



2D U-Net Deep Learning Implementation for Contrast-Enhanced Tumour Identification and Synthesis of Contrast-Enhanced MRI in Glioma Patients

Author: Amritansh Jain (Postgraduate Student at UCL)

Program: MSc in Artificial Intelligence for Biomedicine and Healthcare (2022-23)

Supervisor: Matthew Grech Sollars (Associate Professor in Quantitative Neuroradiology at UCL)

Submission Date: 25th September 2023

Disclaimer: This report is submitted as part of the MSc in Artificial Intelligence for Biomedicine and Healthcare requirement at UCL. It is substantially the result of my work except where explicitly indicated in the text. Either: The report may be freely copied and distributed provided the source is expressly acknowledged.
Or:
The report will be distributed to the internal and external examiners but, after that, may not be copied or distributed except with permission from the author.

Abstract

Magnetic Resonance Imaging (MRI) is an essential standard biomedical imaging technique that is pivotal in diagnosing brain tumours and forecasting their progression and dispersion throughout the human body. It uses magnetic fields and radio waves to produce detailed images to help oncologists or radiologists identify the tumour region in the brain. Both contrast-enhanced (CE) and non-contrast-based MRI scans are used for diagnosis. Still, contrast-based MRIs improve and enhance the visibility of areas with blood-barrier disruptions due to invasive cancerous cells in the brain. It helps in getting an accurate and precise positioning of the tumour cells. Patients are administered contrast agents to get the CE-MRI scans during the scanning process. Gadolinium-based contrast agents (GBCAs) are frequently used in neuro-oncology during MRI examinations to identify, characterise, and monitor various diseases, especially brain tumours, making them a prevalent and obvious choice in this domain. However, gadolinium poses MRI safety issues and discomfort to patients, which is still an area of ongoing research about its harmful impact as a by-product of the contrast agent. There is a significant risk of gadolinium accumulation in the body or its deposition in the brain, potentially leading to adverse side effects and reactions, particularly in pregnant patients, where contrast agents are contraindicated as they can harm the developing foetus. This will lead to diagnostic healthcare inequality between male and female patients. Also, extended acquisition times for CE scans hinder MRI access and may result in motion artifacts, increasing the cost and scanning time.

Several deep learning architectures (neural networks) have been identified and researched, showcasing potential solutions and obviating the need for injecting contrast agents to get enhanced MRI scans. In this study, we have trained a 2D U-Net Convolutional Neural Network (CNN) on the University of California San Francisco Preoperative Diffuse Glioma MRI (UCSF-PDGM) dataset to evaluate the multiple imaging modalities of the MR images based on how well our model performs the tumour identification task with each of the individual modalities as the input and then locating the CE-tumour region. The deep learning model was trained by mapping each MRI sequence to the CE-tumour extracted from the tumour-segmented images in the dataset. The modalities considered for the study from the UCSF-PDGM dataset are T1-weighted images, T2-weighted images, T2/FLAIR-weighted images (Fluid-Attenuated Inversion Recovery sequence), Fractional Anisotropy (FA) map, Mean Diffusivity (MD) map, ASL (arterial spin labeling perfusion) images, and SWI (susceptibility weighted imaging) images. Additionally, based on the outcomes from the previous task, we consolidated the most effective modalities (ASL, T1 and ASL, T1, T2) together to analyse the accuracy of tumour identification after introducing more diversity in the training dataset. We also trained our 2D U-Net CNN model and evaluated the individual modalities to predict the CE T1-weighted MR images. Then, we clubbed the top two MRI modalities to see if there was an increase in the prediction accuracy. The binary cross entropy loss function and dice score evaluation metric has been used to analyse the network's performance in each task during the training phase of the deep learning model. We have also computed IoU (Intersection over Union), PSNR (Peak-Signal-to-Noise-Ration) and F2 score (weighted harmonic mean of the precision and recall) on the test set to help us compare the results between different MRI modalities. In the CE-tumour identification task, the ASL MRI sequence yielded the best results, followed by T1-weighted and T2-weighted MRI sequences. The combined results from the ASL, T1w and T2w modalities were approximately equivalent or in the same range as those from the best three MRI sequences. In predicting post-contrast T1-weighted MRI scans, the pre-contrast T1-weighted MRI sequence produced the best results compared to other modalities followed by the T2-weighted MRI sequence. The combined effects of MRI modalities T1 and T2 were also approximately equivalent to the best results.

Therefore, through this research, we aim to validate the reliability of the UCSF-PDGM dataset, one of the first publicly available datasets with multiple pre-contrast MRI sequences. Also, we pinpointed the most proficient individual modality (ASL) and the combined imaging modalities (T1, T2 and ASL) in this study for the CE-tumour identification task and in the synthesis of CE-T1 MRI scans (T1 and T2). We are confident that our findings and results will surely be a transformative advancement in contrast-enhanced tumour identification, and it will also build a solid foundation for the prediction of CE T1-weighted MR images, capitalising on the intrinsic strengths of a basic 2D U-Net model using the UCSF-PDGM dataset. Therefore, this study supports using artificial intelligence techniques to generate virtual contrast-enhanced MRI scans as a viable alternative to administering gadolinium-based contrast agents (GBACs) in glioma patients to avoid their deposition in the brain.

Table of Contents

| | | |
|----------|---|-----------|
| 1 | Chapter 1: Introduction | 6 |
| 1.1 | Magnetic Resonance Imaging (MRI) | 6 |
| 1.2 | Gadolinium-Based Contrast Enhancing Agents (GBCAs)..... | 6 |
| 1.3 | Gadolinium-Deposition in the Brain | 7 |
| 1.4 | Integration of Deep Learning to Mitigate GBCAs | 9 |
| 2 | Chapter 2: Background – Related Work | 10 |
| 2.1 | Study 1: Synthesizing MR Image Contrast Enhancement Using 3D High-Resolution ConvNets by C. Chen et al., 2023 | 10 |
| 2.2 | Study 1: Brain tumour MRI images identification and classification based on the recurrent convolutional neural network by Vankdothu & Hameed, 2022 | 10 |
| 2.3 | Study 3: Magnetic resonance imaging contrast enhancement synthesis using cascade networks with local supervision by Xie et al., 2022..... | 10 |
| 2.4 | Study 4: Deep-learning-based synthesis of post-contrast T1-weighted MRI for tumour response assessment in neuro-oncology: a multicentre, retrospective cohort study by Jayachandran Preetha et al., 2021..... | 11 |
| 2.5 | Study 5: Deep Learning Enables Reduced Gadolinium Dose for Contrast-Enhanced Brain MRI by Gong et al., 2018 | 11 |
| 2.6 | Study 6: Toward deep learning replacement of gadolinium in neuro-oncology: A review of contrast-enhanced synthetic MRI by Moya-Sáez et al., 2023 | 11 |
| 3 | Chapter 3: Materials and Methods | 12 |
| 3.1 | Dataset Description | 12 |
| 3.2 | Imaging Data Extraction | 13 |
| 3.3 | Data Pre-processing | 13 |
| 3.4 | Deep Learning Architecture | 14 |
| 3.4.1 | Model Design..... | 15 |
| 3.4.2 | Training the Model | 16 |
| 3.4.3 | Testing the Model | 17 |
| 4 | Chapter 4: Results | 18 |
| 4.1 | Binary Cross Entropy Loss Function | 18 |
| 4.2 | Model Performance Evaluation Metrics | 19 |
| 5 | Chapter 5: Discussions | 27 |
| 5.1 | Limitations and Suggestions..... | 27 |
| 5.2 | Future Works | 28 |
| 6 | Chapter 6: Conclusions..... | 28 |
| 7 | Code List..... | 28 |
| 8 | References | 28 |

Table of Figures

| | |
|---|----|
| Figure 1.1 MRI images using the SWI modality of two distinct brains to highlight the difference between a non-pathological, glioma-free brain and a brain affected by glioma. The highlighted area in image two is the tumour region. The images are sourced from the UCSF-PDGM dataset..... | 6 |
| Figure 1.2 Non-CE T1-weighted MR image and CE T1-weighted MR Image. In the CE image (image 2), the bright highlighted part is the tumour region enhanced due to the Gadolinium-based contrast-enhancing agent (GBCA)..... | 7 |
| Figure 1.3 Progressive increase in the signal intensity of the globus pallidus (triangular-shaped subcortical structure of the brain) from images a to c and dentate nucleus (cerebellar cluster of neurons) from images d to f can be seen in patients who have a history of being administered with GBCA does in MRI examinations during the diagnosis of lung cancer (Kanda et al., 2014). This meant that for the MRIs taken during the study done by Kanda et al., the output scans were already contrast-enhanced. However, the patients were not administered any contrast-enhancing agent, thus proving gadolinium deposition in the brain. | 8 |
| Figure 1.4 Classification of existing gadolinium-based contrast agents (Pullicino & Das, 2017). | 8 |
| Figure 1.5 T1w, T1w-CE and Tumour Segmentation Images of a patient in UCSF-PDGM dataset. Only the CE Tumour mask (image 4) has been considered in this study to train the 2D U-Net model for CE tumour identification.... | 9 |
| Figure 3.1 Multimodal MRI sequences of a patient with glioblastoma in the UCSF-PDGM dataset. T1 = T1-weighted pre-contrast, T2 = T2-weighted, T2/FLAIR = T2-weighted FLAIR, ASL = arterial spin labeling perfusion, SWI = susceptibility-weighted imaging, DWI = diffusion-weighted imaging, FA = fractional anisotropy map, MD = mean diffusivity map, ADC = apparent diffusion coefficient map (Calabrese et al., 2022). | 13 |
| Figure 3.2 An example of the U-net architecture for images with a resolution of 32x32 pixels. The blue boxes represent multi-channel feature maps, with the number of channels shown at the top of each box. The x-y size is provided at the lower left edge of the box. White boxes indicate copied feature maps. The arrows represent the various operations involved. From the paper (Ronneberger et al., 2015). | 14 |
| Figure 3.3 The 2D U-Net Convolutional Neural Network used in this study for CE-tumour identification and CE-T1 MRI synthesis..... | 15 |
| Figure 3.4 A Sample Training input image (T2 modality) (img. 1) and Training output image (CE tumour mask from tumour segmentation image) (img. 2) in the CE-tumour identification task. These images are the single slices (slice number = 1000) of dimension 240 x 240 in the training dataset that will be mapped to each other during the neural network training process..... | 16 |
| Figure 3.5 A Sample Test input image (img. 1) and Test output image (img. 2) in the CE-tumour identification task from T2-weighted MRI sequence. These images are the single slices (slice number = 650) of dimension 240 x 240 in the test dataset. The trained model will take img. 1 as input to predict img. 2 as the output. | 17 |
| Figure 4.1 Binary Entropy Loss comparison for different MRI modalities researched in this study. Img.1 is associated with the CE tumour identification task, and img.2 is related to the CE-T1 Prediction task. The maximum loss score is highlighted with a red circle, and the minimum loss score is highlighted with a green circle. | 18 |
| Figure 4.2 A Bar chart comparing individual and combined MRI modalities based on Dice Score, PSNR, IoU and F2 score for CE-tumour identification. The best scores are of T1c (highlighted with a green-yellow box), and the worst (highlighted with a red-yellow box) are of FA and SWI mixed with FA, leading to the lowest score in the three metrics. | 20 |
| Figure 4.3 A Bar chart comparing individual and combined MRI modalities based on Dice Score, PSNR, IoU and F2 score for CE-T1 Prediction. The best score is modality T1 (highlighted with a green-yellow box), and the worst modality MD (highlighted with a red-yellow box). | 20 |
| Figure 4.4 T1 MRI sequence used to identify CE Tumour mask. | 21 |
| Figure 4.5 T2 MRI sequence used to identify CE Tumour mask. | 22 |
| Figure 4.6 T2/FLAIR MRI sequence used to identify CE Tumour mask. | 22 |
| Figure 4.7 FA map used to identify CE Tumour mask. | 22 |
| Figure 4.8 MD map used to identify CE Tumour mask. | 22 |
| Figure 4.9 ASL MRI sequence used to identify CE Tumour mask.... | 22 |
| Figure 4.10 SWI MRI sequence used to identify CE Tumour mask.... | 23 |
| Figure 4.11 CE-T1 MRI sequence used to identify CE Tumour mask. | 23 |
| Figure 4.12 Identification of CE tumour mask by our 2D U-Net model trained on ASL + T1 MRI sequences. | 23 |
| Figure 4.13 Identification of CE tumour mask by our 2D U-Net model trained on ASL + T1 + T2 MRI sequences..... | 23 |
| Figure 4.14 Predicting CE T1 MRI from pre-contrast T1 MRI sequence. | 24 |
| Figure 4.15 Predicting CE T1 MRI from pre-contrast T2 MRI sequence. | 24 |
| Figure 4.16 Predicting CE T1 MRI from pre-contrast T2/FLAIR MRI sequence. | 24 |
| Figure 4.17 Predicting CE T1 MRI from pre-contrast FA map. | 24 |
| Figure 4.18 Predicting CE T1 MRI from pre-contrast MD map. In this case, we don't get any reasonable output.... | 24 |
| Figure 4.19 Predicting CE T1 MRI from pre-contrast ASL MRI sequence..... | 25 |
| Figure 4.20 Predicting CE T1 MRI from pre-contrast SWI MRI sequence. | 25 |
| Figure 4.21 Predicting CE T1 MRI by our 2D U-Net model trained on T1 + T2 MRI sequences..... | 25 |
| Figure 4.22 Plotting the data distribution of dice score for different MRI modalities while identifying CE tumours. | 26 |
| Figure 4.23 Plotting the data distribution of dice score for other MRI modalities while predicting CE-T1 MR images. . | 26 |

Table of Tables

| | |
|---|----|
| Table 1 Binary Entropy Loss values for the respective MRI modalities, individual and combined. The red highlight is for maximum loss values, and the green highlight is for the minimum loss value. | 19 |
| Table 2 The computed evaluation metric scores for CE tumour identification. We highlighted the best score with green and the worst with red without considering the score of T1c, as it is our benchmark for this task. | 21 |
| Table 3 The computed evaluation metric score for CE-T1 image prediction. We highlighted the best score with green and the worst with red. | 21 |

1 Chapter 1: Introduction

In the vast and intricate landscape of global health concerns, cancer is one of the leading causes of death worldwide. It is usually terminal, underscoring the importance of prompt and effective prognosis. According to the World Health Organisation (WHO), nearly 10 million souls succumbed to various cancers in 2020 (Ferlay et al., 2021). Current statistical data from Cancer Research UK elucidates that approximately 367,000 people get diagnosed with cancer each year in the UK, which averages to around 1,000 people daily. While brain or intracranial tumours constitute a small fraction of the total cancer incidences, their malignant characterisation arises from the inferior survival rate, especially in glioma patients (McKinney, 2004). Gliomas are predominantly high-grade brain tumours found in adults, primarily men (McKinney, 2004). It is a tumour originating from the brain's glial progenitor cells that support nerve cells. It is often diagnosed in older adults. However, paediatric populations are not spared. According to the American Cancer Society, brain and spinal cord tumours are children's second most common cancers, surpassed only by leukaemia, the chief paediatric malignancy (Subramanian et al., 2023). In 2022, it had been projected that 72,360 individuals aged 40 and over in the United States would be diagnosed with a primary brain tumour. Current estimates suggest that 700,000 Americans live with primary brain tumours, of which benign variants comprise 71% and malignant variants comprise 29% (Rahman & Islam, 2023). Therefore, pioneering advancement in diagnosing brain tumours is paramount to improving patient prognosis and survival outcomes. Figure 1.1 below displays the MRI scans (SWI modality) of a patient without a tumour and another patient with a tumour, respectively.

