

DEEP LEARNING BASED TECHNIQUES FOR THE REDUCTION OF POST-LABELLING DELAY IN MULTI-PLD ASL MRI

A Project Report

*Submitted to the APJ Abdul Kalam Technological University in partial
fulfillment of requirements for the award of degree*

Bachelor of Technology

in

Computer Science and Engineering

by

ANLIYA JOY (TKM21CS024)

IRISH ANN RAJAN (TKM21CS064)

NANDANA V (TKM21CS092)

NINO JAGADISH (TKM21CS099)



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

TKM COLLEGE OF ENGINEERING, KOLLAM

APRIL 2025

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

TKM COLLEGE OF ENGINEERING, KOLLAM

2024 - 2025



CERTIFICATE

This is to certify that the report entitled **DEEP LEARNING BASED TECHNIQUES FOR THE REDUCTION OF POST-LABELLING DELAY IN MULTI-PLD ASL MRI** submitted by **ANLIYA JOY (TKM21CS024)**, **IRISH ANN (TKM21CS064)**, **NANDANA V (TKM21CS092)** and **NINO JAGADISH (TKM21CS099)** to APJ Abdul Kalam Technological University in partial fulfillment of the B.Tech. degree in Computer Science and Engineering is a bonafide record of the project work carried out by him/her under our guidance and supervision. This report in any form has not been submitted to any other University or Institute for any purpose.

Internal Supervisor

Head of the Department

Project Coordinator

External Examiner

DECLARATION

We hereby declare that the project report entitled **DEEP LEARNING BASED TECHNIQUES FOR THE REDUCTION OF POST-LABELLING DELAY IN MULTI-PLD ASL MRI**, submitted for partial fulfillment of the requirements for the award of degree of Bachelor of Technology of the APJ Abdul Kalam Technological University, Kerala is a bonafide work done by us under supervision of Dr. MANU J PILLAI, Project Coordinator and Associate Professor, Department of Computer Science and Engineering, TKMCE and Dr. SHYNA A, Assistant Professor, Department of Computer Science and Engineering, TKMCE. This submission represents our ideas in our own words and where ideas or words of others have been included, we have adequately and accurately cited and referenced the original sources. We also declare that we have adhered to ethics of academic honesty and integrity and have not misrepresented or fabricated any data or idea or fact or source in my submission. We understand that any violation of the above will be a cause for disciplinary action by the institute and/or the University and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been obtained. This report has not been previously formed the basis for the award of any degree, diploma or similar title of any other University.

Kollam
26-03-2025

ANLIYA JOY
IRISH ANN RAJAN
NANDANA V
NINO JAGADISH

ACKNOWLEDGEMENT

We take this opportunity to express our deepest sense of gratitude and sincere thanks to everyone who helped us complete this work successfully.

We express our sincere gratitude to **Dr. Sajeeb R**, Principal, TKMCE, for providing us with all the necessary facilities and support for conducting the seminar.

We extend our sincere thanks to **Dr. Aneesh G Nath**, Head of Department of Computer Science and Engineering, TKM College of Engineering, Kollam, for providing us with all the necessary facilities and support.

We would like to place on record our sincere gratitude to our project coordinator, **Dr. Manu J Pillai**, Associate Professor, Department of Computer Science and Engineering, TKMCE, and our project guide, **Dr. Shyna A**, Assistant Professor, Department of Computer Science and Engineering, TKMCE, for their invaluable guidance and mentorship throughout this work.

We also extend our sincere appreciation to **Dr. Kesavadas C**, Professor and Head of Department of Radiology, Sree Chitra Tirunal Institute for Medical Sciences & Technology (SCTIMST), Thiruvananthapuram, for providing valuable insights and suggestions regarding multi-PLD ASL MRI, which greatly contributed to our research.

Finally, we thank our family and friends who contributed to the successful fulfillment of this project.

ANLIYA JOY

IRISH ANN RAJAN

NANDANA V

NINO JAGADISH

ABSTRACT

Arterial Spin Labeling (ASL) is a cutting-edge MRI technique used to measure Cerebral Blood Flow (CBF) and Arterial Transit Time (ATT), key parameters for assessing brain function and diagnosing neurological conditions. ASL quantifies CBF, which represents the volume of blood supplied to the brain per unit of time, and ATT, which refers to the time it takes for labelled arterial blood to travel to brain tissue. Both CBF and ATT are vital for detecting disorders like stroke, Alzheimer's, and other vascular diseases, as they provide critical insights into cerebral perfusion. A crucial component of ASL is the Post-Labeling Delay (PLD)—the time between the labelling of arterial blood and the acquisition of the MRI image. This timing influences the accuracy of CBF and ATT measurements. Currently, ASL is often performed using a single PLD, which assumes a uniform ATT across all patients. However, this method has a significant drawback: it overlooks individual variations in ATT, which can lead to inaccurate or biased CBF estimates, potentially affecting diagnosis. To address this limitation, the use of multi-PLD in ASL has been proposed. Multi-PLD approaches involve capturing ASL signals at multiple time points, which better accounts for variability in ATT and enables more accurate and reliable estimations of both CBF and ATT. Traditionally, these parameters are estimated using non-linear fitting methods, which involve adjusting a mathematical model to fit the observed data. While effective, these techniques are computationally intensive and prone to inaccuracies, especially when dealing with noisy or incomplete data. Emerging trends in deep learning are transforming the field by offering new ways to estimate CBF and ATT. Deep learning architectures, such as Convolutional Neural Networks (CNNs), can automatically learn complex patterns from multi-PLD ASL data, significantly improving the accuracy and efficiency of these estimations. Unlike traditional methods, deep learning can handle large datasets and learn subtle variations in blood flow dynamics, making it a promising alternative for clinical use. A novel deep learning architecture is proposed to estimate CBF and ATT from ASL data. By incorporating multi-PLD data into a deep learning framework, the study aims to overcome the limitations of single-PLD techniques and traditional non-linear fitting, paving the way for more accurate and reliable diagnostic tools in neuroimaging.

CONTENTS

Acknowledgement	i
Abstract	ii
List of Figures	v
List of Tables	vi
List of Abbreviations	vii
1. Introduction	1
1.1. Motivation	2
1.2. Problem Statement	3
1.3. Objectives	3
1.4. Organization of the report	4
2. Literature Survey	5
3. Deep Learning Based Techniques for the Reduction of PLD in Multi PLD ASL MRI	8
3.1. Proposed Methodology	8
3.2. Custom Model Generated as a Part of Methodology	14
3.3. Existing Model Implemented as a Part of Methodology	18
4. Experimental Result and Analysis	20
4.1 Evaluation Metrics	20
4.2. Performance Overview	21
4.3 Statistical Analysis	24

CONTENTS

5.	Conclusion	27
6.	References	29
7.	Appendix	32

LIST OF FIGURES

1.	Architecture Diagram of Proposed Model	9
2.	Perfusion Maps at Different PLDs for a Particular subject	12
3.	ResBlock	13
4.	DINO-TransformerRegNet	15
5.	VGG-Att3DUNet model	16
6.	CNN-Trans3DNet Model	17
7.	U-Net Model	18
8.	Bland-Altman Plot of ResED3D Vs Other Models for CBF_MSE and ATT_MSE	24
9.	Bland-Altman Plot of ResED3D Vs Other Models for CBF_PSNR and ATT_PSNR	25
10.	Bland-Altman Plot of ResED3D Vs Other Models for CBF_SSIM and ATT_SSIM	25
11.	Bland-Altman Plot of ResED3DV's Other Models for CBF_CCC and ATT_CCC	26

LIST OF TABLES

1. Quantitative Evaluation of Model Performance	22
2. Qualitative Comparison of Ground Truth and Predicted CBF & ATT Maps	23

LIST OF ABBREVIATIONS

- **ASL** : Arterial Spin Labeling
- **ATT** : Arterial Transit Time
- **CBF** : Cerebral Blood Flow
- **PLD** : Post Labeling Delay
- **MRI** : Magnetic Resonance Imaging
- **PWI**: Perfusion-Weighted Image
- **CNN** : Convolutional Neural Network
- **H-CNN**: Hierarchical Convolutional Neural Network
- **ResNet** : Residual Network
- **ROI** : Region of Interest
- **MSE** : Mean Squared Error
- **PSNR** : Peak Signal-to-Noise Ratio
- **SSIM** : Structural Similarity Index
- **CCC** : Concordance Correlation Coefficient
- **DINO** : Distillation with NO labels.

Chapter 1

Introduction

Arterial Spin Labeling (ASL) is a non-invasive MRI technique used to quantify Cerebral Blood Flow (CBF) and Arterial Transit Time (ATT), essential biomarkers for assessing neurological conditions such as stroke, dementia, and cerebrovascular diseases. Unlike traditional perfusion imaging methods that require exogenous contrast agents, ASL utilizes magnetically labeled arterial blood water as an endogenous tracer, providing a safer and repeatable alternative for clinical and research applications.[5]

ASL operates on the principle of label-control image acquisition. The arterial blood is magnetically labeled, and a control image is acquired without labeling. By subtracting the control image from the labeled image, a perfusion-weighted image (PWI) is obtained, which serves as the basis for CBF estimation. A critical parameter in this process is the PLD, which represents the time interval between labeling and image acquisition.

However, most ASL protocols rely on a single PLD as it simplifies data acquisition and processing. The drawback of this approach is that it assumes a uniform ATT across individuals, which can lead to CBF estimation errors, particularly in patients with altered cerebral hemodynamics. To address these limitations, researchers introduced the multi-PLD ASL approach, which acquires data at multiple PLDs to better account for variations in ATT and improve the accuracy of CBF quantification. Compared to single-PLD ASL, multi-PLD ASL provides additional physiological parameters, such as ATT estimation, leading to more accurate perfusion measurements. However, this method is time-consuming, requires longer scan durations, and increases the risk of motion artifacts, making widespread clinical adoption challenging.

Researchers have explored different methods to estimate CBF and ATT from multi-PLD ASL data, including kinetic model fitting and weighted delay techniques. While these methods improve estimation accuracy, they are computationally expensive and require extensive post-processing. Recently, deep learning techniques have emerged as a promising solution due to their capability to efficiently model complex perfusion dynamics. CNNs [1], hierarchical CNNs (H-CNNs) [2], U-Net [5], and physics-informed

unsupervised learning networks [13] have been extensively explored for ASL MRI processing, demonstrating improvements in accuracy and scan time reduction. These models leverage multi-PLD ASL data to learn intricate relationships between signal intensity and perfusion parameters, allowing for high-quality CBF and ATT estimation with minimal input data.

