldi, A., Smith, S., 2021. Accurate brain age prediction with lightweight deep neural networks. Medical Image Analysis 68, 101871. doi:10.1016/j.media.2020.101871.

Plis, S., Hjelm, D., Salakhutdinov, R., Allen, E., Bockholt, H., Long, J., Johnson, H., Paulsen, J., Turner, J., Calhoun, V., 2014. Deep learning for neuroimaging: a validation study. Frontiers in Neuroscience 8, 229. doi:10.3389/fnins.2014.00229.

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.

Qin, Q., Alsop, D., Bolar, D., Hernandez-Garcia, L., Meakin, J., Liu, D., Nayak, K., Schmid, S., van Osch, M., Wong, E., Woods, J., Zaharchuk, G., Zhao, M., Zun, Z., Guo, J., Group, I., 2022. Velocity-selective arterial spin labeling perfusion mri: A review of the state of the art and recommendations for clinical implementation. Magnetic Resonance in Medicine 88, 1528–1547.

Rempp, 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.

Ruopp, M., Perkins, N., Whitcomb, B., Schisterman, E., 2008. Youden index and optimal cut-point estimated from observations affected by a lower limit of detection. Biometrical Journal 50, 419–430. doi:10.1002/bimj.200710415.

Sadaghiani, S., Tackett, W., Tisdall, M., Detre, J., Dolui, S., 2023. Reliability of periventricular white matter cerebral blood flow using different asl protocols. Proceedings on CD-ROM - International Society for Magnetic Resonance in Medicine. Scientific Meeting and Exhibition/Proceedings of the International Society for Magnetic Resonance in Medicine, Scientific Meeting and Exhibition doi:10.58530/2022/4875.

Selvaraju, R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D., 2017. Grad-cam: Visual explanations from deep networks via gradient-based localization. Proceedings of the IEEE International Conference on Computer Vision (ICCV) , 618–626.

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.

Taso, M., Aramendía-Vidaurreta, V., Englund, E., Francis, S., Franklin, S., Madhuranthakam, A., Martirosian, P., Nayak, K., Qin, Q., Shao, X., Thomas, D., Zun, Z., Fernández-Seara, M., Group, I., 2023. Update on state-of-the-art for arterial spin labeling (asl) human perfusion imaging outside of the brain. Magnetic Resonance in Medicine 89, 1754–1776. doi:10.1002/mrm.29609.

Wang, Z., Das, S.R., Xie, S.X., et al., 2013. Arterial spin labeled mri in prodromal alzheimer's disease: A multi-site study. NeuroImage: Clinical 2, 630–636. doi:10.1016/j.nicl.2013.04.014.

Williams, D., Detre, J., Leigh, J., Koretsky, A., 1992. Magnetic resonance imaging of perfusion using spin inversion of arterial water. Proc Natl Acad Sci U S A 89, 212–216. doi:10.1073/pnas.89.1.212.

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.

Woods, J., Achten, E., Asllani, I., Bolar, D., Dai, W., Detre, J., Fan, A., Fernandez-Seara, M., Golay, X., Gunther, M., Guo, J., Hernandez-Garcia, L., Ho, M., Juttukonda, M., Lu, H., MacIntosh, B., Madhuranthakam, A., Mutsaerts, H., Okell, T., Parkes, L., Pinter, N., Pinto, J., Qin, Q., Smits, M., Suzuki, Y., Thomas, D., Van Osch, M., Wang, D., Warnert, E., Zaharchuk, G., Zelaya, F., Zhao, M., Chappell, M., Group, I., 2024. Recommendations for quantitative cerebral perfusion mri using multi-timepoint arterial spin labeling: Acquisition, quantification, and clinical applications. Magnetic Resonance in Medicine 92, 469–495.

Xie, D., Li, Y., Yang, H., et al., 2020. Denoising arterial spin labeling perfusion mri with deep machine learning. Magnetic Resonance Imaging 68, 95–105. doi:10.1016/j.mri.2020.01.005.

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.

Zhang, L., Xie, D., Li, Y., et al., 2022. Improving sensitivity of arterial spin labeling perfusion mri in alzheimer's disease using transfer learning of deep learning-based asl denoising. Journal of Magnetic Resonance Imaging 55, 1710–1722. doi:10.1002/jmri.27984.

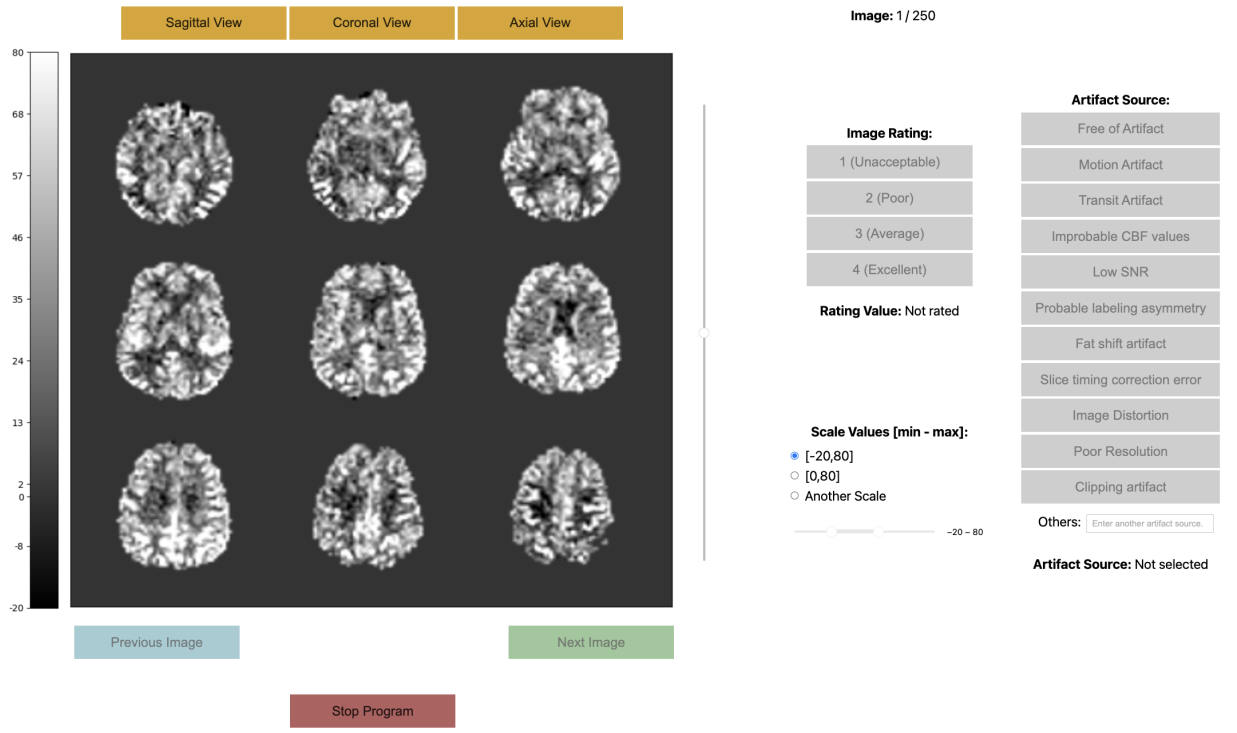


Figure 14: Example of the ASL CBF rating tool.

Appendix A: ASL CBF Rating Tool

Here, we detail the functionality and features of the web-based ASL CBF rating tool developed to simplify the rating task. This tool is a Python notebook designed to be used with Google Collaboratory, 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/ASL-CBF-Rating-Tool