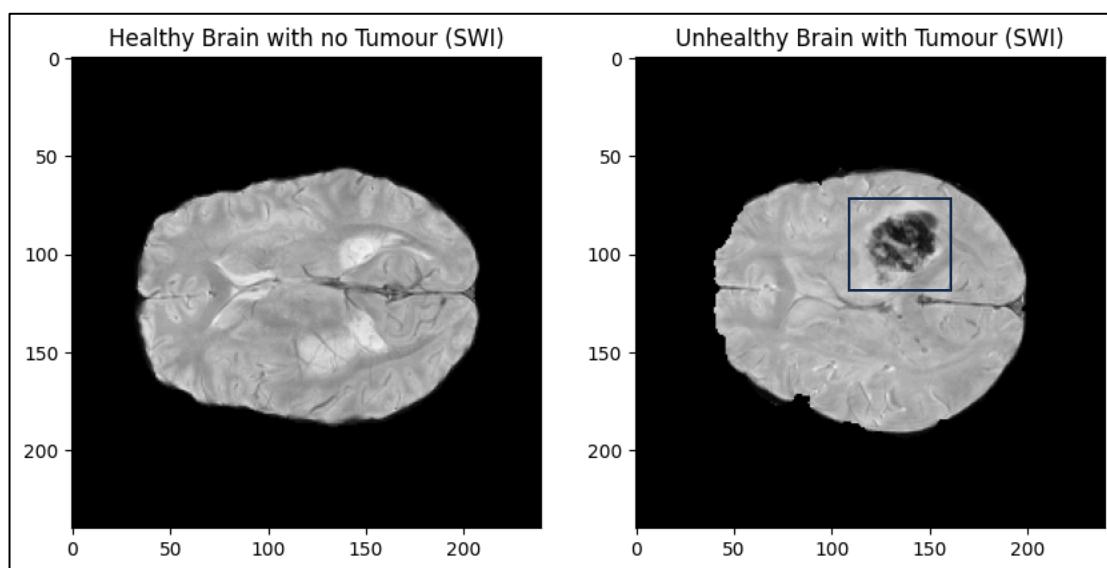


Figure 1.1 MRI images using the SWI modality of two distinct brains to highlight the difference between a non-pathological, glioma-free brain and a brain affected by glioma. The highlighted area in image two is the tumour region. The images are sourced from the UCSF-PDGM dataset.

1.1 Magnetic Resonance Imaging (MRI)

Numerous sophisticated computational methods and biomedical imaging techniques are available to obtain images for diagnosing cancerous diseases. These diagnostic tools include X-rays, Ultrasonic Imaging (UI), Positron Emission Tomography (PET), Computed Tomography (CT), Magnetic Resonance Spectroscopy (MRS), Single-Photon Emission Computed Tomography (SPECT), and Magnetic Resonance Imaging (MRI) (Aamir et al., 2022). In most cases, Magnetic Resonance Imaging (MRI) is the preferred medical imaging technique by oncologists and radiologists for detecting and visualising the precise location of brain tumours and producing a high-resolution, detailed brain image (Mohammed et al., 2019) (Ranjbarzadeh et al., 2023). MRI, a non-invasive diagnostic tool, renders intricate pictures of nearly all internal anatomical components. The MRI machine generates these images using a robust magnet with radiofrequency waves without emitting ionising radiation (<https://www.hopkinsmedicine.org>). Over 3.08 million MRI scans were performed in England between 2015 and 2016 (Pullicino & Das, 2017). MRI also facilitates the acquisition of images through various imaging modalities (multiparametric MRI sequences), yielding a more refined dataset on brain tumours. This expanded information helps improve precision in detecting the tumour region in the brain.

1.2 Gadolinium-Based Contrast Enhancing Agents (GBCAs)

Another distinguishing characteristic of MRI, which emphasises its significance in brain tumour detection, is its capability to employ contrast-enhancing agents. The contrast agents are injected into the patient's body before taking an MRI scan. The contrast agent enhances the clarity and resolution of the acquired MRI images, providing more

accurate interpretations (Figure 1.2). In 1988, the first FDA-approved (US Food and Drug Association) contrast agent explicitly designed for Magnetic Resonance Imaging (MRI) became available for clinical use. It was gadopentetate dimeglumine (Magnevist), a gadolinium-based contrast agent (Lohrke et al., 2016). The contrast agents contain paramagnetic ions (gadolinium) that can influence the proton relaxation time (T_1 relaxivity), thus increasing the visibility of organs and tissues of the human body (enhancement effect) (Lohrke et al., 2016). This breakthrough made gadolinium the most effective paramagnetic ion for contrast agents, and since then, it has been used as a base for contrast agents in MRI for more than 20 years. Gadolinium-based contrast agents (GBCAs) are currently the most commonly administered contrast-enhancing agent during MRI examinations. Annual global consumption of GBCAs exceeds 30 million doses; since its inception, over 300 million doses have been administered worldwide (Guo et al., 2018). In the United States, contrast-enhancing agents are used in approximately 70% of MRI procedures (Davenport, 2018). This shows the oncologists' confidence in prescribing contrast agents during MRIs for their diagnosis, as these contrast agents are not intended to have any pharmacological implications. Until 1994, three MRI contrast agents had been approved for clinical applications in the United States. From 1995 to 2017, the FDA endorsed six other MRI contrast agents (Ibrahim et al., 2023). Among the nine gadolinium-based contrast agents (GBCAs) sanctioned by the United States FDA for CE-MR imaging, seven are approved explicitly for CE-MR imaging of the central nervous system (CNS). These include gadoterate meglumine (Dotarem), gadobutrol (Gadavist), gadopentetate dimeglumine (Magnevist), gadobenate dimeglumine (MultiHance), gadodiamide (Omniscan), gadoversetamide (OptiMARK), and gadoteridol (ProHance). The two contrast agents, gadofosveset trisodium (Ablavar) and gadoxetic acid (Eovist/Primovist), are not approved for CE-MR imaging of the CNS due to some of the properties that make them unsuitable for MR imaging (Kanal et al., 2014).

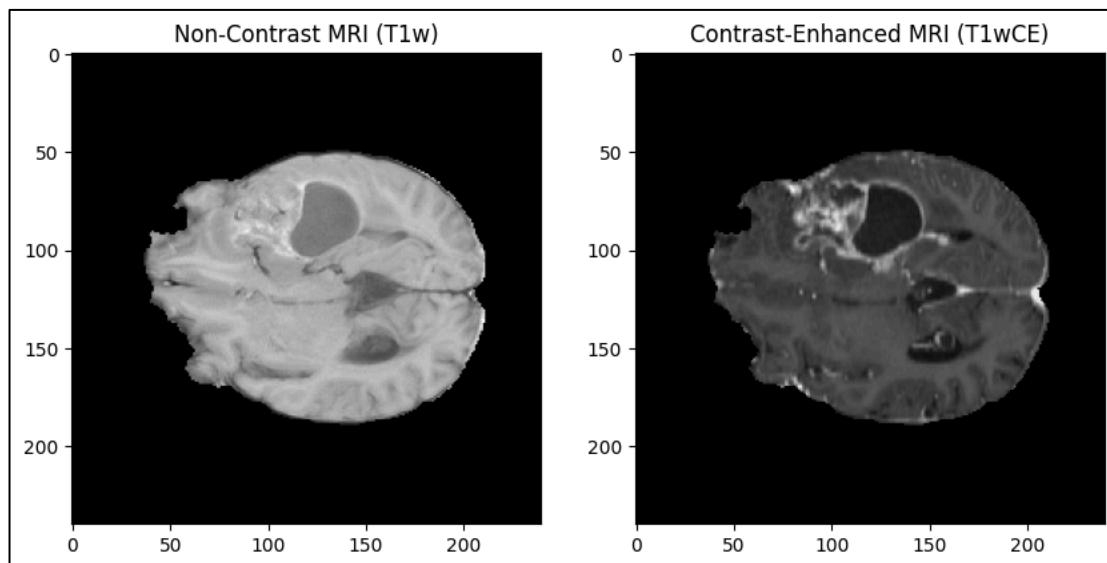


Figure 1.2 Non-CE T_1 -weighted MR image and CE T_1 -weighted MR Image. In the CE image (image 2), the bright highlighted part is the tumour region enhanced due to the Gadolinium-based contrast-enhancing agent (GBCA).

1.3 Gadolinium-Deposition in the Brain

Gadolinium (Gd) is a known heavy metal in the lanthanide series and can be toxic to humans. To make it safe for human use, it is combined with organic ligands or chelating agents to create gadolinium-based contrast agents (GBCAs) (Kanal et al., 2014). Despite these measures, there are still concerns and drawbacks related to using gadolinium. The most major adverse association of gadolinium is with Nephrogenic Systemic Fibrosis (NSF). Although rare, NSF is a life-threatening condition linked to administering Gadolinium-Based Contrast Agents (GBCAs) in patients with severe renal (kidney) impairment (Pullicino & Das, 2017). The inception of NSF can be traced back to 1997, when 15 renal dialysis patients were diagnosed with dermatological manifestations similar to scleromyxoedema, a known skin condition. However, upon further clinical and histopathological evaluation, it became apparent that the observed differences were significant enough to justify its classification as a new clinicopathological entity (Cowper et al., 2000). The association of GBCAs with NFS was only confirmed in 2006 when it was established that renal disease patients undergoing MRI using GBCAs have triggered the development of Nephrogenic Systemic Fibrosis (Grobner, 2006). After that, the European Medicine Authority recommended that high-risk GBCAs, which are gadoversetamide, gadodiamide and gadopentetic acid, during MRI procedures should be avoided in patients with severe renal failure (kidney failure). In contrast, low-risk gadolinium contrast agents should only be used with minimum recommended doses. Later, the FDA issued a similar warning related to the association of GBCAs regarding NSF. It was also observed that most patients diagnosed with NSF had been exposed to GBCAs (Bardin & Richette, 2010). In 2014, a study performed by Kanda et al. demonstrated enhanced signal activities in some parts of the brain (dentate nucleus and globus pallidus) on the un-enhanced T_1 -weighted MR images of a patient (with normal renal function) who had previously received multiple doses of GBCAs during MRI procedure establishing the fact of gadolinium deposition in the brain (Kiviniemi et al., 2019). The results from the study by Kanda et al. have been shown in Figure 1.3.

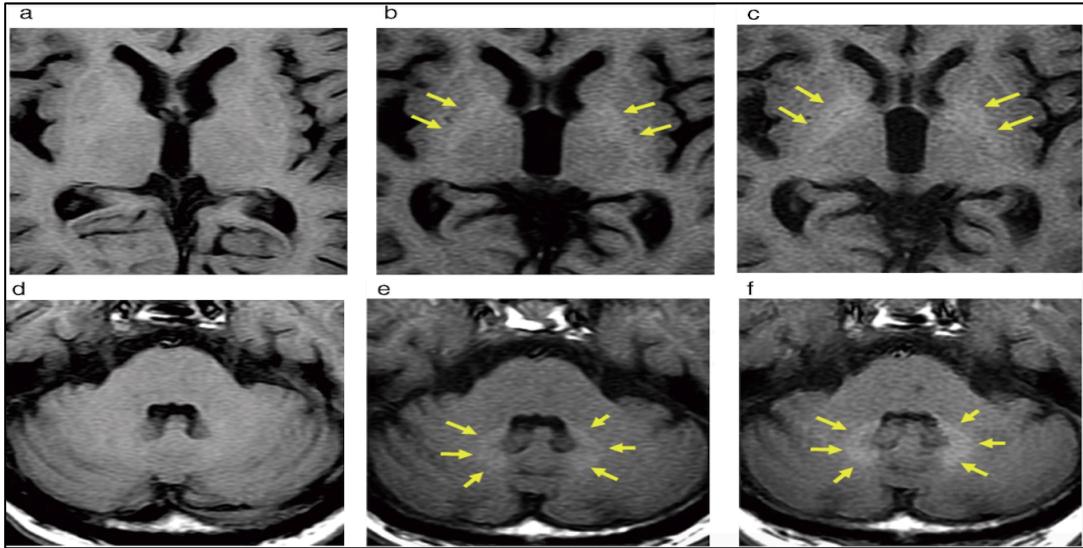


Figure 1.3 Progressive increase in the signal intensity of the globus pallidus (triangular-shaped subcortical structure of the brain) from images a to c and dentate nucleus (cerebellar cluster of neurons) from images d to f can be seen in patients who have a history of being administered with GBCA does in MRI examinations during the diagnosis of lung cancer (Kanda et al., 2014). This meant that for the MRIs taken during the study done by Kanda et al., the output scans were already contrast-enhanced. However, the patients were not administered any contrast-enhancing agent, thus proving gadolinium deposition in the brain.

Therefore, it has been recognised that with repeated usage of GBCAs, gadolinium can cross the blood-brain barrier and get deposited in the brain and other organs, even in patients with normal renal functions (Guo et al., 2018). The pieces of evidence have been gradually increasing since 2014 (Kanda et al., 2014). Pathological analysis has also discovered gadolinium deposits in the brain regions and extracranial tissues such as the liver, skin and bone (Kanda et al., 2014). The actual risks of gadolinium residual are still unknown and are being researched further, suggesting a precautionary proactive approach rather than a reactive approach.

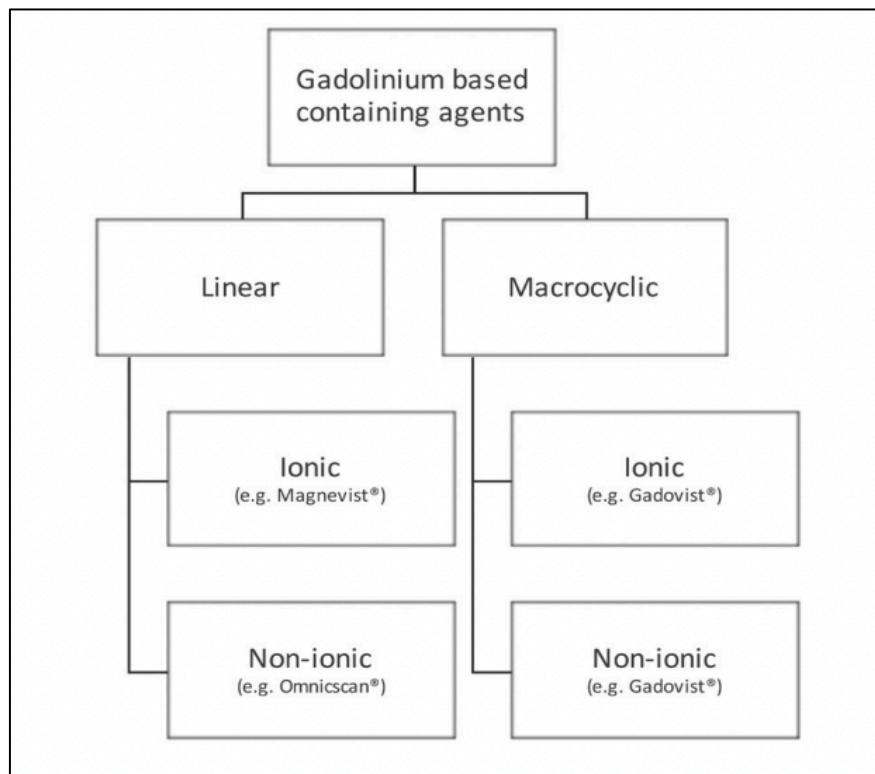


Figure 1.4 Classification of existing gadolinium-based contrast agents (Pullicino & Das, 2017).