In this project, various deep learning models are being explored to further improve the efficiency of multi-PLD ASL processing by reducing the number of PLDs required while still generating high-quality CBF and ATT maps. The aim is to demonstrate that deep learning approaches can retain essential spatial information for accurate image reconstruction, even with reduced input data.

With an understanding of the challenges in the traditional methods used for CBF and ATT estimation in multi-PLD ASL technique, the next section discusses the motivation behind this work and the need for a custom deep learning architecture to achieve CBF and ATT maps with a reduced number of PLDs.

1.1 Motivation

The inspiration for this research comes from the shortfalls of existing ASL methods in measuring CBF and ATT, both of which are vital for diagnosing neurological disorders. Conventional approaches use a single PLD, with the presumption of equal ATT in all patients, resulting in inaccuracy in the measurement of CBF. This can compromise the diagnostic utility of ASL in clinical settings where accuracy is paramount for diagnoses such as stroke or dementia.

In addition, current approaches like non-linear fitting are time-consuming and have the potential for errors, again motivating the need for advancement. With increasing computing abilities of deep learning, much potential exists to overcome these limitations by using sophisticated architectures to more efficiently and accurately estimate CBF and ATT. This research suggests new deep-learning architectures to address these issues. By combining multi-PLD data into a deep learning system, the objective is to improve the accuracy of CBF and ATT measurements, providing more accurate tools for medical diagnosis and enhancing patient outcomes in the process.

Building on these motivations, the following section defines the specific challenge addressed in this work, outlining the core problem and its significance in ASL MRI.

1.2 Problem statement

Develop a deep learning model for ASL MRI that reduces the number of PLD required, thereby decreasing scan time and computational complexity, while maintaining or improving the accuracy of CBF and ATT estimation. Aiming to address limitations in current ASL techniques and surpass existing deep learning approaches in efficiency and precision, making ASL MRI more suitable for real-time clinical applications.

To address this problem this work establishes key objectives that guide the development of deep learning models for ASL MRI ensuring accurate and efficient CBF and ATT estimation.

1.3 Objectives

This study aims to improve CBF and ATT mapping in ASL MRI by developing deep-learning models that use fewer PLDs. The objectives focus on reducing scan time, maintaining accuracy, and enhancing clinical efficiency in neuroimaging.

- The model should reduce the ASL MRI scan time required for accurate CBF and ATT estimation. Shortening scan duration enhances patient comfort and improves clinical efficiency by minimizing the need for multiple PLD acquisitions.
- The model should employ advanced techniques of deep learning to resample an accurate CBF and ATT map from fewer samples acquired at PLD efficiently mitigating the problems concerning the undersampling and poor resolution.
- By drastically cutting the scanning time and the required resources, this approach opens the possibility of the greater application of multi-PLD ASL in clinical settings; these savings in volume constitute the foundation upon which to broaden ASL applications into everyday medical practice and will support fields ranging from neurology to radiology to emergency medicine.

With these objectives defined, the structure of this report is outlined in the following section, providing a detailed breakdown of each chapter and its contributions to the study.

1.4 Organisation of the Report

Chapter 2 Literature Survey, reviews existing research on Arterial Spin Labeling (ASL), focusing on Cerebral Blood Flow (CBF) and Arterial Transit Time (ATT) estimation techniques. It identifies the challenges and gaps in traditional approaches and highlights advancements in deep learning models applied to Multi-PLD ASL MRI. Chapter 3 Deep Learning Based Techniques for the Reduction of PLD in Multi-PLD ASL MRI, presents the proposed methodology, detailing the dataset collection, preprocessing techniques, deep learning model architecture, training, and evaluation strategies used for PLD reduction. Chapter 4 Experimental Analysis, discusses the experimental setup, performance evaluation metrics, and comparative analysis of the proposed system with conventional approaches. Chapter 5 Results and Discussion, analyzes the experimental results, highlighting the effectiveness of the proposed model in improving CBF and ATT accuracy while reducing PLD. Finally, Chapter 6 Conclusion and Future Work, summarizes the key findings, identifies the strengths and limitations of the proposed approach, and suggests potential directions for future research.

The following chapter reviews relevant literature, highlighting traditional approaches, their limitations, and the advancements in machine learning techniques that address these challenges. This analysis forms the foundation for the proposed methodology, ensuring a structured and informed approach to reducing PLD in Multi-PLD ASL MRI using deep learning.

Chapter 2

Literature Survey

In recent years, several deep learning approaches have been proposed to automate the generation of CBF and ATT maps from multi-delay ASL MRI. These methods aim to reduce scan time, improve accuracy, and minimize motion artifacts in clinical applications. This chapter reviews key studies that have contributed to the advancement of CBF and ATT estimation using deep learning techniques, highlighting their methodologies, advantages, limitations, and potential improvements.

Nicholas J. Luciw et al.[2] proposed a method to automate the generation of CBF and ATT maps using deep learning from multi-delay ASL MRI. Their approach involved using a U-Net and a CNN to estimate CBF and ATT by training models with distinct architectures. The input data consisted of a clinical dataset with ASL MRI acquired with six PLDs on 1.5T and 3T GE systems in 99 adults, including individuals with and without cognitive impairment. The models were trained and validated using 5-fold cross-validation to compare their performance. CNNs were applied to model the voxel-wise CBF and ATT, while a modified U-Net architecture provided better results by minimizing mean absolute error (MAE). Monte Carlo dropout was used to estimate model uncertainty and robustness.

The study demonstrated several advantages. U-Net outperformed CNN, with MAE values of 8.4 ± 1.4 mL/100 g/min for CBF and 0.22 ± 0.09 s for ATT. The mean processing time for U-Net was only 10.77 seconds, compared to 10 minutes and 41 seconds for the reference pipeline, indicating a significant reduction in computation time. Additionally, the model maintained its performance even when fewer PLD images were used, suggesting adaptability to different PLD schedules. However, the method also had some limitations. Despite the use of Monte Carlo dropout to estimate uncertainty, there was still variability in predictions across different datasets. Another drawback was the limited generalization to diverse populations, as the model's performance on a more diverse clinical population or varied acquisition sequences remains unexplored.

To address the limitations of prolonged scan times and variability in predictions, Zhang et al.[16] proposed a deep learning-based approach to reduce the number of PLDs

while accurately estimating CBF and ATT in multi-delay ASL MRI. Conventional ASL MRI requires multiple PLDs for accurate estimation, but this increases scan time, leading to a higher likelihood of motion artifacts and a reduction in signal-to-noise ratio (SNR) due to T1 decay of labeled spins.

To address these challenges, Zhang et al. trained two separate deep networks—one for estimating CBF and ATT using data from a single PLD and another for estimation with two PLDs. The models were trained and tested using the Human Connectome Project multiple-PLD ASL MRI dataset. The deep learning model used in this approach was a CNN, which effectively captured spatial features from ASL MRI data to estimate CBF and ATT. The results demonstrated that ATT and CBF can be reliably estimated even with just one PLD using the proposed deep networks. The proposed method achieved a MAE of 6.5 mL/100 g/min for CBF and 0.18 s for ATT when using a single PLD, and the accuracy improved further when using two PLDs, achieving a MAE of 5.2 mL/100 g/min for CBF and 0.12 s for ATT.

The approach had several advantages. It significantly reduced scan time, making it more feasible for clinical use while maintaining accurate estimation of CBF and ATT. Additionally, reducing the number of PLDs minimized motion artifacts, improving the robustness of the results. However, a key disadvantage was the dependence on deep networks trained on a specific dataset, which may not generalize well across different populations or scanner types without additional fine-tuning.

To overcome the limitations of long scan times in multi-PLD pCASL despite improved accuracy, Huang et al.[15] proposed a hierarchical 3D convolutional neural network (H-CNN) to estimate ATT and CBF maps from a reduced number of PLDs and averages, significantly reducing the overall scan time while maintaining accuracy. The dataset used in this study was a clinical dataset consisting of 48 subjects (38 females and 10 males) aged 56–80 years, with 45 subjects in the training group and 3 in the validation group. H-CNN estimates ATT through shallow shared layers and CBF through deeper layers of the network.

The proposed H-CNN model reduced the scan time of multi-PLD pCASL by enabling estimation from a smaller number of PLDs without significant discrepancies. The method was validated against conventional nonlinear model fitting using the MAE. The H-CNN achieved MAEs of 32.69 ms for ATT and 3.32 mL/100 g/min for CBF when using the full dataset with six PLDs and six averages. Furthermore, it demonstrated that using

only three of the six PLDs still provided reliable estimations, with MAEs of 231.45 ms for ATT and 9.80 mL/100 g/min for CBF. The primary advantage of this method was the substantial reduction in scan time while maintaining high accuracy. However, the disadvantage lies in the possibility of limited generalisability when using fewer PLDs compared to the full dataset.

To address the limited generalizability of the H-CNN, Ishida et al.[14] proposed a method comparing supervised and physics-informed unsupervised deep neural networks (DNNs) for estimating CBF and ATT in ASL MRI. The supervised DNN was trained using simulated data with known ground truth values, while the physics-informed unsupervised DNN incorporated a kinetic model (KM) of ASL, based on the modified Bloch equation, directly into the loss function, enabling training without ground truth.

The dataset used included simulated data for training and validation and in vivo data for real-world validation, where Rician noise was added to the original images to simulate realistic conditions. The supervised DNN demonstrated higher accuracy in CBF and ATT estimation with lower bias in CBF estimation, making it more reliable for clinical applications. The unsupervised DNN, on the other hand, did not require ground truth data, providing the advantage of applicability in scenarios where labeled data is unavailable. It also exhibited better noise immunity in ATT estimation. However, the supervised DNN required high-quality labeled data for training, making it resource-intensive, while the unsupervised DNN showed slightly higher positive bias in CBF estimation and its performance was more dependent on model assumptions.

Despite significant advancements in automating CBF and ATT estimation using deep learning models, several gaps remain in the literature. Many of the existing approaches suffer from limitations such as prolonged scan times, dependence on large training datasets, lack of generalizability to diverse populations. Additionally, while models such as supervised DNNs provide high accuracy, they require high-quality labeled data for training, which may not always be available in real-world clinical settings.

To address these gaps, the next chapter presents the proposed methodology to balance high accuracy and reduced scan time while maintaining robustness across different acquisition parameters and ensuring better generalizability across diverse populations.