Most cases of gadolinium deposition in the brain have been associated with instances where linear GBCAs have been used during the MRI procedure. The classification of GBCAs is shown in Figure 1.4. On the contrary, the administration of macrocyclic GBCAs has not been found to cause the same impact even with high doses (Radbruch et al., 2017). Nevertheless, concerns about any GBCAs used during MRI scans have been raised for any amount of

gadolinium deposition (Murata et al., 2016). Research continues to find gadolinium-free methods of getting enhanced MRI scans.

1.4 Integration of Deep Learning to Mitigate GBCAs

In recent years, machine learning (ML) techniques have paved a potential path in the biomedical and healthcare domain. Multiple studies have successfully identified deep learning (DL) architectures for extracting precise and accurate information from the MRIs of cancer patients, helping oncologists diagnose better. These advances in DL techniques have led to the development of novel methods for performing various tasks related to brain tumour analysis. These tasks include tumour identification, automated tumour segmentation, tumour grade classification, radio genomics, MRI-based survival prediction, and synthesis of contrast-enhanced MR images from non-contrast images, a highly discussed and researched area of interest. The present study draws inspiration from this particular task, supporting the effort to reduce the use of GBCAs by implementing artificial intelligence to produce contrast-enhanced MRIs. For example, the study by M. Gab Allah et al. introduced a convolution neural network (CNN) called the Edge U-Net model, inspired by the basic U-Net architecture to perform brain tumour segmentation in 2022. The study by Vankdethu and Hameed, 2022 also devised a novel combination of algorithms for tumour identification, tumour segmentation (K-means clustering) and tumour classification (Recurrent Convolutional Neural Network – RCNNs). The most relevant study identified that justifies the inspiration of our research is “Can Virtual Contrast Enhancement in Brain MRI Replace Gadolinium?” by Kleesiek et al., 2019 in which the researchers developed a Bayesian DL architecture that predicts CE MRI scans based on the learnings from multiparametric (multimodal) MRI data of non-contrast images as the input to the neural network.

The quality of its training data significantly influences the efficacy of a machine-learning model. In the past few years, the integration of artificial intelligence with magnetic resonance imaging (MRI) has seen a tremendous surge in research due to publicly accessible datasets. However, many datasets containing glioma MRIs offer limited MRI contrasts, usually T2, T2/FLAIR, and pre- and post-contrast T1, with inconsistent magnetic field strengths and acquisition parameters. To address this limitation, we have turned to the University of California San Francisco Preoperative Diffuse Glioma MRI (UCSF-PDGM) dataset recently made public to fuel AI advancements in diffuse glioma research (Calabrese et al., 2022). It encompasses data from 501 patients with verified diffuse gliomas (Calabrese et al., 2022). The dataset includes advanced diffusion and perfusion techniques and diverse imaging modalities. Key modalities analysed in this study are T1-weighted, T2-weighted, FLAIR-weighted, FA map, MD map, ASL image, and SWI, which are bias-corrected. These individual MRI sequences have been trained on a 2D U-Net architecture for CE-tumour identification, focusing exclusively on the tumour mask of the CE-tumour region (Figure 1.5). The evaluation metrics used are Dice Score, Intersection over Union (IoU), Peak-Signal-To-Noise-Ration (PSNR) and F2-score (Precision and Recall score). This gives us the hierarchy of MRI modalities that best identify the CE tumour on this dataset with the 2D U-Net DL architecture. Subsequently, we combined the MRI modalities (a maximum of 3 due to computational limits) and compared the results. Then, we repeated the exact methodology for each MRI modality to generate a post-contrast T1-weighted MR image from the respective non-contrast MR images. Additionally, the DL model was trained using a combination of the top two performing modalities from the previous results to predict T1w-CE MRI, facilitating comparative analysis of the model’s performance when presented with an enriched dataset.

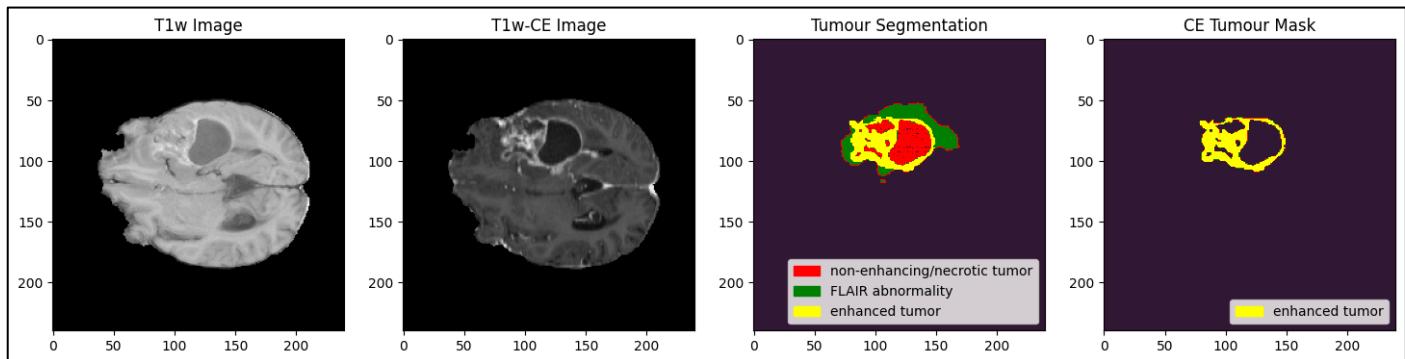


Figure 1.5 T1w, T1w-CE and Tumour Segmentation Images of a patient in UCSF-PDGM dataset. Only the CE Tumour mask (image 4) has been considered in this study to train the 2D U-Net model for CE tumour identification.

To summarise, through this study, we aim to explore the new publicly accessible UCSF-PDGM dataset and provide empirical evidence to avoid the usage of gadolinium-based contrast agents in MRI procedures. The research uses Artificial Intelligence techniques of Deep Learning algorithms, a 2D U-Net convolutional neural network, as a viable alternative to be trained for contrast-enhanced tumour identification and for generating synthesised contrast-enhanced MR images. This recommendation is motivated by the urgent need to prevent potential foreseeable risks associated with gadolinium accumulation in the cerebral structures of patients receiving GBCAs during brain tumour diagnostic procedures and having long-lasting adverse impacts due to gadolinium deposition.

2 Chapter 2: Background – Related Work

Multiple studies over the past few years have explored potential alternatives by examining deep learning methods to minimise or prevent the employment of Gadolinium-based contrast agents (GBCAs), aiming to avoid gadolinium accumulation in the cerebral region. This would allow the oncologists to stop administering GBCAs, thereby circumventing conditions like Nephrogenic Systemic Fibrosis (NSF) in brain tumour patients, especially ones with severe renal impairment. Therefore, in this section, we delineate some seminal works that resonate with our current investigation and are also inclined towards the primary purpose. We have summarized their respective designs with which they achieved the results in either tumour identification, segmentation or synthesising contrast-enhanced MR images.

2.1 Study 1: Synthesizing MR Image Contrast Enhancement Using 3D High-Resolution ConvNets by C. Chen et al., 2023

In this study, the researchers have designed a 3D high-resolution, fully convolutional neural network (FCN) to maintain high-resolution image information throughout processing. The study aimed to generate a contrast-enhanced T1-weighted MRI from pre-contrast MRIs, offering a potential alternative to mitigate gadolinium-based contrast agents (GBCAs) for CE-MRIs. The pre-contrast MRI sequences used in this study as inputs are T1, T2 and ADC (Apparent Diffusion Coefficient map) mapped to the post-contrast T1-MRI sequence as the target output using the design where three parallel subnetworks can generate and process multi-scale information in parallel. A distinguishing feature of this study is incorporating a local loss function specifically designed to augment the contribution from tumour regions, thus ensuring enhanced results for tumours. The study uses a dataset encompassing 426 MRI scans from 300 brain tumour patients. The study achieved a PSNR (Peak-Signal-To-Noise Ratio) of 28.24 dB in the brain and 21.2 dB in the tumour regions. The study mentioned limitations where the dataset didn't have precise tumour masks for training, and the designed model can produce better results by including more training data, especially non-tumorous patients, for adding high diversity. In summation, the results produced from this study have been quite promising and give us the confidence to pursue our current research on the more diverse dataset (UCSF-PDGM).

2.2 Study 1: Brain tumour MRI images identification and classification based on the recurrent convolutional neural network by Vankdothu & Hameed, 2022

In this study, the researchers have proposed an advanced automated system for brain tumour identification and a deep learning-based classification mechanism with enhanced accuracy. The study aims to help clinicians by providing them with an improved medical tool for more accurate brain tumour categorisation and potentially increase the survival rate of patients with a better prognosis. The process has been divided into various phases: image processing, image segmentation, feature extraction, and image classification (Vankdothu & Hameed, 2022). In the first stage of data pre-processing, an adaptive filtering algorithm was used to eliminate the noises from the MRI images before training the machine learning model. The second stage of tumour segmentation uses an improved version of the K-means clustering (IKMC) algorithm, presenting an advantage over the traditional K-means. It also states that the applied segmentation process takes less time and computational cost. The third stage is the feature extraction stage, where a gray-level co-occurrence matrix (GLCM) is used to extract essential features from the MRI images with high accuracy. The fourth and final stage of the system, tumour classification, has been performed using a recurrent convolutional neural network (RCNN) into different tumour types - glioma, meningioma, pituitary, or no tumour. The experiments were done on the Kaggle brain tumour dataset with 2870 training and 394 test images. The proposed RCNN method achieved higher accuracy (95.17%), sensitivity (98.42%) and specificity (89.28%) compared to other classifiers like backpropagation, U-Net and standard RCNN.

2.3 Study 3: Magnetic resonance imaging contrast enhancement synthesis using cascade networks with local supervision by Xie et al., 2022

In this study, the researchers proposed a cascaded sequential deep-learning neural network to eliminate the need for gadolinium-based contrast agents and reduce patient toxicity risk. The initial network employed is a retina U-Net model, which extracts semantic features of tumour regions from non-contrast MRIs. Subsequently, the output from this network with semantic feature maps is concatenated with the non-contrast MRI and introduced to the second network, a sequential synthesis model to generate synthetic contrast-enhanced MR images. The dataset used in this study contains MR images from 369 patients constituting T1-weighted pre- and post-contrast images. The findings indicate that incorporating tumour contours through the retina U-Net phase during training generates more precise predictions of T1-weighted contrast-enhanced MR images. This assertion is further substantiated by significant improvement in the study's quantitative evaluation metrics of normalised mean absolute error, structural similarity index, and Pearson correlation coefficient that were computed. The study also compares disparities between the ground truth and the generated images to see how different the enhanced visuals are with or without the information on tumour contours. The integration of the tumour contour information yielded better results, adding to the proposition of using this design to replace GBCAs during MRI procedures in the clinics.

2.4 Study 4: Deep-learning-based synthesis of post-contrast T1-weighted MRI for tumour response assessment in neuro-oncology: a multicentre, retrospective cohort study by Jayachandran Preetha et al., 2021

In this multi-institutional retrospective cohort study, the researchers have examined two distinct deep-learning methods for their capacity to generate contrast-enhanced T1-weighted MRI scans from pre-contrast T1-weighted, T2-weighted, FLAIR (fluid-attenuated inversion recovery) and ADC (Apparent diffusion coefficient map) MRI sequences as model inputs. The neural network architectures compared in this study include a 3D U-Net convolutional neural network and a 3D conditional General Adversarial Neural Network (CGAN). The collective dataset from all the participating clinics encompassed the data from 2061 glioblastoma patients, resulting in 6929 MRI scans (Moya-Sáez et al., 2023). Therefore, this study uses a cumulative dataset which fulfils the diversity factor and the appropriate size to train the models. During the 3D U-Net CNN model training, a combination of Mean absolute error (MAE) and structural similarity index measure (SSIM) was used. For the CGAN model, the generator's training phase maintained the discriminator's weights as constant. Conversely, during the discriminator's training, the generator remained static while the discriminator's weights underwent updates via backpropagation utilising discriminator loss. To encourage the generator to produce output images that closely resemble the target image in structure, both MAE and SSIM loss were also integrated. To determine whether the inclusion of ADC maps improves the performance of generating synthetic post-contrast T1-weighted sequences, both U-Net and CGAN models were trained and evaluated with and without ADC images as input, in addition to pre-contrast MRI modalities. This study, therefore, validates the feasibility of implementing deep-learning methods to produce post-contrast MR images from pre-contrast non-enhanced MRI sequences.

2.5 Study 5: Deep Learning Enables Reduced Gadolinium Dose for Contrast-Enhanced Brain MRI by Gong et al., 2018

This study proposed a deep learning method to reduce the prescribed GBCA dose levels during MRI procedures while maintaining full-dose contrast-enhanced MRI image quality and contrast information. The approach involves training an encoder-decoder Convolutional Neural Network (2D U-Net CNN) on processed zero-dose pre-contrast MRI and 10% low-dose post-contrast MRI scans, with the target output of a full-dose post-contrast CE-MRI scan. The study utilised only T1-weighted images, and the dataset consisted of 60 MRI brain scans, with 30 patients having mixed brain tumour grades and the remaining 30 being glioma patients. Out of the 60 scans, the deep learning model was trained only on the first 10 cases with diverse conditions, while the remaining 50 were part of the test set. To assess the results, the Peak-Signal-To-Noise Ratio (PSNR) and structural-similarity index (SSIM) were used for quantitative comparison, followed by paired t-tests and noninferiority tests on the quantitative comparison results. The study concluded that gadolinium doses can be reduced to 10% of the current dose level while maintaining decent image quality and contrast information. However, the study's limitation was using a single MRI modality, suggesting more MRI sequences were needed to train the model and improve CE predictions of target MRI scans. Also, using a 3D rather than a 2D model to achieve more enhanced prediction performance.

2.6 Study 6: Toward deep learning replacement of gadolinium in neuro-oncology: A review of contrast-enhanced synthetic MRI by Moya-Sáez et al., 2023

This review study congregates different studies (some have been mentioned above as well), discussing the advanced deep learning techniques used in those studies to synthesise contrast-enhanced MR images from pre-contrast MRI sequences. It also compares and discusses the evaluation methods and limitations or shortcomings in the respective reviewed studies to make them suitable for clinical deployability. It also sheds some light on the future research trends in this domain. Twelve manuscripts have been reviewed in this study collected from Scopus and Web of Science (WOS) repositories. The participant studies have been categorised into two groups: one that proposes a reduction in gadolinium dose levels (4 studies), and the other suggests bypassing the gadolinium injection completely (8 studies). This paper helped our research by collectively comparing deep learning methods that have been proposed and published based on their effectiveness. This allowed us to set a marking position for the results that our study should aim for while performing the contrast-enhanced tumour identification and synthesis of CE-MR images.

3 Chapter 3: Materials and Methods

3.1 Dataset Description

The current study uses - The University of California San Francisco Preoperative Diffuse Glioma MRI (UCSF-PDGM) dataset. In this dataset, 544 subjects were reviewed, of which 44 were excluded (severe image artifacts due to patient motion/hardware settings), and the remaining 501 were included. These 501 patients had been diagnosed with histopathologically proven (WHO grade 2-4) diffuse gliomas (Calabrese et al., 2022). The imaging protocol is highly standardised, and the preoperative brain tumour MRI scans were obtained in a 3-T three-dimensional imaging format, encompassing diffusion MRI, perfusion MRI, and tumour segmentations. The dataset also includes genetic data related to the tumour, treatment records, and survival statistics, making it an all-inclusive and comprehensive resource for analysis (Calabrese et al., 2022). The population within the dataset includes individuals who have undergone preoperative MRI, had their initial tumour removed, and underwent genetic testing for the tumour at a single medical centre between 2015 and 2021 (Calabrese et al., 2022). In the dataset, patients who had undergone any previous treatment for brain tumours were not included in the sample. However, having a history of tumour biopsy was not considered a reason for exclusion from the dataset (Calabrese et al., 2022). About the study participant demographics of the 501 patients, 11% had grade II tumours, 9% had grade III tumours, and 80% had grade IV tumours. Interestingly, there was a higher percentage of males (56%, 60%, and 60% for grades II-IV, respectively) than females in all the tumour grades (Calabrese et al., 2022).

In the dataset, the image acquisition of all the preoperative MRIs has been performed on a 3.0 Tesla Scanner (Discovery 750, GE Healthcare) and a dedicated 8-channel head coil (Invivo) (Calabrese et al., 2022). The multiple imaging modalities included in the dataset are 3D T2-weighted, FLAIR-weighted, susceptibility-weighted (SWI), diffusion-weighted (DWI), pre-and post-contrast T1-weighted images (Figure 1.2), 3D arterial spin labeling (ASL) perfusion images, and 2D 55-direction high angular resolution diffusion imaging (HARDI) based images (Calabrese et al., 2022). The HARDI data were also eddy current corrected. They processed the HARDI images using the Eddy and DTIFIT modules, which gave the final diffusion-weighted images (DWI) and their diffusivity maps of mean diffusivity (MD), axial diffusivity (AD), radial diffusivity (RD), and fractional anisotropy (FA). Two gadolinium-based contrast agents were used during the image acquisition: gadobutrol at 0.1 mL/kg and gadoterate at 0.2 mL/kg (Calabrese et al., 2022). The dataset also includes the tumour-segmented images with three major tumour labels: enhancing tumour, non-enhancing/necrotic tumour, and surrounding FLAIR abnormality. An example of the tumour segmentation image available in the dataset is shown in image 3 of Figure 1.5. All the image files are in NIfTI format. The complete UCSF-PDGM dataset is publicly available through The Cancer Imaging Archive (<https://doi.org/10.7937/tcia.bdgf-8v37>) (Calabrese et al., 2022). The description of the respective modalities present in the dataset has been mentioned below, along with a sample voxel displayed from the UCSF-PDGM dataset in Figure 3.1 for each of the non-contrast MRI modalities:

- **T1-weighted imaging:** A structural technique with a high-resolution depiction of brain anatomy, strongly contrasting grey and white matter, reflecting differences in the interaction of water with surrounding tissues (tissue T1 relaxation times) (UK Biobank Brain Imaging).
- **T2-weighted imaging:** A primary pulse sequence depicting differences in the T2 relaxation time of various tissues (Y. Chen et al., 2018).
- **FLAIR-weighted imaging:** A distinct fluid-attenuated inversion recovery sequence characterised by an extended inversion time. This sequence effectively suppresses the signal from cerebrospinal fluid in the resultant images. In FLAIR images, the brain tissue exhibits characteristics similar to the T2-weighted MRI sequence: the grey matter is brighter compared to the white matter, yet the cerebrospinal fluid (CSF) presents as dark, in contrast to its more glowing appearance in conventional T2-weighted images (<https://radiopaedia.org/articles/fluid-attenuated-inversion-recovery>).
- **Arterial Spin Labeling (ASL) Imaging:** It measures tissue perfusion (delivery of oxygen and nutrients to tissue through blood flow) using freely diffusible magnetically labelled arterial blood water protons as the intrinsic tracer (Petcharunpaisan, 2010).
- **Susceptibility Weighted Imaging (SWI):** It uses the magnetic susceptibility differences of multiple compounds like deoxygenated blood, blood products, iron and diamagnetic calcium (Halefoglu & Yousem, 2018).
- **Diffusion Weighted Imaging (DWI):** A method of signal contrast generation based on the differences in Brownian motion of the molecules driven by thermal energy known as diffusion. In our case, it evaluates the molecular function and micro-architecture of the human brain and its tissue (Baliyan et al., 2016). The UCSF-PDGM dataset has various quantitative diffusivity maps created using the mapping techniques of Apparent Diffusion Coefficient (ADC) and Diffusion Tensor Imaging (DTI) via diffusivity (diffusion sensitivity) and diffusion coefficient to understand the flow of direction of water molecules during the diffusion phenomena (O'Donnell & Westin, 2011) (Jacobs et al., 2008). In this dataset, we have the Mean Diffusivity (MD) map, Fractional Anisotropy (FA) map, Axial Diffusivity (AD) map and Radial Diffusivity (RD) map. All the images of DWI and diffusivity maps are eddy current corrected (Calabrese et al., 2022).

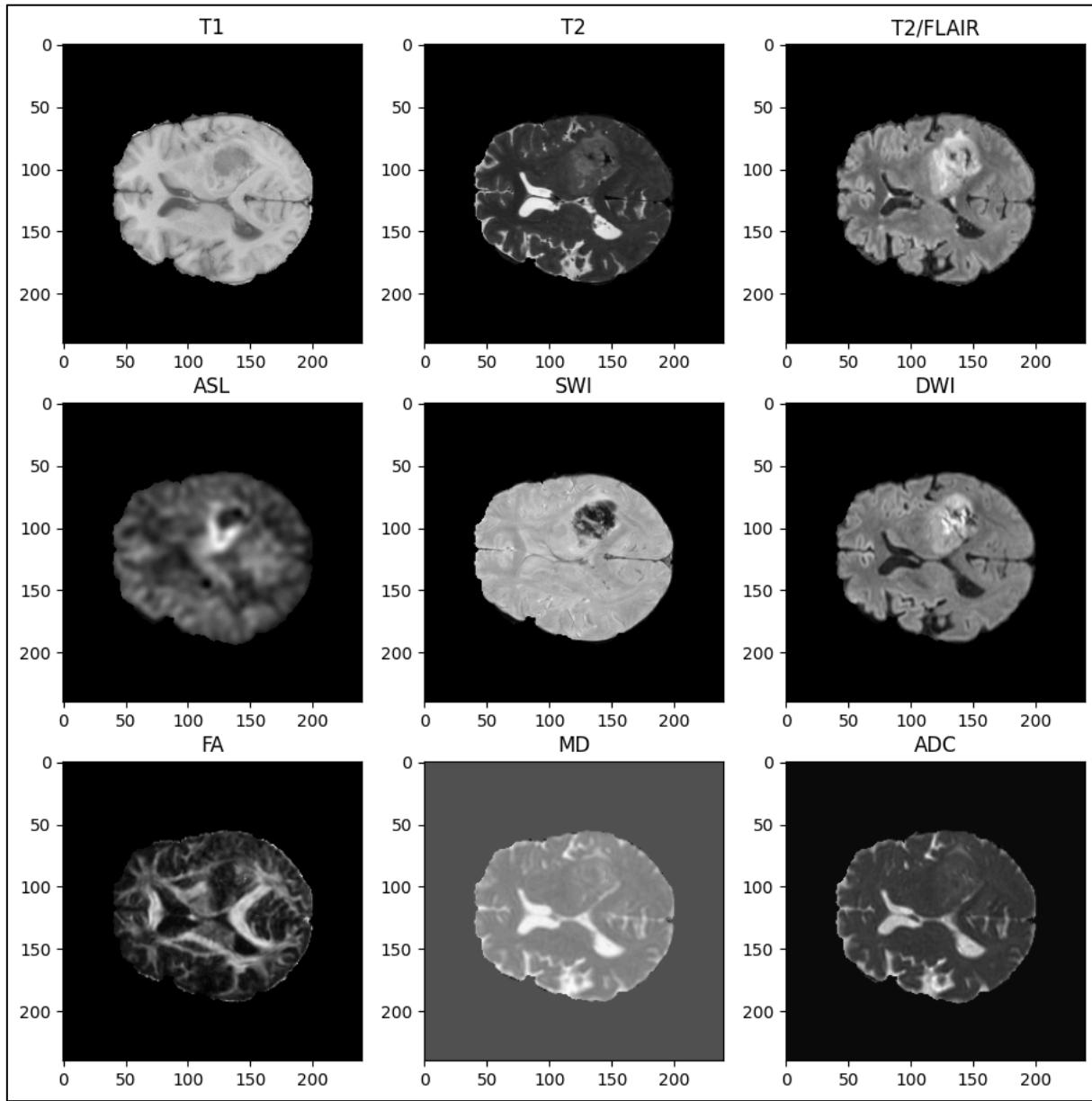


Figure 3.1 Multimodal MRI sequences of a patient with glioblastoma in the UCSF-PDGM dataset. T1 = T1-weighted pre-contrast, T2 = T2-weighted, T2/FLAIR = T2-weighted FLAIR, ASL = arterial spin labeling perfusion, SWI = susceptibility-weighted imaging, DWI = diffusion-weighted imaging, FA = fractional anisotropy map, MD = mean diffusivity map, ADC = apparent diffusion coefficient map (Calabrese et al., 2022).

3.2 Imaging Data Extraction

The UCSF-PDGM dataset is a comprehensive collection of MRI data, totalling 157 Gigabytes (157G) in size, that has been processed and analysed as part of this study. Each patient is assigned a dedicated folder containing MRI data files in various modalities, with each file following the naming convention UCSF-PDGM_XXXX_nifti, where 'XXXX' represents a unique patient ID. In this study, we have trained a 2D U-Net Convolutional Neural Network utilising the MRI data from 390 patients. The focus of the study was twofold: one was to identify contrast-enhanced tumours, and the second was to synthesise post-contrast T1-weighted MR images. Out of 501 patients, 111 (6+105) were not included for specific reasons. Six patients had discrepancies in their file naming conventions, which we excluded to maintain data integrity. Additionally, 105 patients were excluded due to the unavailability of tumour segmentation files or the indexing error of contrast-enhanced tumour segment labels. As a result, our study was streamlined to a refined set of 390 patients after excluding a list of pre-defined patient IDs.

3.3 Data Pre-processing

Among the 390 patients considered for the study, the MRI data comprises NIfTI format brain images across varied MRI sequences. Each brain image in the dataset is in grayscale with a singular channel. The MRI images exhibit a dimensional structure of $240 \times 240 \times 155$, where 240×240 represents the dimensions of one 2D brain image, while 155 represents the total number of image slices present in the entire voxel. However, in this study, we have only

extracted a subset of 15 image slices uniformly spread around the central slice (slice 77). We iterated through each eligible patient's directory and loaded the individual MRI sequence and its corresponding tumour segmentation image data for the same patient. We did not use all three tumour segment labels but only the contrast-enhanced ones. Therefore, we extracted a binary mask (CE-tumour label) representing the tumours' presence or absence. This selective approach resulted in an MRI dataset comprising 5850 2D images of the respective MRI sequence and its corresponding binary tumour mask. These images were consolidated into a comprehensive matrix (arrays) with a dimension of 5850 x 240 x 240 for both the feature (individual MRI sequence's matrix) and label (binary CE-tumour mask matrix) matrices. After constructing the required dataset, we normalised the intensity value of each extracted voxel from the image data. The normalisation process rescaled the pixel values to a range between 0 and 1 using an image's maximum and minimum pixel values. The formula used to normalise the pixel value is:

$$\text{Normalised value} = \frac{\text{Original pixel value} - \text{Minimum pixel value}}{\text{Maximum pixel value} - \text{Minimum pixel value}}$$

In this investigation, following the normalisation of the MRI data, we constructed the matrices X (representing input features) and Y (representing output labels). This was accomplished by reshaping pre-existing input features and target label arrays containing two-dimensional (2D) slices of MRI images. The reshaping introduced a fourth dimension, transforming each image into a three-dimensional (3D) volume comprising a single channel, equivalent to one voxel per image slice. The resulting array dimensions of both datasets were 5850 x 240 x 240 x 1. Subsequently, to prepare the data for our machine learning model, we then partitioned it. We initially divided the comprehensive dataset into a train-test subset and an unseen dataset, with the former accounting for 98% and the latter comprising 2%. We subdivided the primary 98% portion into a training dataset (70%) and a testing dataset (30%). We formulated the unseen dataset to evaluate the proficiency of our deep learning model when confronted with entirely unfamiliar MRI data. When considering utilising two distinct sequences, namely T1 and ASL, and three MRI sequences, including T1, T2, and ASL, the total number of compiled images amounted to 11700 and 17550, respectively. The data split for these cases adhered to the same ratio as observed in the individual MRI sequences ((98% train-test → 70% training data, 30% test data) and (2% unseen data)). The same methodology has been used to prepare the dataset for predicting contrast-enhanced T1-weighted MRI scans. Instead of extracting the CE-tumour mask, we use the CE T1-w images for the same 15 slices corresponding to the respective MRI sequence for the image-to-image translation. The data split ratio was the same as in the case of the CE-tumour identification task.

3.4 Deep Learning Architecture

Convolutional neural networks (CNNs) typically perform classification tasks with a single target or class label. In 2015, U-Net architecture was introduced by Ronneberger et al. to modify the way traditional CNNs work. The aim was to make the new architecture capable of solving visual tasks, especially in biomedical image processing, where each pixel in the image holds an important feature and its respective correlation to the target class label. As the name suggests, a U-Net model is a U-shaped neural network with two main sections—the contracting path, the bottleneck and an expansive path. A sample U-Net model is shown in Figure 3.2.

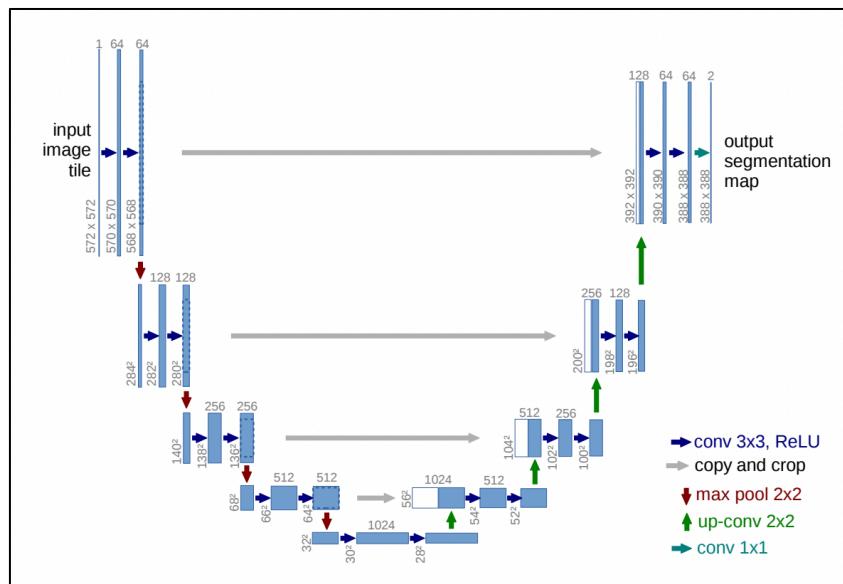


Figure 3.2 An example of the U-net architecture for images with a resolution of 32x32 pixels. The blue boxes represent multi-channel feature maps, with the number of channels shown at the top of each box. The x-y size is provided at the lower left edge of the box. White boxes indicate copied feature maps. The arrows represent the various operations involved. From the paper (Ronneberger et al., 2015).