Chapter 3

Methodology

Reducing scan time in ASL MRI, particularly for non-invasive CBF and ATT estimation, is critical for clinical applications. Current deep learning methods have minimized scan time by reducing the number of PLDs to as few as two or three, but this often compromises the accuracy of CBF and ATT measurements. Developing a model that enhances estimation accuracy without increasing scan duration is essential to advance clinical reliability and efficiency in ASL MRI diagnostics.

Incorporating models that can generalize well across varying patient conditions and scanner settings is crucial for real-world deployment. Robust performance across diverse datasets will ensure consistent and reliable perfusion measurements, ultimately supporting wider adoption of ASL MRI in both research and clinical environments.

This advancement not only enhances the feasibility of ASL MRI in routine clinical practice but also broadens its applicability to patient populations where prolonged scan times are impractical, such as pediatric or critically ill patients.

3.1 Proposed Methodology

A novel architecture, entitled "ResED3D"- Ordinary Differential Equation based encoder-decoder architecture, is introduced for the estimation of CBF and ATT. The comprehensive framework integrates specialized modules designed to efficiently estimate perfusion parameters, even with a reduced number of PLDs. A detailed diagram of the architecture is provided in Fig. 1, illustrating the sequential data flow through each module and emphasizing the technical rationale behind every design choice, ultimately presenting an innovative and efficient approach for CBF and ATT prediction using PWI inputs.

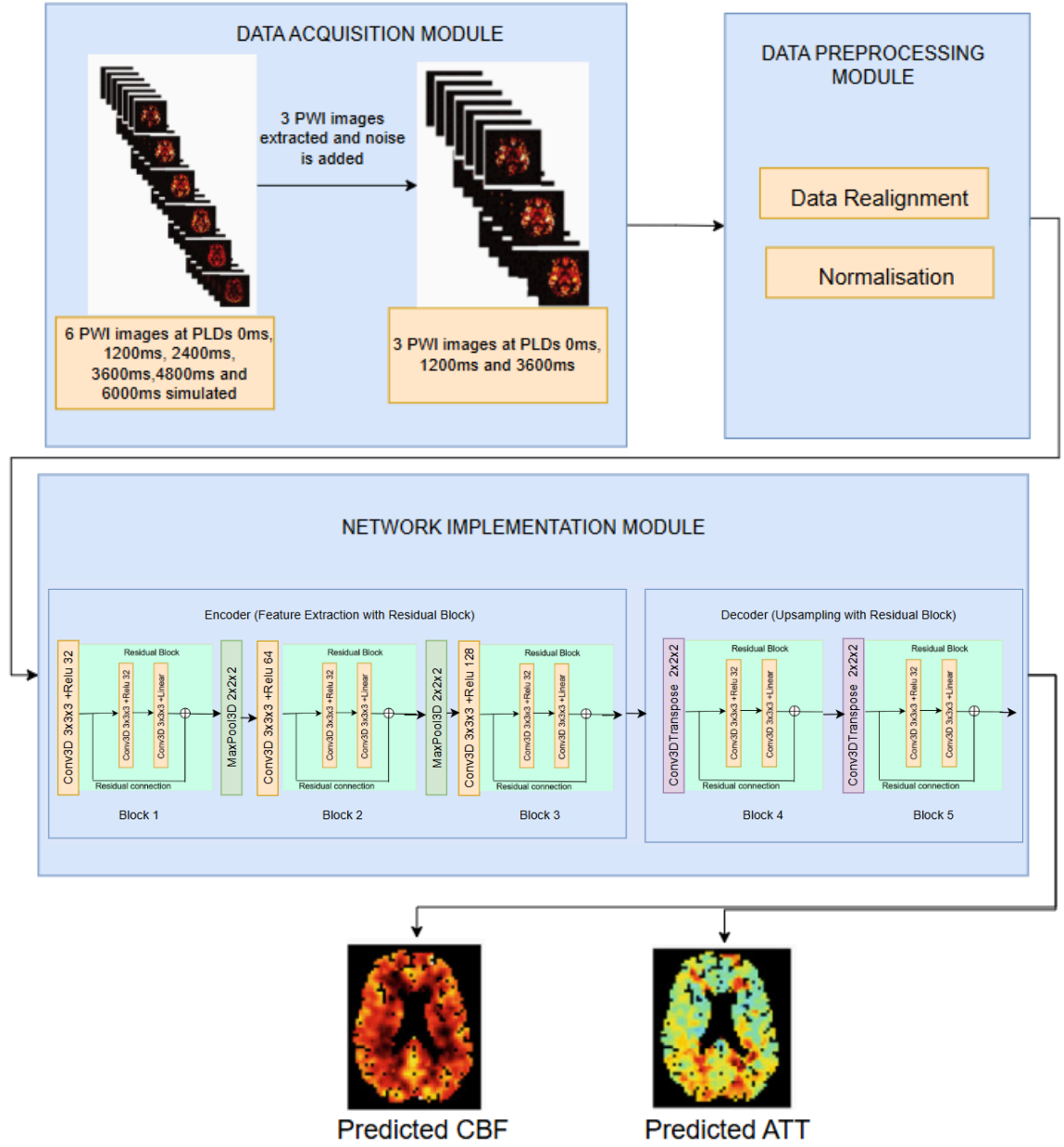


Fig.1 Architecture Diagram of Proposed Model

1. Data Acquisition Module

In this study, the data acquisition process involves collecting ASL MRI scans at multiple PLD times, allowing for the capture of blood flow dynamics over time. This multi-PLD approach supports the estimation of perfusion parameters such as CBF and ATT.

Since acquiring real multi-PLD ASL MRI datasets with multiple delay times is expensive, time-consuming, and requires specialized equipment, simulating the dataset was

opted.

The simulated dataset was generated by following a well-defined phantom algorithm designed to mimic realistic ASL MRI data, ensuring that the generated CBF and ATT values fall within clinically accepted ranges for gray matter and white matter of the brain. The ground truth CBF and ATT values were referenced from multiple peer-reviewed clinical articles [19][20] and journals, ensuring that the dataset reflects physiologically meaningful perfusion characteristics. The simulation involved six post-labeling delays (PLDs) at 0 ms, 1200 ms, 2400 ms, 3600 ms, 4800 ms, and 6000 ms. From these, three PLDs 0 ms, 1200 ms, and 3600 ms were selected to form the ASL standard curve, as they provide essential perfusion information. This selection was made to reduce the number of PLDs, thereby minimizing the overall scan time in a clinical setting while maintaining the accuracy of perfusion measurements. As a result, three PWI were generated per subject which were merged to form a final PWI image.

To enhance the realism of the dataset, Gaussian noise was added using standard deviation values derived from empirical studies, effectively simulating the acquisition noise typically observed in ASL MRI scans. This simulated dataset was used for training and evaluating the deep learning model. Additionally, spatial and temporal variations were incorporated to account for inter-subject differences, maintaining a diverse and representative dataset. This approach follows the method suggested by Maier et al.[18] for generating synthetic ASL data, which was adopted for the data simulation process. This allows for the evaluation of the deep learning model's performance in predicting CBF and ATT under realistic conditions.

Algorithm for Generating MultiPLD ASL Phantom with Noise Addition

Input: SNR_value (10,15,20)

Output: Simulated ASL phantom dataset with noise

```
1: procedure Generate_ASL_Phantom(SNR_value)
2:   data  $\leftarrow$  importdata('brainphantom.mat')
3:   mask  $\leftarrow$  importdata('mask_rt.mat')
4:   Define CBF_GM_range, CBF_WM_range, ATT_GM_range, ATT_WM_range,
   T1_GM, T1_WM
5:   Set interp_  $\leftarrow$  0.33, num_subjects  $\leftarrow$  24, num_pwi_images  $\leftarrow$  6
6:   PLD_values  $\leftarrow$  linspace(0, 6000, 6) / 60000
7:   for subject_id = 1 to num_subjects do
8:     Generate random CBF_GM, CBF_WM, ATT_GM, ATT_WM
```

```

9:   Initialize  $M0$ ,  $T1$ ,  $CBF$ ,  $ATT$  maps based on data, mask
10:  Downsample  $M0\_ds$ ,  $T1\_ds$ ,  $CBF\_ds$ ,  $ATT\_ds$  using interp_
11:  Set ASL parameters: params.T1t, params.T1b, params.lambda, params.alpha,
    params.PLD, params.tau
12:   $PWI\_images \leftarrow []$ 
13:  for pwi_idx = 1 to num_pwi_images do
14:    params_current  $\leftarrow$  params
15:    params_current.PLD  $\leftarrow$  params.PLD[pwi_idx]
16:    Generate  $PWI$  using CBF2PWI( $CBF\_ds$ ,  $ATT\_ds$ ,  $M0\_ds$ , params_current)
17:    Normalize  $PWI$  values:
18:     $PWI\_noisy \leftarrow \text{abs}(PWI)$ 
19:     $PWI\_normalized \leftarrow (PWI\_noisy - \min(PWI\_noisy)) / (\max(PWI\_noisy) -$ 
20:     $PWI\_noisy) \leftarrow \text{abs}(PWI)$ 
21:     $PWI\_normalized \leftarrow (PWI\_noisy - \min(PWI\_noisy)) / (\max(PWI\_noisy) -$ 
22:     $\min(PWI\_noisy))$ 
23:    Append  $PWI\_normalized$  to  $PWI\_images$ 
24:  end for
25:  Store full 6-PLD dataset in data_phantom_full
26:  Save index6_subject_{subject_id}_full.mat
27:  Select 3 PLDs (0s, 1200s, 3600s) and extract relevant  $PWI\_images$ 
28:  Save index3_subject_{subject_id}_selected.mat
29:   $\max\_deltaM \leftarrow \max(PWI\_images)$ 
30:   $SNR\_value \leftarrow SNR\_values(\text{randi}(\text{length}(SNR\_values)))$ 
31:   $\sigma \leftarrow \max\_deltaM / SNR\_value$ 
32:   $\text{noise} \leftarrow \sigma \times \text{randn}(\text{size}(\text{selected\_PWI\_images}))$ 
33:  Save index3_subject_{subject_id}_noisy_SNR{SNR_value}.mat
34: end for
35: end procedure