The contracting path, also known as the encoder, employs convolutional layers to derive features and spatial context from the input image. The deeper the convolutional layer, the more intricate features and semantics it can capture. Following each convolution layer is an activation function that introduces non-linearity to the deep learning model, enabling it to comprehend complex, non-linear relationships between inputs and outputs. This is very crucial for tasks like ours of tumour identification, where the image data have intricate hierarchical structures. It also considers the gradient flow stability and avoids vanishing or exploding gradients. Furthermore, a max pooling operation is implemented to down-sample the feature map, decreasing the spatial dimensions of the input image. This spatial reduction enhances the model's ability to concentrate on more expansive contextual features and achieve computational efficiency.

After a contracting path, the U-Net architecture leads to the bottleneck section - the neural network's most profound segment, between the encoder and decoder paths. This section efficiently captures the most abstract representation of input features and boasts the most diminutive spatial dimensions among all network layers. The bottleneck section bridges the contracting and expanding paths, taking abstracted features and preparing them for refinement in the expanding course through up-sampling.

The decoder, also known as the expanding path, is the final component of the U-Net architecture. Its primary function is to restore the image data to its original size while the network learns to localise the input data with the help of previous layers. To accomplish this, the expanding path uses upsampling layers before each convolutional layer, which enhances the spatial resolution of the feature maps. This approach enables the network to capture the image's most intricate details, resulting in a segmentation map that aligns with the input image's spatial resolution. At each level of upsampling expansion, we incorporate a concatenation layer, which effectively merges the feature map of the corresponding contracting path layer. This is crucial in avoiding any spatial dimension mismatch that may occur. Combining the high-level contextual information from the deeper layer with the low-level, broader feature information from the earlier layers of the contracting path ensures that the convolutional layer can access information from both image processes. The expanding path in the U-Net architecture ends with a final convolutional layer to generate the resultant output image with the required number of output channels. Generally, a sigmoid (for binary segmentation task) or a softmax (for multiclass segmentation task) is used as an activation function for this final convolutional layer.

3.4.1 Model Design

In this research, we utilised a 2D U-Net Convolutional Neural Network (CNN), a deep learning architecture, to identify contrast-enhanced tumours and predict contrast-enhanced T1-weighted MRI. The 2D U-Net architecture is depicted in Figure 3.3. This architecture is segmented into three pivotal components: the contracting path, a bottleneck, and an expanding path. The input image dimensions passed to the U-Net model stand at 240 x 240 x 1, where 240 x 240 represents the 2D image size, whilst 1 refers to the singular channel, as all input images are grayscale. The model encapsulates as deep as eight layers of two-dimensional convolution as part of the configuration.

The contracting block comprises six convolution layers and three max-pooling layers. Two 2D convolution layers are stacked in one block, provisioned with 64 filters and a 3 x 3 kernel size. The convolution layers use the Rectified Linear Unit (ReLU) as the activation function, mathematically represented by $f(x) = \max(0, x)$, infusing non-linearity into the network. Notably, no additional padding has been added to the input feature map. Moreover, each convolution layer uses the He-normal weight initialiser as the kernel initialiser. After these layers, a max-pooling layer with a stride of 2 x 2 is introduced, which downsamples the spatial dimension of the input image by half, resulting in a feature map of size 120 x 120. This block of 2 convolution layers and 1 Max-pooling layer is repeated three times in the contracting path. In block two, the convolution layers use the same parameter attributes as in the previous setup but with a different filter size of 128. The down-sampled feature map size was 60 x 60 after the max-pooling layer. In block three, the filter size is increased to 256, keeping all the other parameters the same. The max-pooling layer condenses the feature map to a size of 30 x 30.

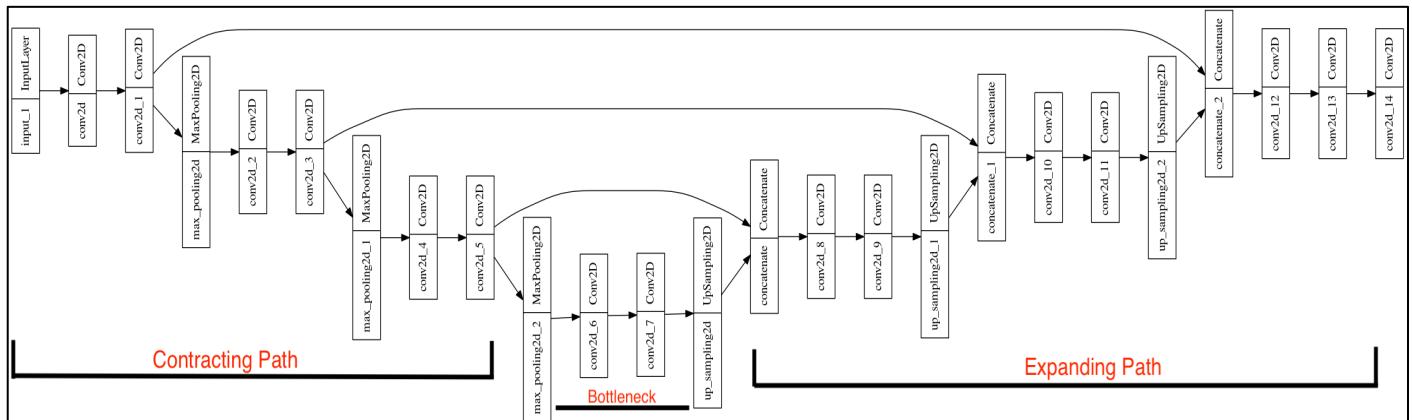


Figure 3.3 The 2D U-Net Convolutional Neural Network used in this study for CE-tumour identification and CE-T1 MRI synthesis.

The bottleneck component of the 2D U-Net consists of two 2D convolutional layers with a kernel size of 3x3 and 512 filters each. The expanding path also consists of three blocks, each with an up-sampling layer, followed by a concatenation layer and two 2D convolution layers. The up-sampling layer increases the spatial dimension of the feature map by a factor of two, doubling the sizes from $30 \times 30 \rightarrow 60 \times 60 \rightarrow 120 \times 120 \rightarrow 240 \times 240$. The concatenation layer combines the up-sampled feature map with the corresponding convolution layer in the contracting path. After this, we have two consecutive two-dimensional convolution layers similar to the contracting path, sharing the same parametric attributes of kernel size, padding, and kernel initialiser. The only different configuration is the filter size, which goes in the reverse order of the contracting path, from 256 \rightarrow 128 \rightarrow 64, since we are re-tracing the size to generate the output with the exact spatial dimensions as the original input image. Our final two-dimensional output convolution layer is at the end of the 2D U-Net architecture. This layer adds the desired singular grayscale channel to the spatial dimension of the output image. After the final convolution layer in the third block of the expanding path, the image size was 240 \times 240 \times 64. The final output convolution layer converted this to 240 \times 240 \times 1 (same as the original input image).

3.4.2 Training the Model

The training phase is a crucial element in creating a deep learning model. Proper selection of the loss function and performance metric is essential to track progress accurately after each iteration. Our study meticulously defined the model and compiled it with specific attributes. The Adam optimiser was utilised as the learning rate optimisation algorithm with a default learning rate of 0.001. The model's performance was monitored via the binary cross-entropy as the loss function and the dice coefficient (or dice score) as the accuracy metric of choice. This enabled us to identify CE-tumour or predict CE T1-weighted MR images from different or combined MRI modalities of pre-contrast MR images. In brief, our 2D U-Net deep learning neural network consists of 1 input layer, 12 convolution layers (which include the contracting path, bottleneck, and expanding path), 3 max-pooling layers, 3 up-sampling layers, 2 concatenation layers, and one output convolution layer. The network consists of 7,781,761 trainable parameters, and we trained it with a batch size of 16 for 100 epochs. Our neural network was trained using TensorFlow and Keras framework, CUDA 11.7, and Python 3.10.0 on a GNU/Linux (x86_64) server with an NVIDIA Corporation's GPU with graphics processor GM200, specifically the GeForce GTX TITAN X, which has 12GB memory and 3072 cores.

During the training phase of CE-tumour detection using MRI scans, the average time to train with one type of MRI scan was 166 minutes using a dataset of 4013 images. However, when two types of MRI scans, T1 and ASL, were used, the training time increased to 317 minutes with a dataset of 8026 images. Using three types of MRI scans, T1, T2, and ASL, prolonged the training time even further to 563 minutes with a dataset of 12039 images. When predicting CE-T1 images from individual MRI scans, the mean training duration was 165 minutes with a dataset of 4013 images. However, when both T1 and T2 modalities were integrated, the training time increased to 318 minutes with a dataset of 8026 images. All the MRI images in these datasets were 240 \times 240 \times 1 in size.

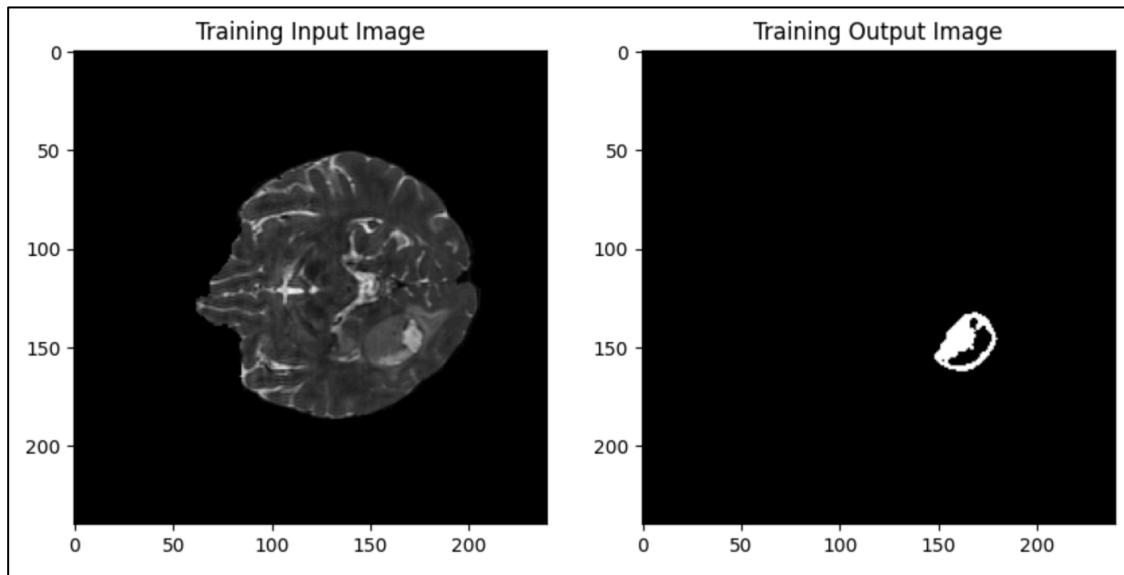


Figure 3.4 A Sample Training input image (T2 modality) (img. 1) and Training output image (CE tumour mask from tumour segmentation image) (img. 2) in the CE-tumour identification task. These images are the single slices (slice number = 1000) of dimension 240 \times 240 in the training dataset that will be mapped to each other during the neural network training process.

3.4.3 Testing the Model

The development of our deep neural network involves a vital testing phase. This phase allows us to assess the model's ability to generalise and avoid overfitting during training. Testing the model on new and unseen data gives us valuable insights into its real-world performance and lets us compare different models and configurations. To ensure accurate results, the test set must be kept separate from the training data to preserve its integrity. By analysing errors based on the test set, we can identify any shortcomings in the model and improve its accuracy. In our study, we evaluated the performance of our 2D U-Net model by fitting it to the test set using the same attributes of the Adam optimiser, binary entropy loss function, and dice score accuracy metric. We also assessed the model's performance using metrics of IoU (Intersection over Union), PSNR (Peak-Signal-To-Noise-Ratio), and F2 Score (Precision and Recall). By computing the average values of these metrics on our test set, we gained valuable insights into the model's performance.

For the specific task of CE-tumour identification, the test suite consisted of 1720 MR images evaluated individually based on each MR sequence. The dataset was expanded to 3440 MR images when the MRI modalities of T1 and ASL were combined. A comprehensive test dataset was created by integrating T1, T2, and ASL modalities, encompassing 5160 MR images at its peak. When synthesising CE-T1 images, an individual MRI modality's test set consisted of 1720 MR images, while the conjunctive modality of T1 and T2 employed a dataset constituting 3440 MR images.

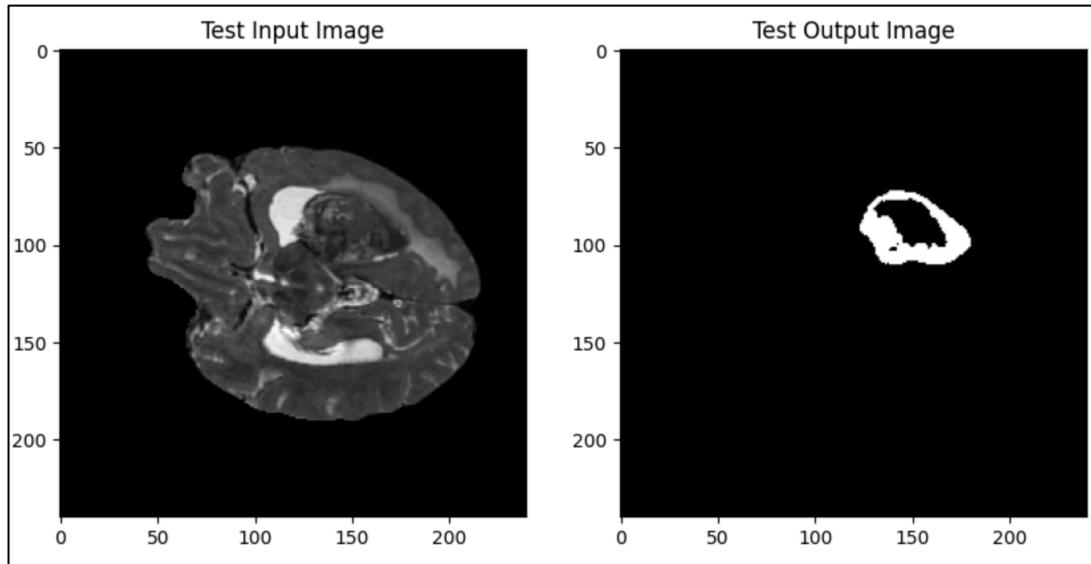


Figure 3.5 A Sample Test input image (img. 1) and Test output image (img. 2) in the CE-tumour identification task from T2-weighted MRI sequence. These images are the single slices (slice number = 650) of dimension 240 x 240 in the test dataset. The trained model will take img. 1 as input to predict img. 2 as the output.

4 Chapter 4: Results

In this section, we will showcase the results obtained from our 2D U-Net deep learning method. This method was used to identify CE (contrast-enhancing) tumours and create CE-T1 MR images from a wide range of MRI modalities, either separately or in combination. To evaluate our model's performance, we calculated the binary cross entropy as the loss criterion, and we also analysed its efficacy using different assessment metrics such as Dice Score, IoU, PSNR, and F2 Score. Based on these evaluations, we identified the best modality combinations compared to the results obtained from individual MRI sequences. For CE-tumor identification, we combined ASL + T1 and ASL + T1 + T2 modalities based on the ranking of individual performance, while for CE-T1 prediction, we combined T1 + T2. The motivation was to check whether adding more diversity to the dataset improves the model's performance, and therefore we captured the evaluation scores of these combinations as well. It is important to note that we have also captured the analytical results in the CE-tumour identification task using the post-contrast T1 image. Due to its contrast-enhanced nature, it is ideally the best input to identify CE-tumour. Therefore, this particular result has been used as a benchmark for comparison in the CE-tumour identification task.

4.1 Binary Cross Entropy Loss Function

Binary Cross Entropy, or the Logarithmic Loss function (Log loss), is primarily used for binary classification problems where it measures the difference between the actual distribution (ground truth) and the estimated distribution (model's prediction). It tracks a model's incorrect labelling of the data class and penalises the model if a probability error occurs in classifying the labels (Amber Roberts, 2023). Hence, Binary Cross Entropy computes the negative average of the log-likelihood of the corrected predicted probabilities (Shipra Saxena, 2021). The below-mentioned equation defines the log loss computation.

$$\text{Binary Cross Entropy loss} = - \frac{1}{N} \sum_{i=1}^N y_i \cdot \log(p(y_i)) + (1 - y_i) \cdot \log(1 - p(y_i))$$

In the above equation, y_i is the actual class and $\log(p(y_i))$ is the probability of that class. Here, $p(y_i)$ is the probability of 1, meaning a perfect prediction and $(1-p(y_i))$ is the probability of 0, indicating an incorrect prediction for all N points (Amber Robert, 2023).

While training our 2D U-Net model on the MRI data of one or more MRI sequences, we used the binary cross-entropy loss for our two-fold task (CE-tumour identification and CE-T1 MR image synthesis). Figure 4.1 showcases the loss calculated on the Test dataset for the MRI modalities included in this study. As anticipated, T1-CE exhibited the lowest loss (highlighted in green) while identifying CE-tumour. However, ASL (highlighted in brown) outperformed the other modalities and emerged as the best individual modality for predicting the CE-tumour mask. On the other hand, the FA map (highlighted in red) had the highest loss value among all the MRI sequences. When predicting CE-T1, the loss values for all MRI sequences are similar, but the T1 and T2 dataset combined has the lowest loss value (highlighted in green) and is the most effective modality. On the other hand, the Mean Diffusivity map has the highest loss value (highlighted in red) and performs the worst. The individual computed loss values are mentioned in Table 1 highlighted with the same colours.

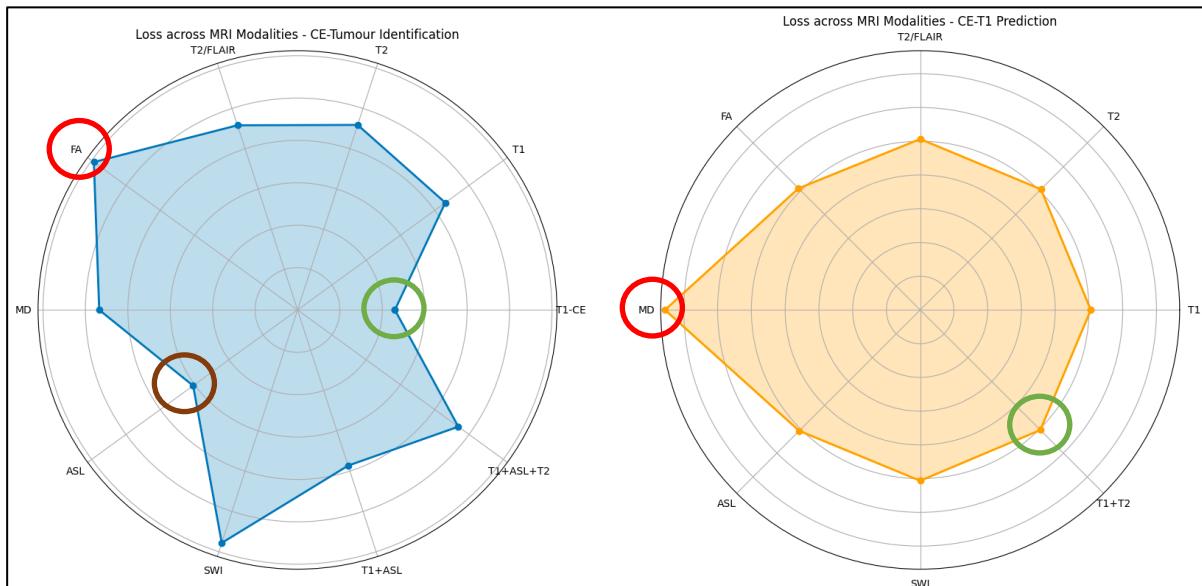


Figure 4.1 Binary Entropy Loss comparison for different MRI modalities researched in this study. Img.1 is associated with the CE tumour identification task, and img.2 is related to the CE-T1 Prediction task. The maximum loss score is highlighted with a red circle, and the minimum loss score is highlighted with a green circle.

Table 1 Binary Entropy Loss values for the respective MRI modalities, individual and combined. The red highlight is for maximum loss values, and the green highlight is for the minimum loss value.

| MRI Modality | Loss – CE Tumor Identification | Loss – CE T1 Prediction |
|--------------|--------------------------------|-------------------------|
| T1 | 0.008613954 | 0.126335993 |
| T2 | 0.009181632 | 0.126489386 |
| T2/FLAIR | 0.009162759 | 0.126624793 |
| FA | 0.011871704 | 0.12740764 |
| MD | 0.009351098 | 0.188964084 |
| ASL | 0.006084108 | 0.126698762 |
| SWI | 0.011562707 | 0.126608655 |
| T1+ASL | 0.007726193 | - |
| T1+ASL+T2 | 0.00937504 | - |
| T1+T2 | - | 0.125854507 |
| T1-CE | 0.004593843 | - |

4.2 Model Performance Evaluation Metrics

As mentioned earlier, we employed multiple performance metrics to evaluate the proficiency of our 2D U-Net deep learning model on the test dataset. These metrics encompassed the dice similarity coefficient, Intersection over Union (IoU), Peak Signal to Noise Ratio (PSNR), and F2 score. The dice score was computed as a model evaluation component via the binary cross-entropy loss. Conversely, the three metrics were evaluated per image within the test dataset, with the average results being analysed for the respective metrics.

The Dice Similarity Coefficient (Sørensen-Dice similarity coefficient) quantifies the similarity between two data samples. To elaborate, it is a statistical validation metric that evaluates the model's performance to reproduce the segmentation and the spatial overlap accuracy with the ground truth of the input MR image (Zou et al., 2004). It has been one of the most broadly used metrics for evaluating binary image segmentation tasks. In this study, we computed the dice coefficient separately based on the equation:

$$DSC = \frac{2 |X \cap Y|}{|X| + |Y|}$$

where $|X \cap Y|$ is the cardinality of the intersection of sets X and Y. The value of the dice score ranges between 0 and 1, where 1 indicates perfect overlap between the ground truth and the prediction, whereas a score of 0 indicates no overlap. The Intersection over Union (IoU) metric, also known as the Jaccard Index, is a helpful tool for calculating the similarity between two images based on the object of interest. It measures the overlap between a predicted truth and the ground truth. A score of 1 indicates that the projected and the ground truth match perfectly, while 0 means no overlap between them. To use the IoU metric in our study, we have used the **jaccard_score** function from the scikit-learn library. The next metric used in the study is the Peak-Signal-to-Noise-Ratio (PSNR), which calculates the PSNR ratio in decibels to measure the quality of the image between the original image and the resultant image. The higher the PSNR value, the better the output image's quality (Nadipally, 2019). The PSNR value is computed using the **psnr** function in the scikit-image library. The last metric used in this study is the F2 score or F-beta score. The F2 score is a particular case of the F-beta score where the parameter beta = 2. It is a metric that weighs the balance between precision and recall. The **F-measure** is calculated as the harmonic mean of precision and recall, giving each the exact weighting. The **F-beta** score is a generalisation of the F-measure that adds a configuration parameter called beta, which, when equal to 2, gives less weight to precision and more weight to recall in the score calculation (Jason Brownlee, 2020).

$$F_\beta = (1 + \beta^2) \frac{precision * recall}{\beta^2 * precision + recall}$$

The **fbeta_score** function from scikit-learn is used to compute the F2 score where the parameter **beta=2** specifies the F2 score, emphasising recall to avoid false-negative errors of tumour identification.

The above-defined evaluation metrics have been plotted using bar charts in Figure 4.2 and Figure 4.3 to compare the performance scores calculated for individual and combined MRI sequences for our 2D U-Net model. Figure 4.2 plots the comparison for the CE-tumour identification task, and Figure 4.3 plots the comparison for the CE-T1 prediction. In both graphs, the x-axis represents the MRI modality, and the y-axis represents the score in the range of [0,1]. The PSNR values are not in the range of [0,1], but to plot them alongside the other metrics, we have normalised them. That is why the FA map's worst PSNR is 0.00, although its original value is not 0. The actual values are shown in Table 2 and Table 3 to elaborate on the comparison.

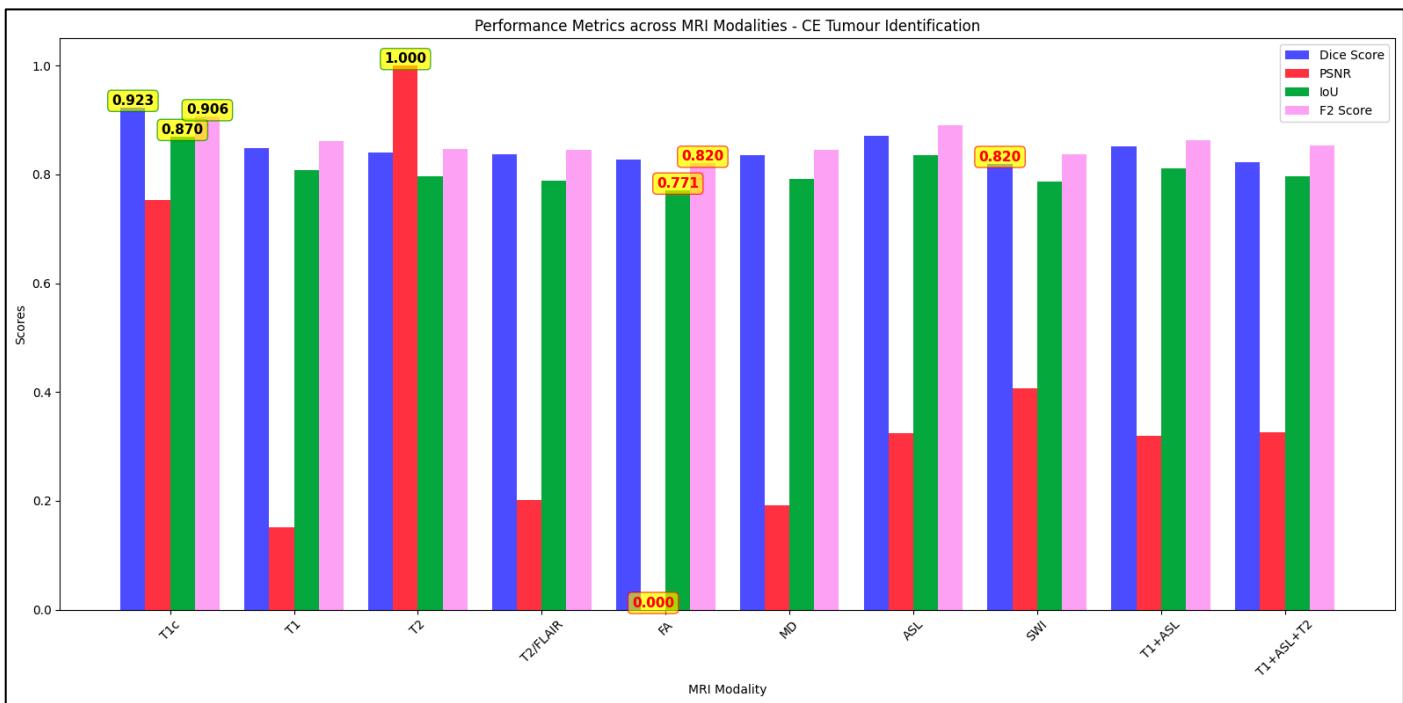


Figure 4.2 A Bar chart comparing individual and combined MRI modalities based on Dice Score, PSNR, IoU and F2 score for CE-tumour identification. The best scores are of T1c (highlighted with a green-yellow box), and the worst (highlighted with a red-yellow box) are of FA and SWI mixed with FA, leading to the lowest score in the three metrics.

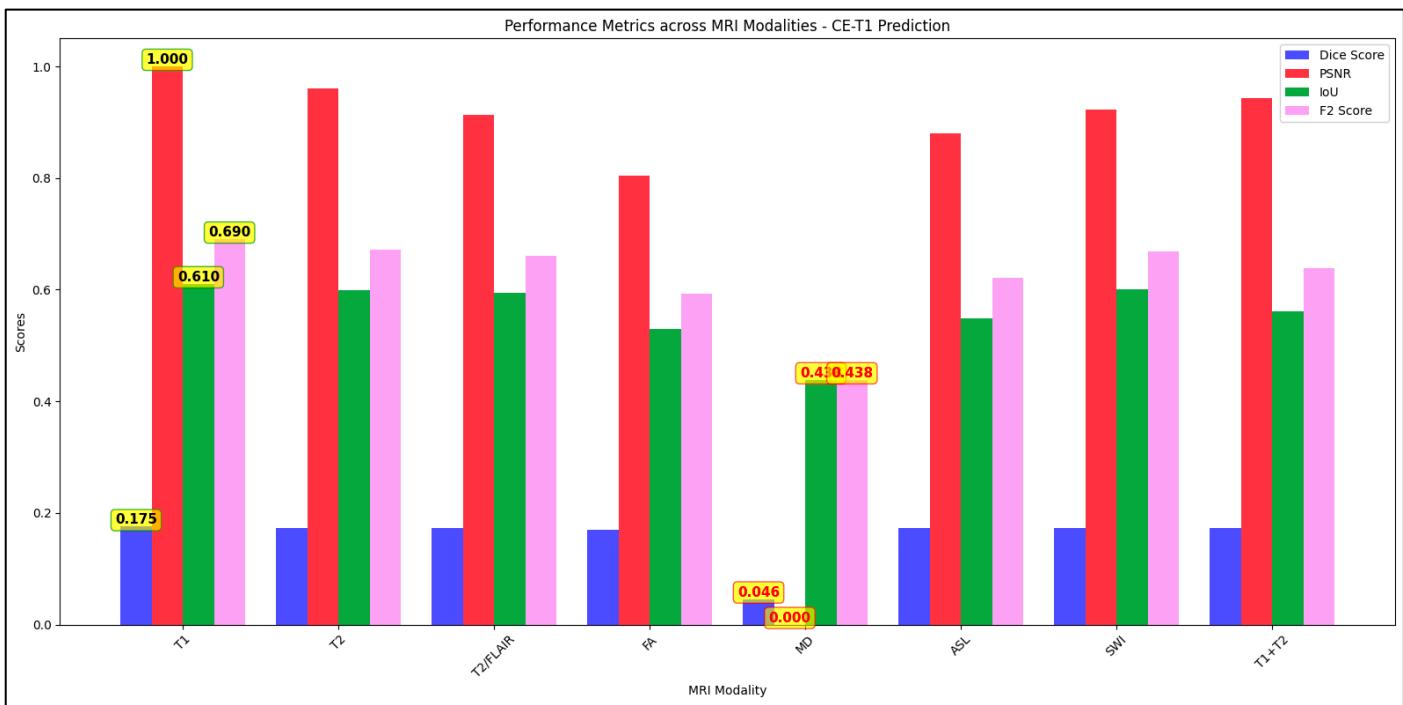


Figure 4.3 A Bar chart comparing individual and combined MRI modalities based on Dice Score, PSNR, IoU and F2 score for CE-T1 Prediction. The best score is modality T1 (highlighted with a green-yellow box), and the worst modality MD (highlighted with a red-yellow box).