```

Algorithm for Generating PWI Images

Input: CBF_ds , ATT_ds , $M0_ds$, *params_current*

Output: Set of PWI images for given PLDs

```

1: procedure CBF2PWI( $CBF\_ds$ ,  $ATT\_ds$ ,  $M0\_ds$ , params_current)
2:    $\Delta M \leftarrow M0\_ds \times \text{params.alpha} \times \exp(-\text{params.PLD} / \text{params.T1b})$ 
3:    $\text{Flow\_term} \leftarrow (CBF\_ds / \text{params.lambda}) \times \exp(-ATT\_ds / \text{params.T1t})$ 
4:   if params.PLD >  $ATT\_ds$  then
5:      $PWI \leftarrow \Delta M \times \text{Flow\_term} \times (1 - \exp((ATT\_ds - \text{params.PLD}) /$ 
6:      $\text{params.T1t}))$ 
7:   else
8:      $PWI \leftarrow 0$ 
9:   end if

```

```

9: return PWI_images
10: end procedure

```

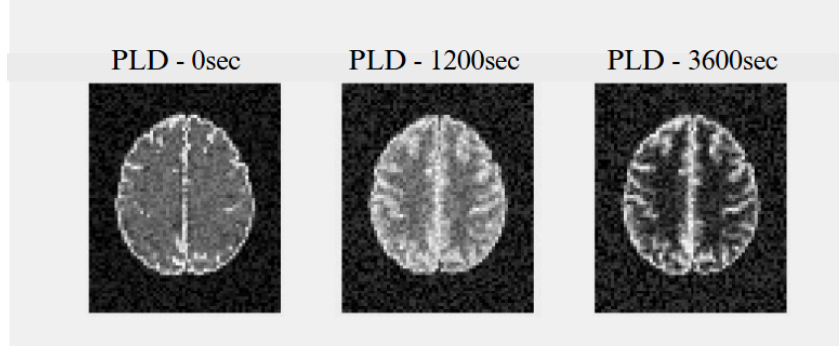


Fig.2 Perfusion Maps at Different PLDs for a Particular subject

Fig 2 illustrates the perfusion maps of slice 23 for a particular subject, captured at different PLDs—0 seconds, 1200 seconds, and 3600 seconds—highlighting the changes in cerebral blood flow over time.

The acquired data is preprocessed as discussed in the next section, which refines the acquired data to enhance its suitability for analysis

2. Preprocessing Module

This module is responsible for formatting and preparation of the pCASL image data for input into the neural network. The primary goal was to ensure that the PWI, CBF, and ATT data were properly formatted, normalized, and ready for model input. The following steps were performed:

- **Data Realignment:** The extracted data included PWI images, represented as a 4D array with dimensions (3, 60, 60, 72) indicating 3 channels with $60 \times 60 \times 72$ voxels. CBF ground truth, 3D array with dimensions (60, 60, 72), representing the target CBF values and ATT ground truth, 3D array with dimensions (60, 60, 72), representing the target ATT values. The PWI data was transposed to change the shape from (3, 60, 60, 72) to (60, 60, 72, 3) to match the expected input shape for the convolutional model.
- **Normalization:** All three datasets (PWI, CBF, and ATT) were normalized to the range [0, 1] using min-max normalization:

$$\text{Normalized data} = \frac{\text{data} - \min(\text{data})}{\max(\text{data}) - \min(\text{data})}$$

- Data Splitting: The data was split into training and testing sets using an 80:20 ratio. The split was performed as:

Training Data: $X_{\text{train}}, y_{\text{train_cbf}}, y_{\text{train_att}}$

Testing Data: $X_{\text{test}}, y_{\text{test_cbf}}, y_{\text{test_att}}$

- Batch and Input Preparation: The input data (PWI) and the corresponding output labels (CBF and ATT) were converted to NumPy arrays to facilitate efficient data processing during model training.

This comprehensive preprocessing approach ensured that the data was ready for training the deep learning model to accurately predict CBF and ATT maps.

With the data preprocessed and structured, the next step is to train and fine-tune machine learning models for CBF, ATT estimation. The following section details the training process, optimization techniques, and model-specific adaptations used to enhance prediction accuracy.

3. Network Implementation Module

The proposed model is a ResED3D model developed to automate the generation of CBF and ATT maps from multiple PLD ASL MRI data. This model achieves high accuracy compared to other architectures evaluated in this study. The ResED3D model follows an encoder-decoder architecture with residual connections to improve spatial information retention during feature reconstruction.

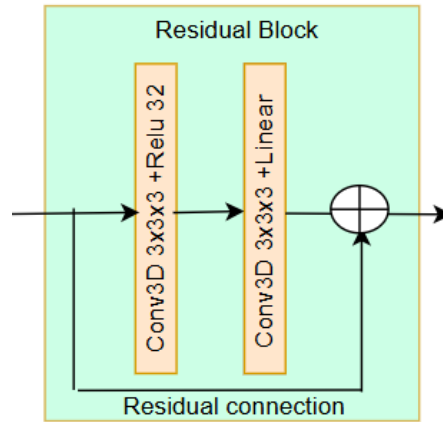


Fig.3 ResBlock

The encoder extracts hierarchical features through a series of 3D convolutional

layers with ReLU activations, batch normalization, and residual blocks that enhance feature representation. Max-pooling operations are applied after each convolutional stage to progressively reduce spatial dimensions and increase feature abstraction. The encoder gradually increases the number of filters from 32 to 128, capturing both low-level and high-level spatial features.

Each ResBlock consists of two 3D convolutional layers, where the first convolutional layer applies a ReLU activation to introduce non-linearity, and the second convolutional layer applies a linear activation for feature refinement. The output of the block is added to the original input through a residual connection, allowing the model to efficiently learn identity mappings and refine feature representations. This approach enhances spatial and contextual awareness, stabilizing feature learning and improving representation capability. The decoder mirrors the encoder's architecture, performing upsampling using transposed convolutions and incorporating ResBlocks to refine the reconstructed feature maps. The decoder progressively reduces the number of filters while restoring the original spatial resolution. The final feature maps pass through two separate 3D convolution layers with a filter size of (1,1,1) to generate the CBF and ATT maps.

Fine-tuning involves training the model on preprocessed PWI input data, with MSE loss for both CBF and ATT predictions. The model is trained for 50 to 2500 epochs with a batch size of 2 using the Adam optimizer with a learning rate of $1e-4$, allowing the model to iteratively refine its predictions and improve generalization.

As part of this study, multiple deep learning architectures were designed and evaluated for estimation of CBF and ATT. This is discussed in the following section.

3.2 Custom Model Generated as a Part of Methodology

The study involved designing and evaluating multiple deep learning architectures for CBF and ATT estimation. The outline of the alternative models considered during the experimentation are discussed below.

3.2.1 DINO-TransRegNet

In the medical imaging domain, acquiring large, well-annotated datasets poses significant challenges due to privacy constraints and the necessity of expert annotations. Conventional

models like HCNN are inherently dependent on supervised learning, making them less effective in scenarios where labeled data is scarce. To overcome this limitation, DINO-TransRegNet model (Self-Distillation with No Labels model-based Transformer Regression Network) was explored, a self-supervised vision transformer that learns meaningful feature representations without relying on extensive labeled datasets. Its ability to extract rich, hierarchical features through self-supervised learning makes it particularly well-suited for medical imaging applications, where annotated data is often limited and difficult to obtain.

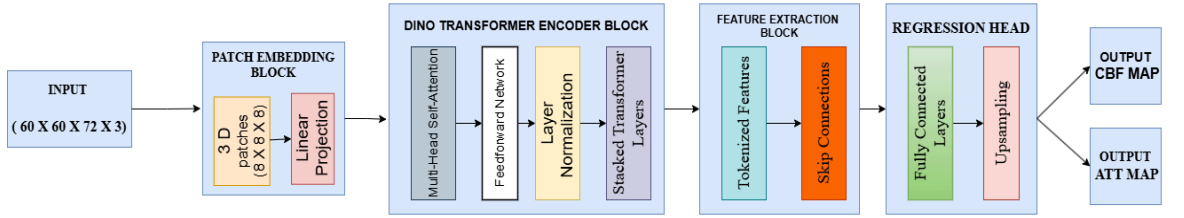


Fig.4 : DINO-TransRegNet

The DINO-TransRegNet is a self-supervised Vision Transformer (ViT) designed to learn meaningful feature representations without requiring labeled data. Unlike traditional convolutional networks, which rely on supervised learning, DINO-TransRegNet leverages a teacher-student training framework to learn representations through contrastive learning. This enables the model to discover rich, high-level features from images by aligning representations across different augmentations of the same input. Its ability to capture long-range dependencies and adapt to diverse image structures makes it highly effective for medical imaging tasks, where labeled datasets are often scarce.

The model uses a ViT backbone to process 3D patch embeddings. The embeddings are passed through a series of self-attention layers, capturing long-range dependencies between patches. The transformer layers process the patch embeddings using multi-head self-attention (MSA) and feedforward layers. Instead of using DINO's original contrastive loss, the model is fine-tuned for a regression task. The final feature representation is passed through two separate fully connected (FC) layers to generate predictions for: CBF , ATT maps. The output has the same spatial dimensions as the input images.

Fine-tuning involves training the model on preprocessed PWI input data, with MSE loss for both CBF and ATT predictions. The model is trained for 50 to 2500 epochs with a

batch size of 2 using the Adam optimizer with a learning rate of $1e-4$, allowing the model to iteratively refine its predictions and improve generalization.

3.3.2. VGG-Att3DUNet model

The VGG-Att3DUNet model follows an encoder-decoder architecture with integrated attention mechanisms to enhance feature extraction and reconstruction for CBF and ATT estimation. The model is designed to efficiently capture spatial dependencies in 3D medical imaging while preserving critical structural details through skip connections and attention refinement.