Table 2 The computed evaluation metric scores for CE tumour identification. We highlighted the best score with green and the worst with red without considering the score of T1c, as it is our benchmark for this task.

| MRI Modality | Dice Score | Mean IoU | Mean PSNR | Mean F2 score |
|----------------------|-------------|-------------|-------------|---------------|
| T1c | 0.92316079 | 0.86962738 | 74.9611377 | 0.90601018 |
| T1 | 0.848762453 | 0.808101019 | 68.42549542 | 0.860876537 |
| T2 | 0.840190768 | 0.796178088 | 77.63102044 | 0.846572433 |
| T2/FLAIR | 0.836364865 | 0.788804539 | 68.97854895 | 0.84503893 |
| FA | 0.826800168 | 0.771095369 | 66.79157076 | 0.820343571 |
| MD | 0.835350394 | 0.791590499 | 68.87436669 | 0.845212871 |
| ASL | 0.871505797 | 0.835052399 | 70.3098744 | 0.88999564 |
| SWI | 0.82005602 | 0.7866867 | 71.2088755 | 0.83706085 |
| T1 + ASL | 0.85171884 | 0.81189651 | 70.2560522 | 0.86233347 |
| T1 + ASL + T2 | 0.82250899 | 0.79588275 | 70.3216274 | 0.85243003 |

Table 3 The computed evaluation metric score for CE-T1 image prediction. We highlighted the best score with green and the worst with red.

| MRI Modality | Dice Score | Mean IoU | Mean PSNR | Mean F2 score |
|-----------------|-------------|-------------|-------------|---------------|
| T1 | 0.17535222 | 0.61036436 | 37.8754496 | 0.69045612 |
| T2 | 0.173564374 | 0.599253392 | 37.26330605 | 0.672194911 |
| T2/FLAIR | 0.172662511 | 0.593691238 | 36.55746932 | 0.660000497 |
| FA | 0.169296578 | 0.529713129 | 34.87678675 | 0.592678406 |
| MD | 0.045757707 | 0.438372093 | 22.61133848 | 0.438372093 |
| ASL | 0.17280668 | 0.549396571 | 36.05473999 | 0.62091422 |
| SWI | 0.173397347 | 0.600534522 | 36.70397357 | 0.668910874 |
| T1 + T2 | 0.17264546 | 0.56091111 | 37.0107426 | 0.6378187 |

According to the scores outlined in Table 2 and Table 3, it has been determined that the three most effective MRI modalities for detecting CE tumours are ASL, T1, and T2, following the same order. Conversely, when producing CE-T1 MR images, T1 and T2 are the top two MRI modalities in the same order. However, there is no massive difference between the performances of other modalities except those that have performed the worst, like the FA map in CE-tumour identification and the MD map in CE-T1 MR image prediction. The captured metric measurements have been impressive while identifying the CE-tumour mask, which justifies our basic 2D U-Net model's ability. Even more encouraging is that all the dice scores are above 0.8 (Table 2), indicating that our model has learned to identify tumours quite efficiently from the original MRI sequence images. To validate the results, we have displayed multiple figures below to help us validate the MRI modalities that produced better results. The MR images used in the figures belong to the Test dataset. For better comparison, we have used the example of the same slice (input image number 4) being predicted in different MRI modalities.

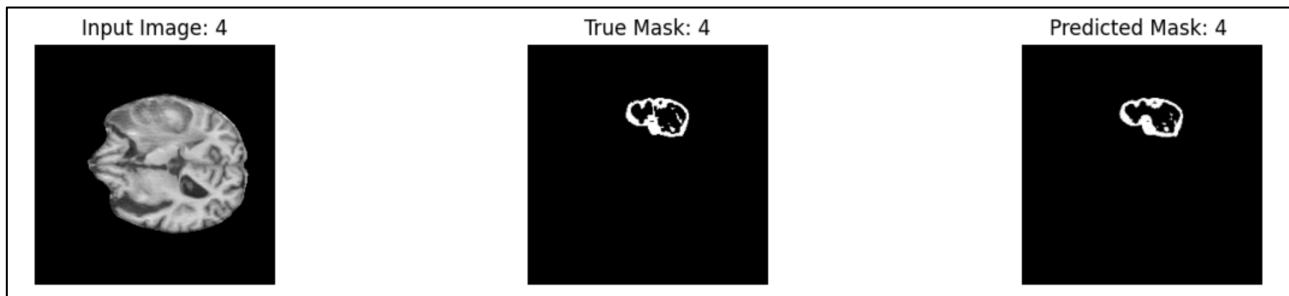


Figure 4.4 T1 MRI sequence used to identify CE Tumour mask.

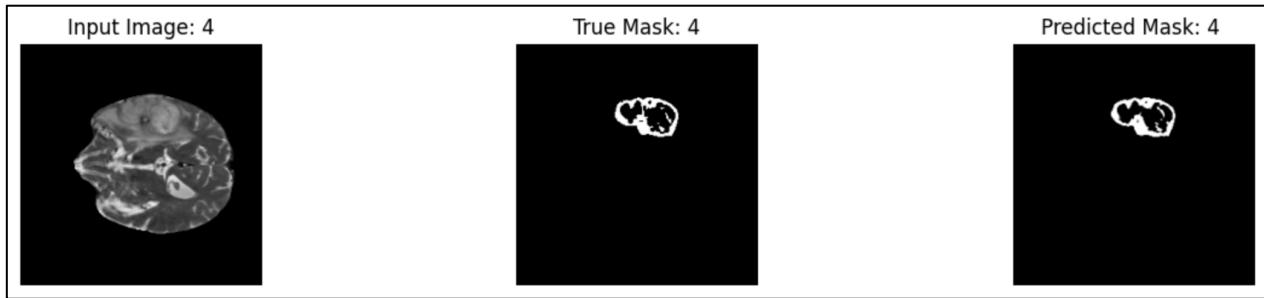


Figure 4.5 T2 MRI sequence used to identify CE Tumour mask.

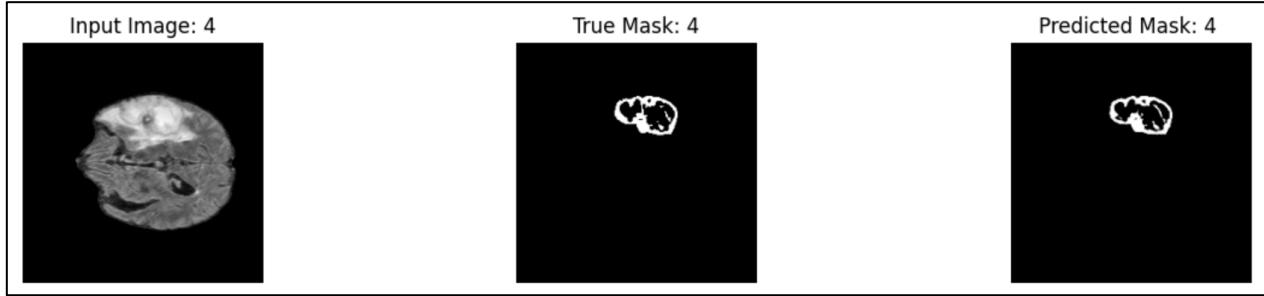


Figure 4.6 T2/FLAIR MRI sequence used to identify CE Tumour mask.

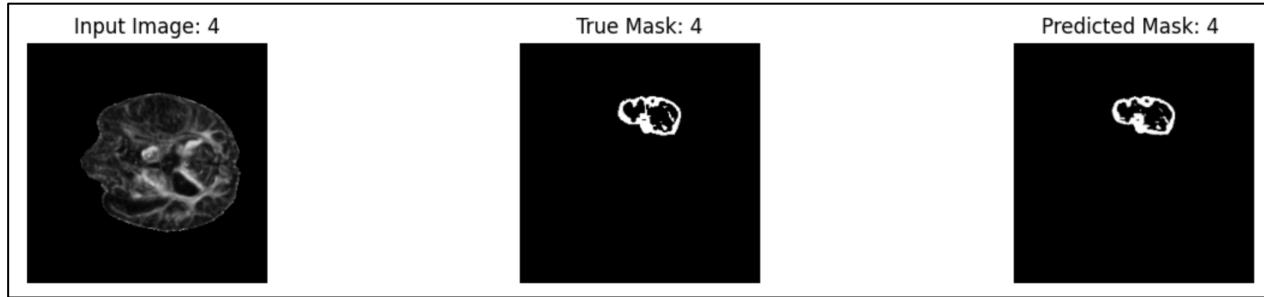


Figure 4.7 FA map used to identify CE Tumour mask.

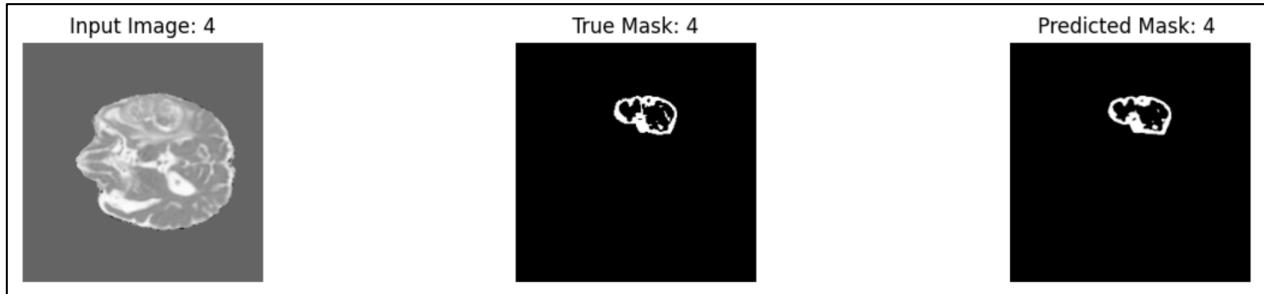


Figure 4.8 MD map used to identify CE Tumour mask.

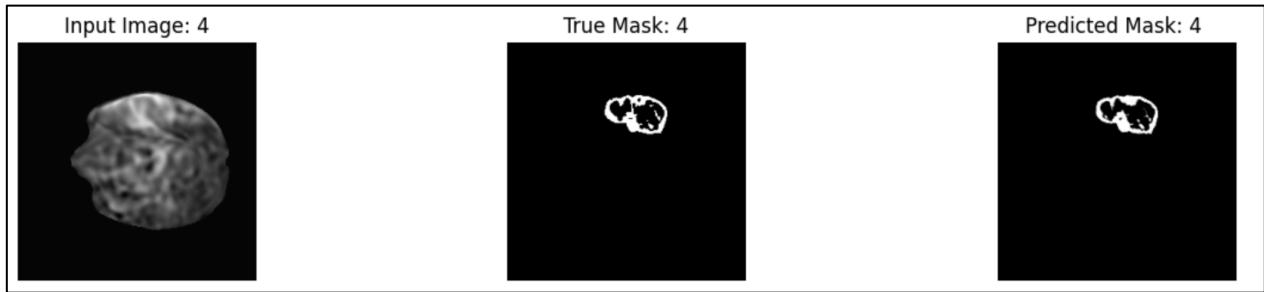


Figure 4.9 ASL MRI sequence used to identify CE Tumour mask.

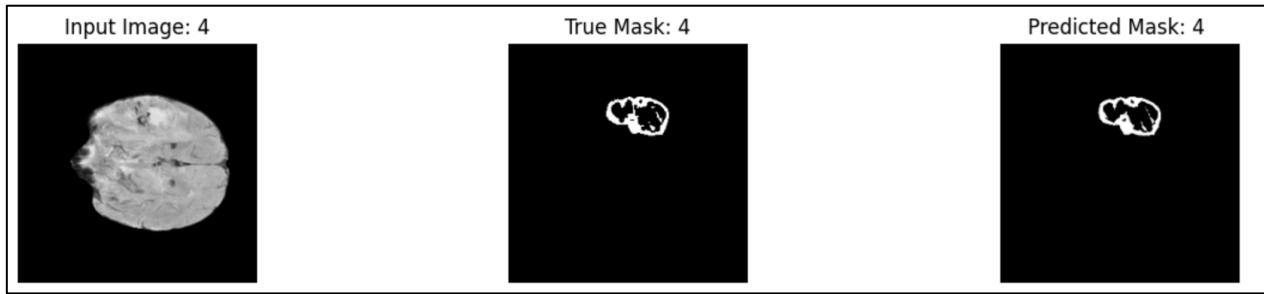


Figure 4.10 SWI MRI sequence used to identify CE Tumour mask.

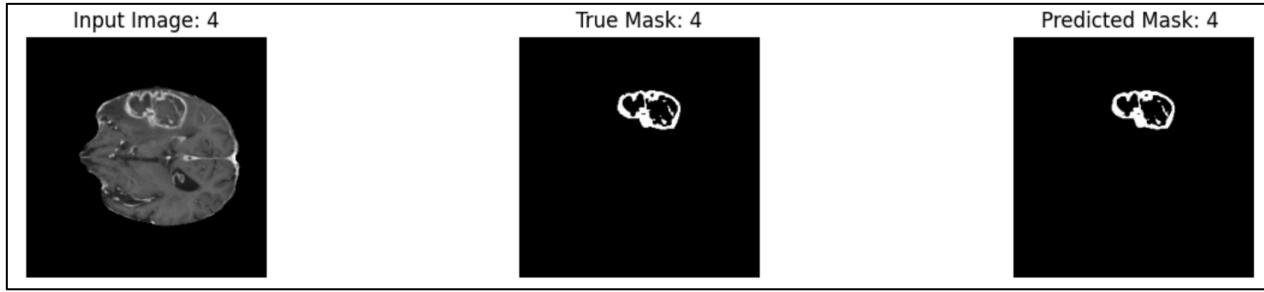


Figure 4.11 CE-T1 MRI sequence used to identify CE Tumour mask.

Upon examining the output images of the CE tumour mask, we have noticed that the CE-T1 MRI image is highly accurate and precise in producing contrast-enhanced tumour visuals, as expected. According to the metric scores, ASL is the second-best option for achieving optimal results. However, visually, most MRI techniques have correctly identified the tumours, with only slight variations in their scores compared to ASL. The 2D U-Net model has shown impressive performance with its simple architecture for identifying CE tumours. Furthermore, when we combined different MRI modalities, we observed positive results. The figures presented below (Figure 4.12 and Figure 4.13) clearly and accurately demonstrate the effects of combining modalities. However, the numerical metric scores are approximately identical to the individual modalities. We had hoped for better scores due to the addition of diversity in the training dataset, but unfortunately, that did not significantly improve the results.