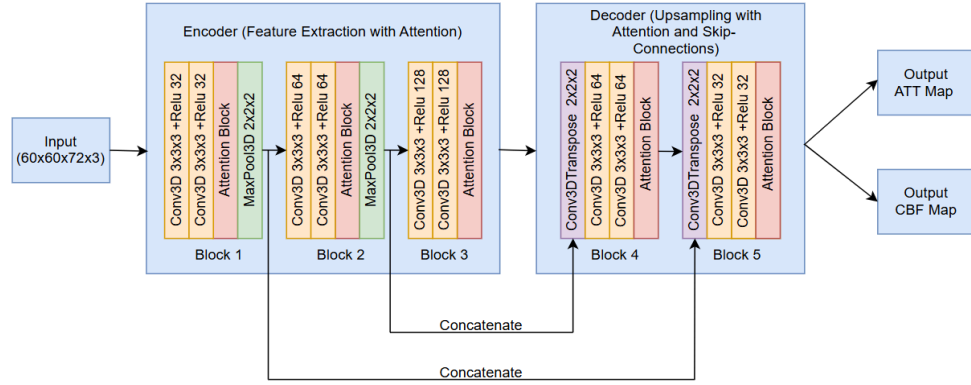


Fig 5. VGG-Att3DUNet model

The encoder consists of multiple 3D convolutional layers with ReLU activations, progressively increasing the number of filters from 32 to 128. Each convolutional block is followed by an attention block, which applies a convolution and a sigmoid activation to generate an attention mask, refining feature representations by emphasizing important spatial regions. Max-pooling layers are applied after certain blocks, reducing spatial dimensions while retaining essential information.

The decoder mirrors the encoder's structure, utilizing 3D transposed convolutions to upsample feature maps and restore spatial resolution. Skip connections concatenate encoder features with decoder outputs at corresponding levels, preserving fine-grained spatial details essential for precise CBF and ATT estimation. Attention mechanisms are also applied in the decoder to selectively enhance reconstructed feature maps. The final stage consists of two separate 3D convolutional layers, each producing a distinct output: one for CBF mapping and the other for ATT estimation. By integrating attention at

multiple levels, the model effectively suppresses irrelevant features and enhances critical regions, leading to improved predictive accuracy for perfusion parameter estimation.

The fine-tuning process involves training the model on preprocessed PWI data, optimizing CBF and ATT predictions using MSE loss. Training spans 50 to 2000 epochs with a batch size of 2, employing the Adam optimizer with an initial learning rate of $1e-4$ that progressively decreases over time.

3.3.4 CNN-Trans3DNet Model

A hybrid CNN-Transformer model was developed to estimate CBF and ATT from PWI data, for multi-PLD ASL MRI scans. The model integrates a 3D Convolutional Neural Network (CNN) for spatial feature extraction and a Transformer-based attention mechanism to capture long-range dependencies.

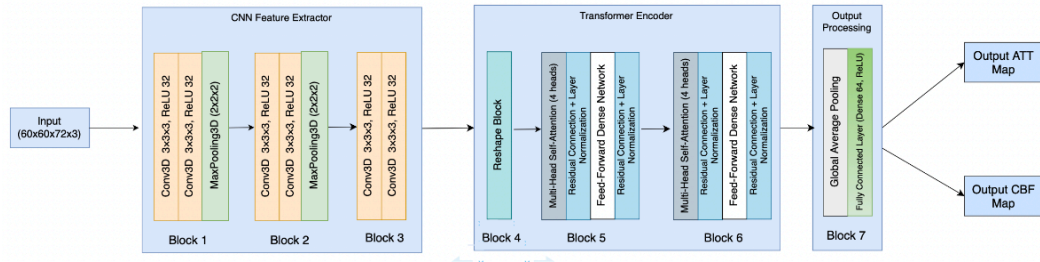


Fig 6. CNN-Trans3DNet Model

The architecture of the model consists of a CNN-Transformer hybrid design. The CNN-based feature extraction is achieved through a series of 3D convolutional layers with ReLU activation functions, followed by max-pooling layers to reduce the spatial dimensions. These features are then reshaped and passed through multi-head self-attention Transformer layers that capture global spatial dependencies. The model ends with a fully connected output layer, which predicts the CBF and ATT maps, reshaped to match the dimensions of the ground truth data.

The Transformer Block helps the model focus on important features by using Multi-Head Self-Attention (MHSA) to capture spatial and contextual details. It includes residual connections and Layer Normalization for stability, followed by a Feedforward Network (FFN) to refine the features. This improves how well the model understands and processes the input, enhancing accuracy in predicting CBF and ATT from PWI scans.

The model is trained from 50 epochs to 2500 epochs, with a batch size of 2, using

the Adam optimizer with a learning rate of $1e-4$. A variety of evaluation metrics, including accuracy and clinical applicability measures, ensure that the model's performance is comprehensively assessed, making it well-suited for real-world applications.

3.3 Existing Model Implemented as a Part of Methodology

The U-Net model used in this work is implemented based on the architecture proposed by Luciw et al.[13] in their research.

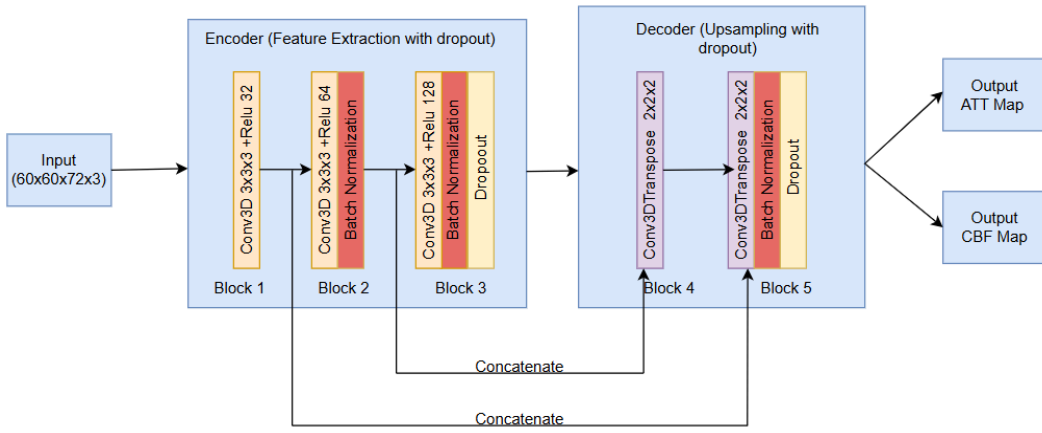


Fig 7. U-Net Model

The model follows a 3D U-Net architecture with downsampling and upsampling blocks, enhanced by skip connections to preserve spatial information during feature reconstruction as shown in Fig 5. The encoder extracts hierarchical features through a series of 3D convolutional layers with ReLU and LeakyReLU activations, batch normalization, and dropout layers that refine feature representations. Max-pooling and strided convolution operations progressively reduce spatial dimensions while increasing feature abstraction. The number of filters increases from the initial 32 to 128 to capture low- to high-level spatial information. Monte Carlo Dropout is applied at each stage to introduce randomness during training, enhancing feature transformation and improving spatial-contextual awareness. The decoder mirrors the encoder's architecture, using transposed convolutions for upsampling and incorporating skip connections to refine the reconstructed feature maps. It gradually reduces the number of filters while restoring the original spatial resolution. The final feature maps pass through two 3D convolutional layers to generate the CBF and ATT maps.

Fine-tuning involves training the model on preprocessed PWI data using Mean Squared Error (MSE) loss for both CBF and ATT predictions. The model is trained for 50 to 2500 epochs with a batch size of 2, using the Adam optimizer with an initial learning rate of $1e-4$, which gradually decays over time. To mitigate overfitting, MC Dropout is applied during training, ensuring that the model captures variability and uncertainty effectively. The training process refines the feature representations by iteratively improving the model's generalization capability, ensuring accurate CBF and ATT predictions on unseen data.

The next chapter provides a comprehensive examination of the experimental results, discussing key findings and the comparative performance of all the implemented models.

Chapter 4

Experimental Result and Analysis

In this chapter, a detailed analysis of the results is provided, comparing the accuracy of CBF and ATT estimation across experimented model architectures and its comparison with the proposed model.

4.1 Evaluation Metrics

The performance of the deep learning models—VGG-Att3DUNet, U-NET, ResED3D, DINO-TransRegNet, and CNN-Trans3DNet—was evaluated over multiple training epochs (50 to 2500) using four key metrics:

- MSE quantifies the average of the squared differences between predicted and actual values. Lower MSE indicates more accurate predictions, ensuring reliable estimation of perfusion parameters.

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

- y_i = Ground truth value
- \hat{y}_i = Predicted value
- N = Total number of samples

- PSNR measures the ratio between the maximum signal value and the noise in the reconstructed image. Higher PSNR suggests better reconstruction quality, indicating that the predicted CBF and ATT maps closely resemble the reference images.

$$PSNR = 10 \times \log_{10} \left(\frac{MAX_I^2}{MSE} \right)$$

- MAX_I = Maximum possible intensity value of the image
 - MSE = Mean Squared Error
- SSIM compares luminance, contrast, and structure between two images. A higher

SSIM value signifies that the model preserves the anatomical structure while predicting perfusion parameters.

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

- μ_x, μ_y = Mean of images x and y
 - σ_x^2, σ_y^2 = Variance of images x and y
 - σ_{xy} = Covariance between images
 - C_1, C_2 = Stabilizing constants
- CCC evaluates agreement between predicted and actual values, considering both precision and accuracy. Higher CCC demonstrates strong alignment with the ground truth, ensuring consistent and reliable predictions.

$$CCC = \frac{2\rho\sigma_x\sigma_y}{\sigma_x^2 + \sigma_y^2 + (\mu_x - \mu_y)^2}$$

- ρ = Pearson correlation coefficient
- σ_x, σ_y = Standard deviations of predicted and ground truth values
- μ_x, μ_y = Mean of predicted and ground truth values

These metrics assess the models ability to estimate CBF and ATT accurately while reducing the number of PLDs. The models are evaluated against the above-mentioned metrics, and their performances are compared in the next section.

4.2 Performance Overview

The proposed model, ResED3D consistently outperforms all other models across multiple evaluation metrics, establishing itself as the most accurate and reliable approach for perfusion parameter estimation. From the beginning of training, ResED3D maintains one of the lowest MSE values, and by 2500 epochs, it achieves the best overall performance with an MSE of 0.00227 for CBF and 0.00178 for ATT. This demonstrates its superior ability to minimize reconstruction errors, surpassing all competing models, including DINO-TransRegNet and U-NET.

In terms of PSNR, ResED3D leads once again, attaining the highest values at 2500

epochs (27.04 for CBF and 28.00 for ATT). This indicates that ResED3D generates the most accurate and high-fidelity reconstructions, outperforming DINO-TransRegNet, which follows closely, and significantly surpassing CNN-Trans3DNet and VGG-Att3DUNet, both of which struggle with weak reconstruction capabilities.

EPOCHS	MODEL	CBF - MSE	ATT - MSE	CBF- PSNR	ATT- PSNR	CBF- SSIM	ATT- SSIM	CBF- CCC	ATT- CCC
50	VGG-Att3DUNet	0.1031	0.1301	19.9532	19.5717	0.6613	0.6848	0.8283	0.8408
	U-NET[13]	0.0131	0.0168	18.9894	17.9907	0.6173	0.5892	0.9239	0.9072
	Proposed Model(ResED3D)	0.0124	0.0196	19.1875	17.4649	0.5084	0.4787	0.9236	0.8965
	DINO-TransRegNet	0.0247	0.0104	20.9713	15.2121	0.6230	0.5781	0.8314	0.8214
	CNN-Trans3DNet	0.0146	0.0066	19.2463	22.5136	0.7817	0.8827	0.9246	0.9671
100	VGG-Att3DUNet	0.0084	0.0110	20.9206	20.9063	0.8187	0.8455	0.8187	0.8455
	U-NET[13]	0.0096	0.0116	20.3025	20.2634	0.8165	0.8297	0.9495	0.9394
	Proposed Model(ResED3D)	0.0096	0.0102	20.1876	20.0680	0.7187	0.7539	0.9499	0.9510
	DINO-TransRegNet	0.0073	0.0101	21.9742	17.5271	0.7237	0.7357	0.8410	0.8617
	CNN-Trans3DNet	0.0071	0.0049	21.5660	25.7743	0.8750	0.9051	0.9633	0.9766
500	VGG-Att3DUNet	0.0043	0.0052	24.4608	23.9634	0.8959	0.9183	0.9051	0.8815
	U-NET[13]	0.0083	0.0102	20.8719	20.4501	0.8544	0.8425	0.9576	0.9496
	Proposed Model(ResED3D)	0.0063	0.0034	22.1167	24.8595	0.9016	0.9130	0.9697	0.9833
	DINO-TransRegNet	0.0072	0.0042	22.9327	18.7521	0.8322	0.8435	0.8753	0.8720
	CNN-Trans3DNet	0.0119	0.0057	19.6965	25.2553	0.8265	0.9211	0.9407	0.9719
1000	VGG-Att3DUNet	0.0047	0.0034	23.9623	25.3417	0.9311	0.9383	0.9765	0.9838
	U-NET[13]	0.0063	0.0053	22.1342	22.9643	0.8997	0.9160	0.9686	0.9700
	Proposed Model(ResED3D)	0.0021	0.0027	27.0687	26.3221	0.9502	0.9447	0.9888	0.9867
	DINO-TransRegNet	0.0064	0.0031	24.3437	20.5471	0.8726	0.8735	0.9072	0.9026
	CNN-Trans3DNet	0.0099	0.0039	20.3713	27.7029	0.8424	0.9656	0.9484	0.9805
1500	VGG-Att3DUNet	0.0057	0.0046	22.8269	23.8127	0.9166	0.9176	0.9718	0.9781
	U-NET[13]	0.0069	0.0054	21.8134	23.5759	0.9034	0.9244	0.9660	0.9728
	Proposed Model(ResED3D)	0.0024	0.0018	26.7547	28.1350	0.9677	0.9698	0.9874	0.9914
	DINO-TransRegNet	0.0054	0.0021	24.3522	22.5728	0.9021	0.9045	0.9117	0.9258
	CNN-Trans3DNet	0.0180	0.0145	19.2929	24.4930	0.7925	0.8817	0.9098	0.9247
2000	VGG-Att3DUNet	0.0065	0.0050	22.2892	23.9896	0.9061	0.9462	0.9679	0.9758
	U-NET[13]	0.0055	0.0053	22.9285	23.8973	0.9109	0.9335	0.9720	0.9728
	Proposed Model(ResED3D)	0.0024	0.0019	26.7085	27.6727	0.9629	0.9663	0.9873	0.9906
	DINO-TransRegNet	0.0043	0.0017	24.5374	23.7341	0.9115	0.9115	0.9562	0.9427
	CNN-Trans3DNet	0.0145	0.0080	21.1986	23.8313	0.8030	0.9351	0.9220	0.9584
2500	VGG-Att3DUNet	0.0071	0.0039	22.3652	25.0701	0.9122	0.9521	0.9654	0.9811
	U-NET[13]	0.0061	0.0051	22.7109	23.8914	0.9099	0.9336	0.9699	0.9745
	Proposed Model(ResED3D)	0.0022	0.0018	27.0499	28.0011	0.9675	0.9691	0.9882	0.9912
	DINO-TransRegNet	0.0040	0.0019	25.4213	25.2541	0.9431	0.9152	0.9681	0.9538
	CNN-Trans3DNet	0.0106	0.0094	20.8410	27.1560	0.8977	0.9196	0.9508	0.9506

Table 1: Quantitative Evaluation of Model Performance

SSIM further highlights ResED3D dominance. It achieves an SSIM score above 0.96, indicating its ability to preserve fine structural details more effectively than other models. While DINO-TransRegNet also performs well, ResED3D remains the most consistent in maintaining spatial integrity. In contrast, CNN-Trans3DNet and VGG-Att3DUNet exhibit lower SSIM scores, reflecting their inability to retain critical structural information in the generated maps.

Finally, ResED3D achieves the highest CCC, with values close to 0.99 for both CBF and ATT. This metric underscores ResED3D’s reliability, demonstrating its strong agreement with ground truth values. While DINO-TransRegNet and U-NET also perform well in terms of CCC, ResED3D remains the most precise and dependable. Overall, ResED3D stands out as the best-performing model, consistently delivering superior accuracy, structural integrity, and reliability in perfusion estimation.



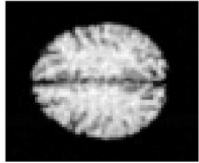
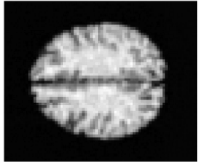


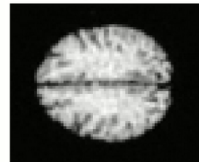

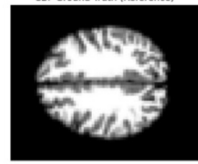
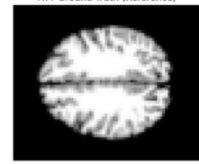
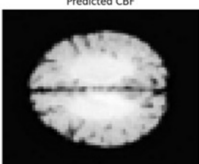
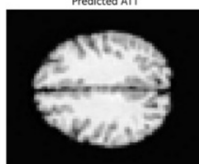
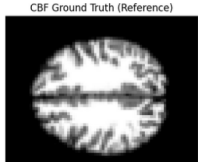
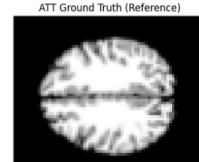
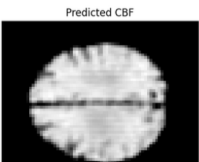
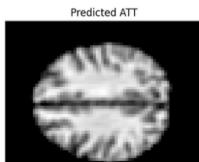
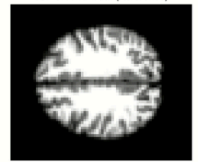
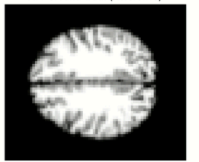
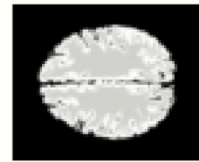
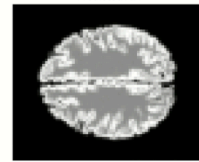
MODEL	HIGHES T ACCUR ACY EPOCH	CBF AND ATT GROUNDTRUTH MAPS		CBF AND ATT PREDICTED MAPS	
Proposed Model (ResED3 D)	2500	<small>CBF Ground Truth (Reference)</small> 	<small>ATT Ground Truth (Reference)</small> 	<small>Predicted CBF</small> 	<small>Predicted ATT</small> 
DINO- TransReg Net	2000	<small>CBF Ground Truth (Reference)</small> 	<small>ATT Ground Truth (Reference)</small> 	<small>Predicted CBF</small> 	<small>Predicted ATT</small> 
VGG-Att 3DUNet	1500	<small>CBF Ground Truth (Reference)</small> 	<small>ATT Ground Truth (Reference)</small> 	<small>Predicted CBF</small> 	<small>Predicted ATT</small> 
U-NET [13]	2000	<small>CBF Ground Truth (Reference)</small> 	<small>ATT Ground Truth (Reference)</small> 	<small>Predicted CBF</small> 	<small>Predicted ATT</small> 
CNN- Trans3D Net Model	1500	<small>CBF Ground Truth (Reference)</small> 	<small>ATT Ground Truth (Reference)</small> 	<small>Predicted CBF</small> 	<small>Predicted ATT</small> 

Table 2 : Qualitative Comparison of Ground Truth and Predicted CBF & ATT Maps

Table 2 presents a comparative analysis of four deep learning models—ResED3D, DINO-TransRegNet, VGG-Att3DUNet, CNN-Trans3DNet and U-NET—by showcasing their highest accuracy epochs, ground truth maps of CBF and ATT, and the corresponding predicted maps. The ResED3D model demonstrates the most precise predictions that closely align with the ground truth maps at epoch 2500. Overall, the visual analysis highlights ResED3D’s superior predictive capability, with DINO, VGG-Att3DUNet, CNN-Trans3D and U-NET also performing reasonably well, albeit with slightly less accuracy.

To further validate these findings, a detailed statistical analysis was conducted to compare the performance of the models in the next section.

4.3 Statistical Analysis

The Bland-Altman plot shown compares the performance of different models against the proposed ResED3D across all key metrics. Each plot visualises the mean of ResED3D and the other model's values on the x-axis and the difference (ResED3D - other model) on the y-axis. The mean difference (bias) is represented by a black dashed line, while the upper and lower limits of agreement (LoA) are shown as red and blue dashed lines, respectively. This analysis helps evaluate agreement, consistency, and systematic bias between the models and the ResED3D reference values.

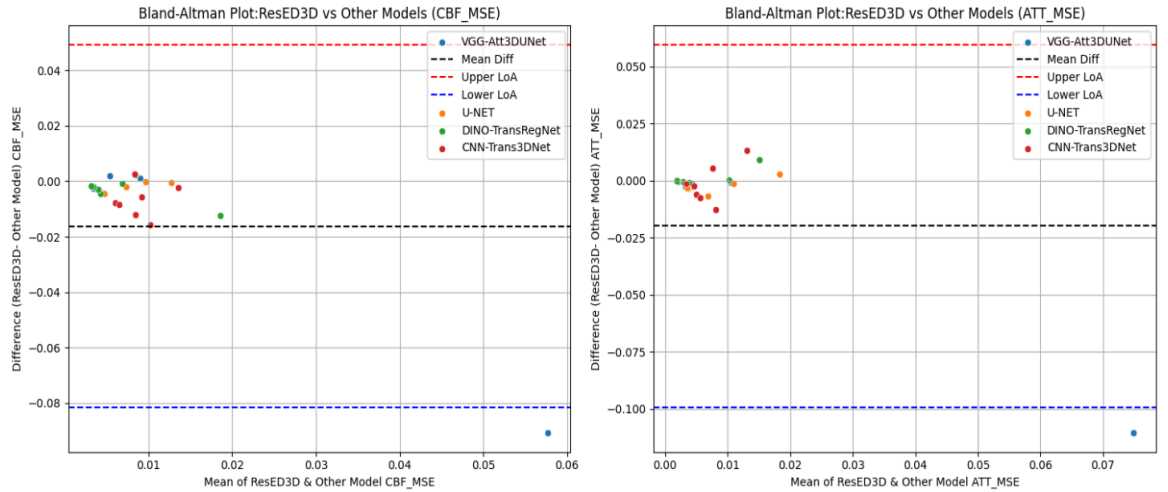


Fig.8: Bland-Altman Plot of ResED3D Vs Other Models for CBF_MSE and ATT_MSE

As shown in Fig.8 the ResED3D model shows the least bias in CBF_MSE, with most differences falling within the limits of agreement. This suggests that ResED3D provides

more stable and precise CBF estimations compared to other models. The spread of differences in other models indicates slightly higher errors in CBF predictions. The ATT_MSE plot reveals that the ResED3D model maintains lower error margins, with minimal deviations across subjects. Some models show larger variations, indicating inconsistent ATT estimations. The close agreement of ResED3D with ground truth suggests its efficiency in modeling ATT.

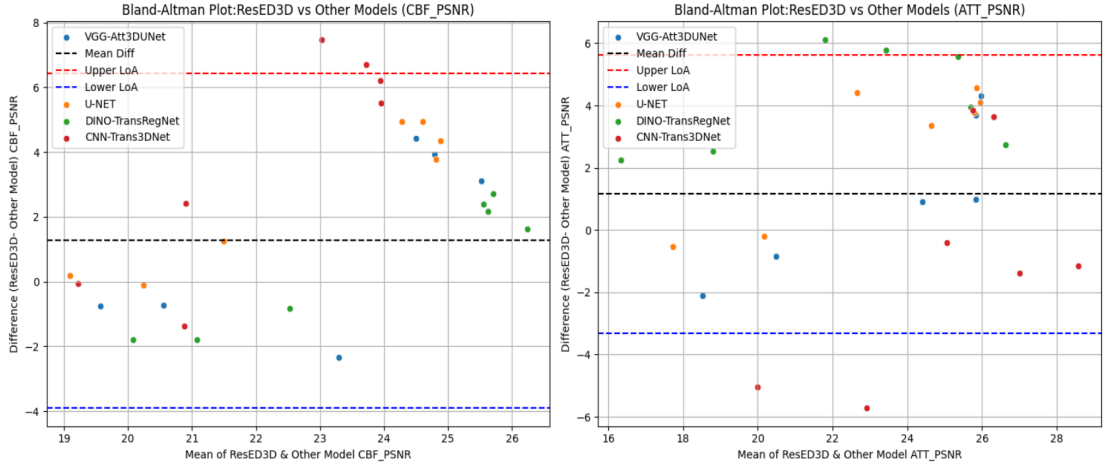


Fig.9 : Bland-Altman Plot of ResED3D Vs Other Models for CBF_PSNR and ATT_PSNR

Higher PSNR values indicate better prediction quality. Fig 9 shows that ResED3D achieves consistently high PSNR values, with minimal deviation. Other models exhibit wider variation, indicating potential noise and distortions in CBF estimations. The ATT_PSNR plot suggests that the ResED3D model provides cleaner reconstructions with higher PSNR values compared to other models. Some models show greater bias and wider limits of agreement, indicating lower-quality ATT reconstructions.

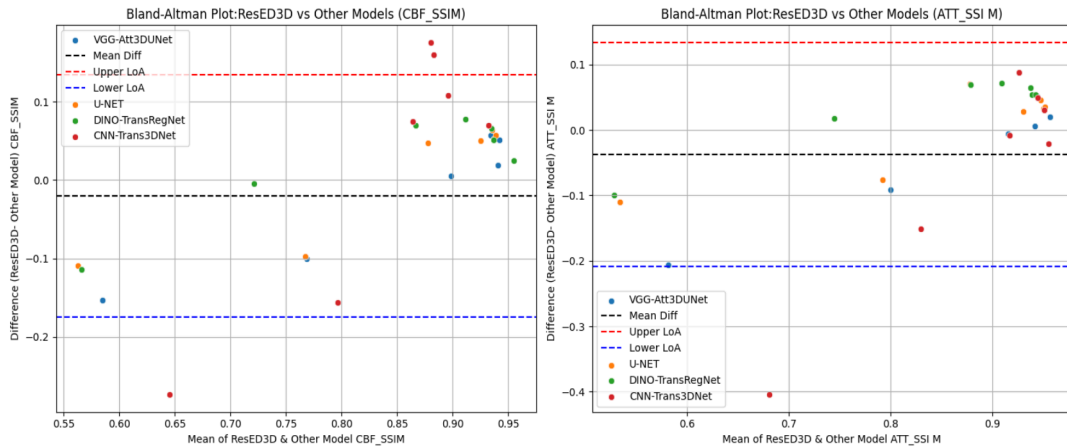


Fig.10 : Bland-Altman Plot of ResED3D Vs Other Models for CBF_SSIM and ATT_SSIM

Another plot for ResED3D Vs Other Models for CBF_SSIM and ATT_SSIM is shown in Fig.10. The CBF_SSIM plot suggests that the ResED3D model maintains structural integrity in CBF predictions, with differences mostly centered around zero. Other models show more scattered deviations, implying less accurate structural preservation. The ATT_SSIM plot further reinforces ResED3D's superiority in preserving spatial structures in ATT maps. The tight clustering of values around zero indicates strong agreement, whereas other models display higher deviations, suggesting structural inconsistencies.

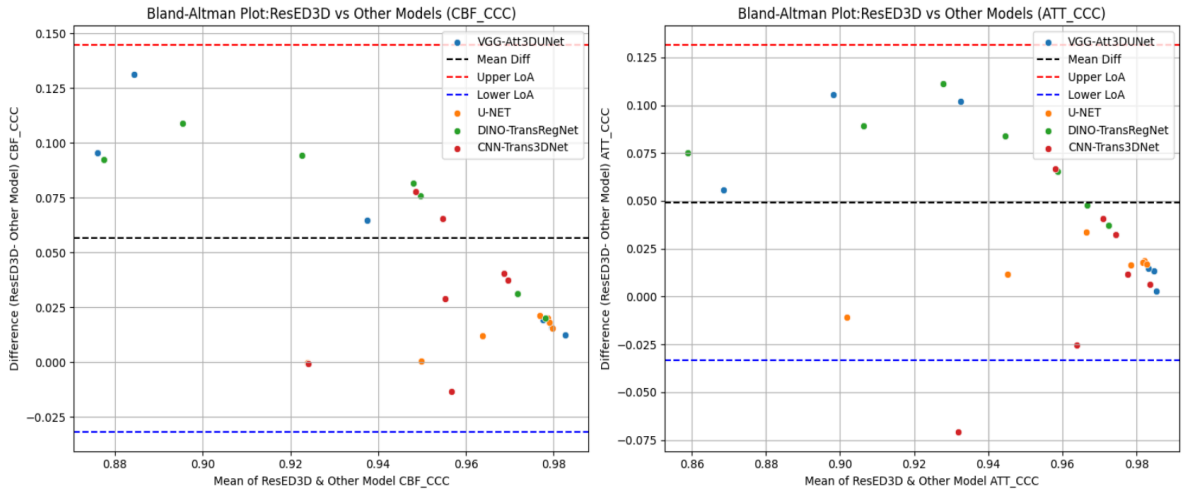


Fig.11 : Bland-Altman Plot of ResED3D Vs Other Models for CBF_CCC and ATT_CCC

To compare the CBF_CCC and ATT_CCC for other models and ResED3D a similar Bland-Altman Plot is shown in Fig. 11. The ResED3D model achieves high CBF_CCC values, indicating strong agreement with ground truth. Other models show slightly lower CCC values, which suggests they introduce some inconsistencies in the CBF predictions. The ATT_CCC plot shows that the ResED3D model maintains high concordance with the ground truth ATT values. Other models exhibit greater variation, indicating lower reliability in ATT predictions.

After conducting a comprehensive evaluation of model performance, the final chapter synthesizes key findings and reflects on the overall contributions of this study. It highlights the impact of deep learning in medical imaging. Additionally, potential directions for future research and improvements are explored, focusing on refining model accuracy, and expanding dataset diversity.

Chapter 5

Conclusion

This study explored multiple deep learning-based techniques for reducing PLD in multi-PLD ASL MRI, addressing the challenges of extended scan times and signal loss due to prolonged acquisition windows. A comprehensive simulation study was conducted to generate MRI data, allowing controlled analysis of different PLD configurations and their impact on perfusion parameter estimation. Several advanced models, including the ResED3D, DINO-TransRegNet, VGG-Att3DUNet, and CNN-Trans3DNet, were developed and evaluated for their effectiveness in estimating CBF and ATT.

Among the models tested, the proposed model achieved MSE values of 0.0022 for CBF and 0.0018 for ATT, PSNR values of 27.0499 for CBF and 28.0011 for ATT, SSIM values of 0.9675 for CBF and 0.9691 for ATT, and CCC values of 0.9882 for CBF and 0.9912 for ATT at 2500 epochs, proving to be a better model with superior accuracy and robustness. This demonstrates the high accuracy in estimating both CBF and ATT, outperforming other architectures by effectively capturing the smooth temporal variations in perfusion dynamics. Its ability to model continuous transformations within the feature space resulted in improved generalization and reduced errors compared to discrete CNN-based approaches. The incorporation of deep learning techniques, particularly those leveraging attention mechanisms and hybrid architectures, showed promising potential for optimizing ASL imaging by minimizing the need for multiple PLDs while preserving estimation accuracy. These findings highlight the feasibility of deep learning in accelerating ASL acquisition protocols, paving the way for more efficient and clinically viable perfusion imaging techniques.

Additionally, this study provided insights into how different model architectures influence perfusion parameter estimation, emphasizing the importance of designing networks capable of handling both spatial and temporal dependencies in ASL data. The comparative evaluation of various models highlighted the trade-offs between computational

complexity, interpretability, and performance, offering guidance for future research in the field. These findings reinforce the feasibility of deep learning in accelerating ASL acquisition protocols, paving the way for more efficient and clinically viable perfusion imaging techniques. Future work could focus on further refining these models through larger, more diverse datasets, integrating domain-specific constraints, and exploring hybrid approaches that combine deep learning with traditional physiological modeling for enhanced robustness and reliability. Future work could also focus on determining the optimal PLD at which accurate perfusion measurements can be obtained, ensuring minimal scan time while maintaining precision.

Chapter 6

References

- [1] Østergaard, L., et al. "High-resolution measurement of cerebral blood flow using intravascular tracer bolus passages." *Magnetic Resonance in Medicine* 36.2 (1996): 715-725.
- [2] Gulani, V., et al. "Gadolinium deposition in the brain: Summary of evidence and recommendations." *Radiology* 285.2 (2017): 536-554.
- [3] Detre, J. A., et al. "Perfusion imaging using arterial spin labeling." *Proceedings of the National Academy of Sciences* 89.7 (1992): 2842-2845.
- [4] Alsop, D. C., et al. "Recommended implementation of ASL perfusion MRI for clinical applications." *Magnetic Resonance in Medicine* 73.3 (2015): 102-116.
- [5] Zhou, J., et al. "Quantitative perfusion imaging using FAIR in stroke patients." *Magnetic Resonance in Medicine* 45.3 (2001): 541-549.
- [6] Borogovac, A., & Asllani, I. "Arterial spin labeling: Techniques and applications." *Journal of Neuroscience Methods* 205.1 (2012): 50-64.
- [7] Buxton, R. B., et al. "A kinetic model for quantitative ASL." *Magnetic Resonance in Medicine* 40.3 (1998): 383-396.
- [8] MacIntosh, B. J., et al. "Arterial transit time mapping in Alzheimer's disease." *Journal of Magnetic Resonance Imaging* 31.6 (2010): 1471-1478.
- [9] Chappell, M. A., et al. "Partial volume correction in ASL." *NeuroImage* 49.2 (2010): 1114-1121.
- [10] Hirschler, L., et al. "Multi-delay ASL for cerebrovascular assessment." *NeuroImage* 220 (2020): 116738.

- [11] Fan, A. P., et al. "Accelerated multi-PLD ASL using deep learning." *Scientific Reports* 10 (2020): 69393.
- [12] Liu, P., et al. "Deep learning for ASL quantification." *Magnetic Resonance in Medicine* 87.1 (2022): 291-308.
- [13] Luciw NJ, Shirzadi Z, Black SE, Goubran M, MacIntosh BJ. Automated generation of cerebral blood flow and arterial transit time maps from multiple delay arterial spin-labeled MRI. *Magn Reson Med*. 2022 Feb;88(1):406-417. doi: 10.1002/mrm.29193.
- [14] Ishida, S., Fujiwara, Y., Takei, N., Kimura, H., & Tsujikawa, T. (2024). Comparison between supervised and physics-informed unsupervised deep neural networks for estimating cerebral perfusion using multi-delay arterial spin labeling MRI. *NMR in Biomedicine*. <https://doi.org/10.1002/nbm.5177>
- [15] KimD, LipfordME, HeH,etal. Parametric cerebral blood flow and arterial transit time mapping using a 3D convolutional neural network. *Magn Reson Med*. 2023;90:583-595. doi: 10.1002/mrm.29674
- [16] Li Y, Wang Z. Deeply Accelerated Arterial Spin Labeling Perfusion MRI for Measuring Cerebral Blood Flow and Arterial Transit Time. *IEEE J Biomed Health Inform*. 2023 Dec;27(12):5937-5945.
- [17]Shou, Q., Zhao, C., Shao, X., Jann, K., Kim, H., Helmer, K. G., Lu, H., & Wang, D. J. (2023). Transformer-based deep learning denoising of single and multi-delay 3D arterial spin labeling. *Magnetic Resonance in Medicine*, 91(3), 803–818. <https://doi.org/10.1002/mrm.2988>
- [18] Maier, O., Kober, T., & Weiskopf, N. (2020). *Non-linear fitting with joint spatial regularization in Arterial Spin Labeling*. arXiv preprint arXiv:2009.05409. <https://doi.org/10.48550/arXiv.2009.05409>
- [19] Bladt P, van Osch MJP, Clement P, Achten E, Sijbers J, den Dekker AJ. Supporting measurements or more averages? How to quantify cerebral blood flow most reliably in 5 minutes by arterial spin labeling. *Magn Reson Med*.

2020;84:2523–2536. <https://doi.org/10.1002/mrm.28314>

[20] Alsop, D.C., et al., 2015. Recommended implementation of arterial spin-labeled perfusion MRI for clinical applications: A consensus of the ISMRM perfusion study group and European consortium for ASL in dementia. *Magn.Reson.Med.* 73 (1), 102–116. <http://dx.doi.org/10.1002/mrm.25197>.

Chapter 7

Appendix: Accomplished Programme Outcomes

1. PO1 - Engineering Knowledge

1. Mathematics & Science Integration: The project integrates mathematical and scientific principles, including signal processing, MRI physics, and deep learning, to estimate cerebral blood flow and arterial transit time accurately.
2. Algorithmic Development: Advanced deep learning models are employed to analyze multi-PLD pCASL data, extracting relevant features to enhance parameter estimation accuracy.
3. Specialization Application: The research applies medical image processing and deep learning techniques to enhance ASL-based perfusion imaging, contributing to non-invasive cerebral hemodynamics analysis.

2. PO2 - Problem Analysis

1. Problem Identification: The study tackles the limitations of multi-PLD ASL, including extended scan durations and undersampling issues, by developing a deep learning model to generate accurate CBF and ATT maps.
2. Literature Review & Analysis: A thorough review of conventional and deep learning-based ASL image processing methods is conducted to establish a basis for model improvement.
3. Handling Complexity in Neuroimaging: The model adapts deep learning techniques to account for timing delays and varying post-labeling delays (PLDs) in pCASL, ensuring robust parameter estimation.
4. Data-Driven Analysis: The methodology employs advanced statistical evaluation, including SSIM and PSNR analysis, to validate the performance of deep learning models against conventional fitting techniques.

3. PO3 - Design and Development of Solutions

1. Innovative Model Design: The research introduces various deep learning models for efficient feature extraction and parameter estimation.
2. Optimization Strategies: Hyperparameter tuning and adaptive loss functions are employed to refine model accuracy.
3. Clinical Relevance: The approach aimed to facilitate faster and more reliable assessment of cerebral perfusion, benefiting medical applications such as stroke diagnosis and dementia studies.
4. Societal Consideration: Improved scan time efficiency reduces patient discomfort and enhances accessibility to ASL-based neuroimaging, promoting early disease detection.
5. Environmental Consideration: Efficient model selection and preprocessing reduce computational load, supporting energy-efficient processing.

PO4. Investigation:

1. Research-Based Analysis: The study incorporates domain knowledge in medical physics, ASL-MRI, and deep learning to improve parameter mapping.
2. Performance Metrics: RMSE, mean absolute error (MAE), and structural similarity index (SSIM) are used to compare estimated CBF/ATT maps against ground truth values.
3. Reliable Conclusions: The study confirms the feasibility of deep learning models in reducing scan time while maintaining accuracy in perfusion parameter estimation.

PO5. Modern Tool Usage:

1. Deep Learning Frameworks: The project utilizes TensorFlow, PyTorch, and Keras for model development and training.
2. Medical Imaging Datasets: Performed MRI simulation using Matlab to generate synthetic ASL data for controlled experimentation.
3. Computational Resources: Google Colab is utilized for model training and evaluation, enabling efficient deep learning experiments with GPU/TPU acceleration.

PO6. Societal Impact:

1. Impact on Medical Imaging: The study advances AI-driven neuroimaging, aiding in faster, more accessible perfusion analysis.
2. Healthcare Contributions: Reducing scan time in ASL imaging minimizes patient exposure and enhances diagnostic throughput in clinical settings.

3. Efficient Resource Use: The optimized model minimizes the need for multiple ASL acquisitions, conserving scanner usage and improving workflow.

PO7. Sustainability:

1. Optimized Model Design: The research employs lightweight models to minimize computational energy consumption during training and inference.
2. Sustainable Imaging Practices: The reduction of scan time leads to lower MRI operational costs and resource usage.

PO8. Ethics:

1. Transparent Model Development: The methodology ensures unbiased training and validation to avoid overfitting or misleading clinical interpretations.
2. Social Responsibility: The study prioritizes ethical AI applications in medical imaging, promoting trust and transparency in deep learning solutions.
3. Professional Ethics: All code samples and data are used in adherence to ethical research practices and industry standards.

PO9. Teamwork:

1. Interdisciplinary Collaboration: The research encourages experts in medical imaging, AI researchers, and clinicians to collaborate to enhance model accuracy and ensure clinical applicability.
2. Expertise Integration: Team members bring specialized knowledge in deep learning, MRI physics, and neuroimaging data interpretation to develop a robust and clinically relevant model.

PO10. Communication:

1. Technical Documentation: The methodology, preprocessing techniques, and model architecture are clearly documented for reproducibility.
2. Visualization Techniques: Graphical representations of estimated CBF and ATT maps enhance interpretability.
3. Scientific Dissemination: Findings are shared through research papers, conference presentations, and academic discussions.

PO11. Project Management:

1. Workflow Planning: The project follows a structured workflow, including data preprocessing, model training, evaluation, and refinement.
2. Risk Mitigation: The study anticipates challenges such as data diversity and model generalization, adjusting training strategies accordingly.

3. Budget-Conscious Execution: The project is conducted with efficient resource allocation, optimizing cost while achieving high-performance results.

PO12. Life-Long Learning:

1. Continuous Knowledge Development: The research involves ongoing learning in deep learning, medical physics, and neuroimaging techniques.
2. Self-Directed Learning: Extensive literature review and experimentation enhance the understanding of deep learning applications in ASL imaging.
3. Technical Growth: The hands-on implementation of deep learning models enhances skills in neural network design, medical data processing, and AI-based diagnostic tools.