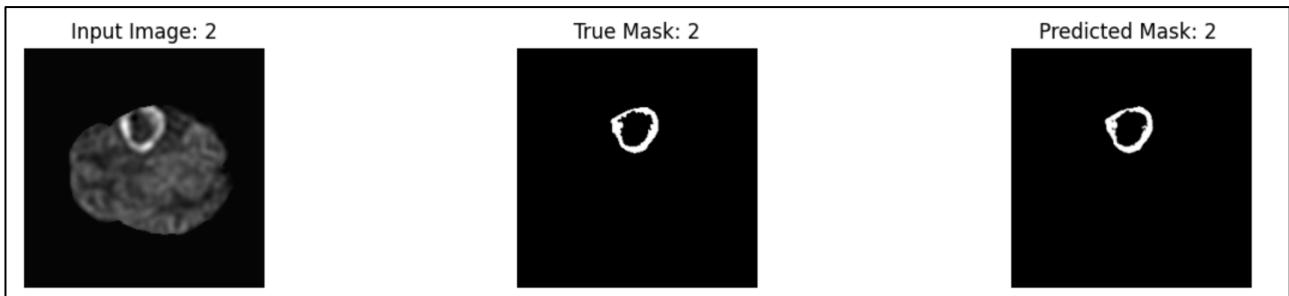


Figure 4.12 Identification of CE tumour mask by our 2D U-Net model trained on ASL + T1 MRI sequences.

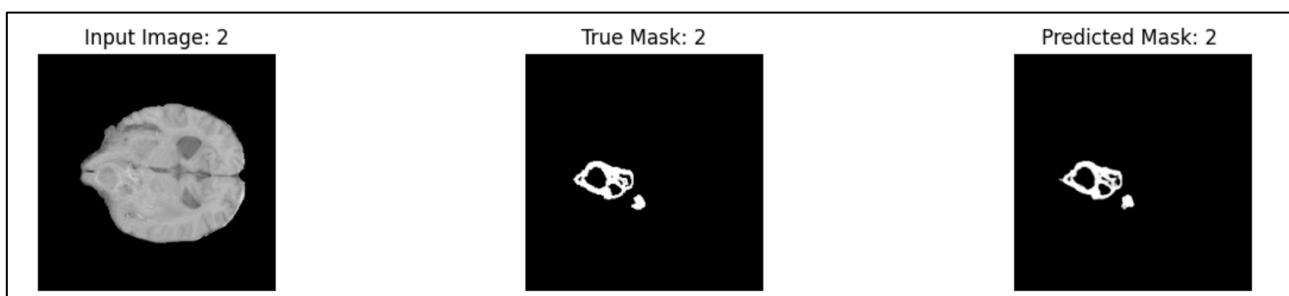


Figure 4.13 Identification of CE tumour mask by our 2D U-Net model trained on ASL + T1 + T2 MRI sequences.

We aimed to use our deep-learning model to predict CE-T1 MR images. Despite the relatively low metrics results shown in Table 3, we made predictions to see if we could at least distinguish the contours of the brain image and improve our model for future projects. This will help us achieve better future outcomes in predicting CE-T1 MR images.

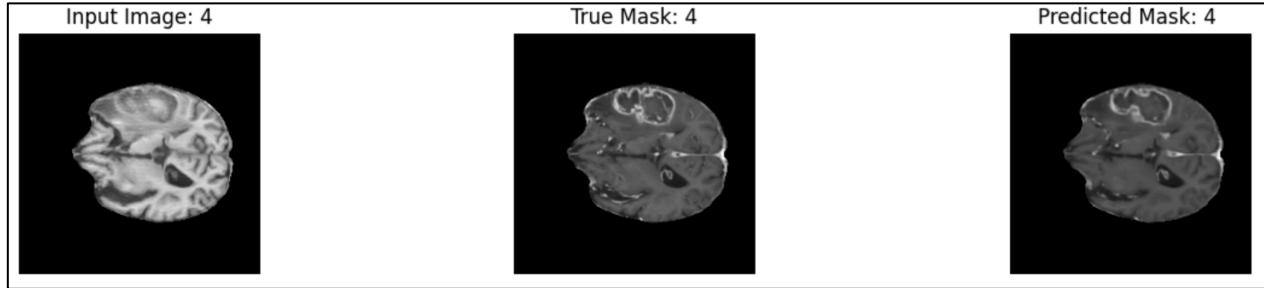


Figure 4.14 Predicting CE T1 MRI from pre-contrast T1 MRI sequence.

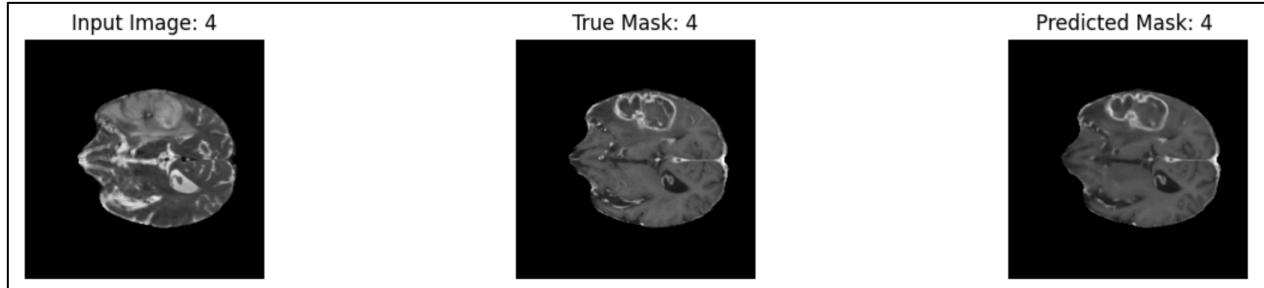


Figure 4.15 Predicting CE T1 MRI from pre-contrast T2 MRI sequence.

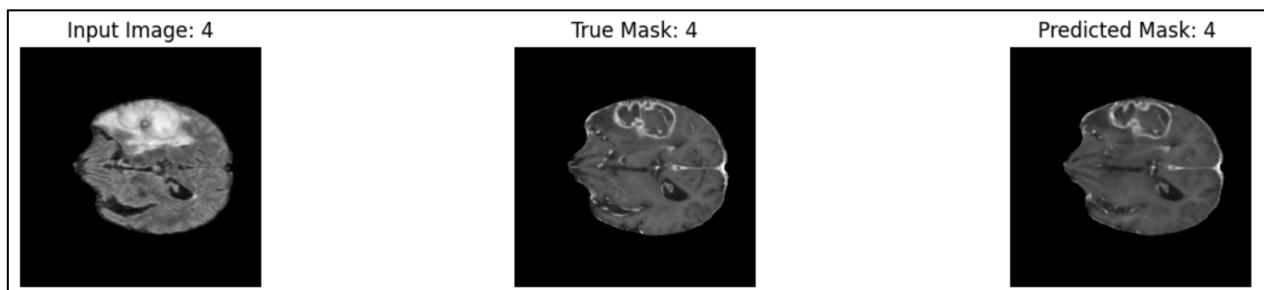


Figure 4.16 Predicting CE T1 MRI from pre-contrast T2/FLAIR MRI sequence.

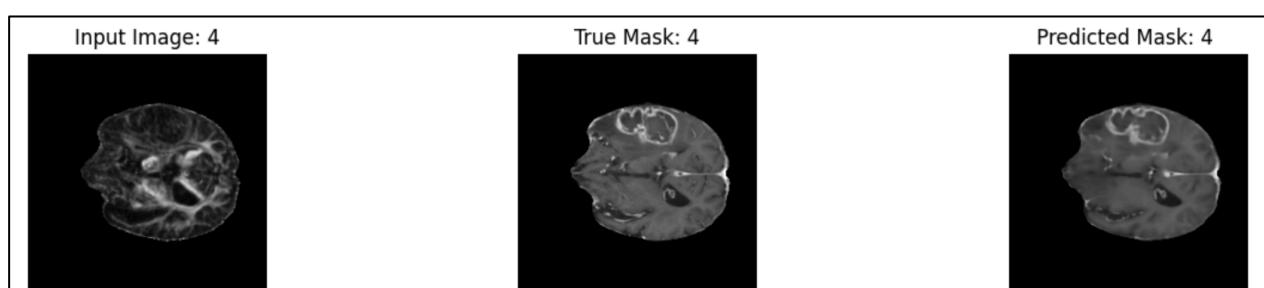


Figure 4.17 Predicting CE T1 MRI from pre-contrast FA map.

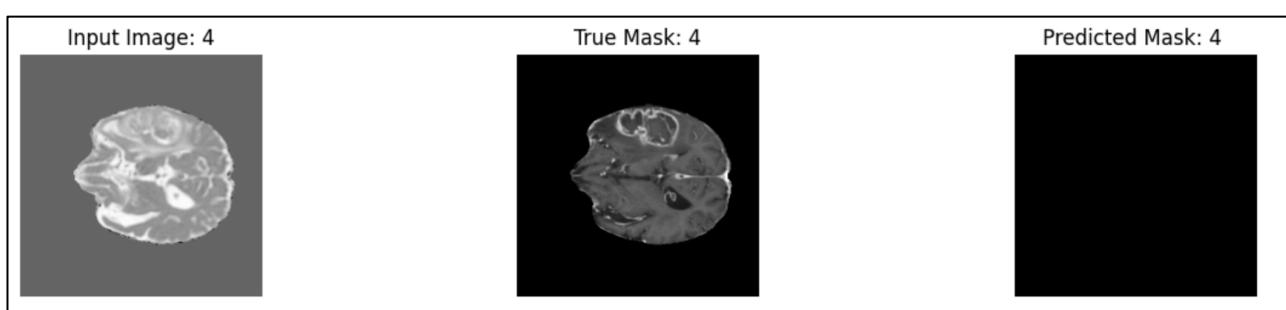


Figure 4.18 Predicting CE T1 MRI from pre-contrast MD map. In this case, we don't get any reasonable output.

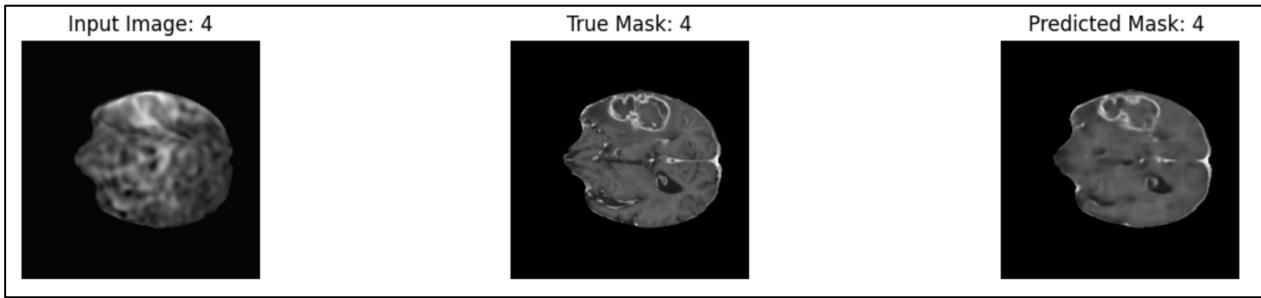


Figure 4.19 Predicting CE T1 MRI from pre-contrast ASL MRI sequence.

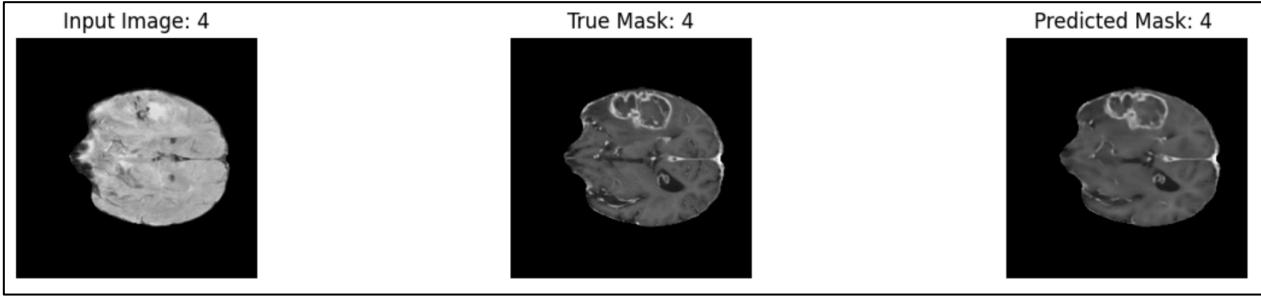


Figure 4.20 Predicting CE T1 MRI from pre-contrast SWI MRI sequence.

Our analysis showed that the MRI modalities yielded less than optimal results than the CE-T1 MRI image, with low scores for dice, IoU, PSNR, and F2. Despite this, we did observe some encouraging yet ambiguous progress in the output image. The predicted output image shows a distinct contrast enhanced tumour section. It's worth noting that the MD map did not produce any output image, highlighting the need for additional scrutiny in the training dataset associated with this modality. Combining the modalities of T1 and T2 to predict the CE-T1 image gave us similar metric measurements. The expected output image in Figure 4.21 is promising but could be more exact. The intricate details of the image are way off compared to the actual CE-T1 MRI.

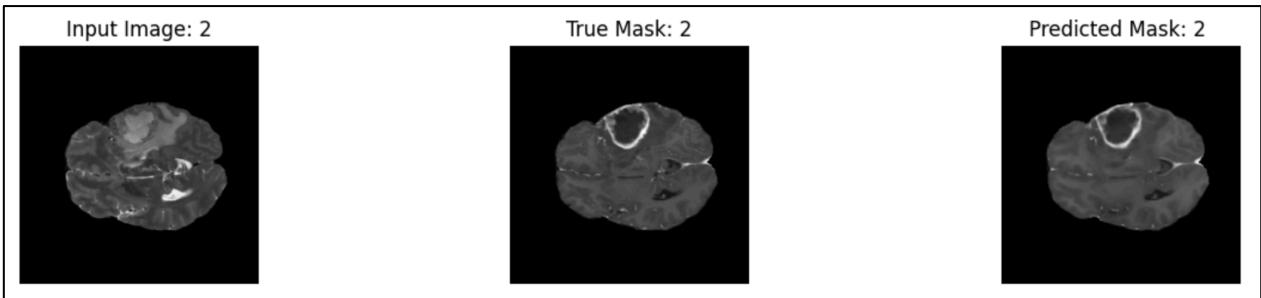


Figure 4.21 Predicting CE T1 MRI by our 2D U-Net model trained on T1 + T2 MRI sequences.

Ultimately, we tested our 2D U-Net model on a dataset it had not seen before to ensure it was adequately evaluated. We assessed its performance on individual images and compared the results across different modalities using the dice score metric. The dataset comprised 117 image slices for each MRI modality in the CE tumour identification task. We recorded the computed dice scores of the unseen data in a CSV file and visualised the distribution on a graph. This allowed us to confirm that the top-performing MRI modalities on the test set were also the best on the unseen data. The worst-performing modality had consistent performance across both datasets. Our findings were consistent with Figure 4.22, which showed that the contrast enhanced T1 MR images had the best dice score distribution plot, followed by the ASL MRI modality, then T1, and finally T2. Our results were therefore validated.

Moreover, Figure 4.23 showcases the distribution of dice scores computed on the CE-T1 MR image synthesis task's unseen data. This graph validates our earlier findings on the test dataset. The MD map demonstrates the weakest performance with the lowest dice score distribution. At the same time, the graphs for the remaining modalities fall within the range we identified after scrutinising our tabular data (Table 3) on the test dataset. T1 and T2 exhibit slightly superior predicted dice scores than other modalities, but the variance is negligible.

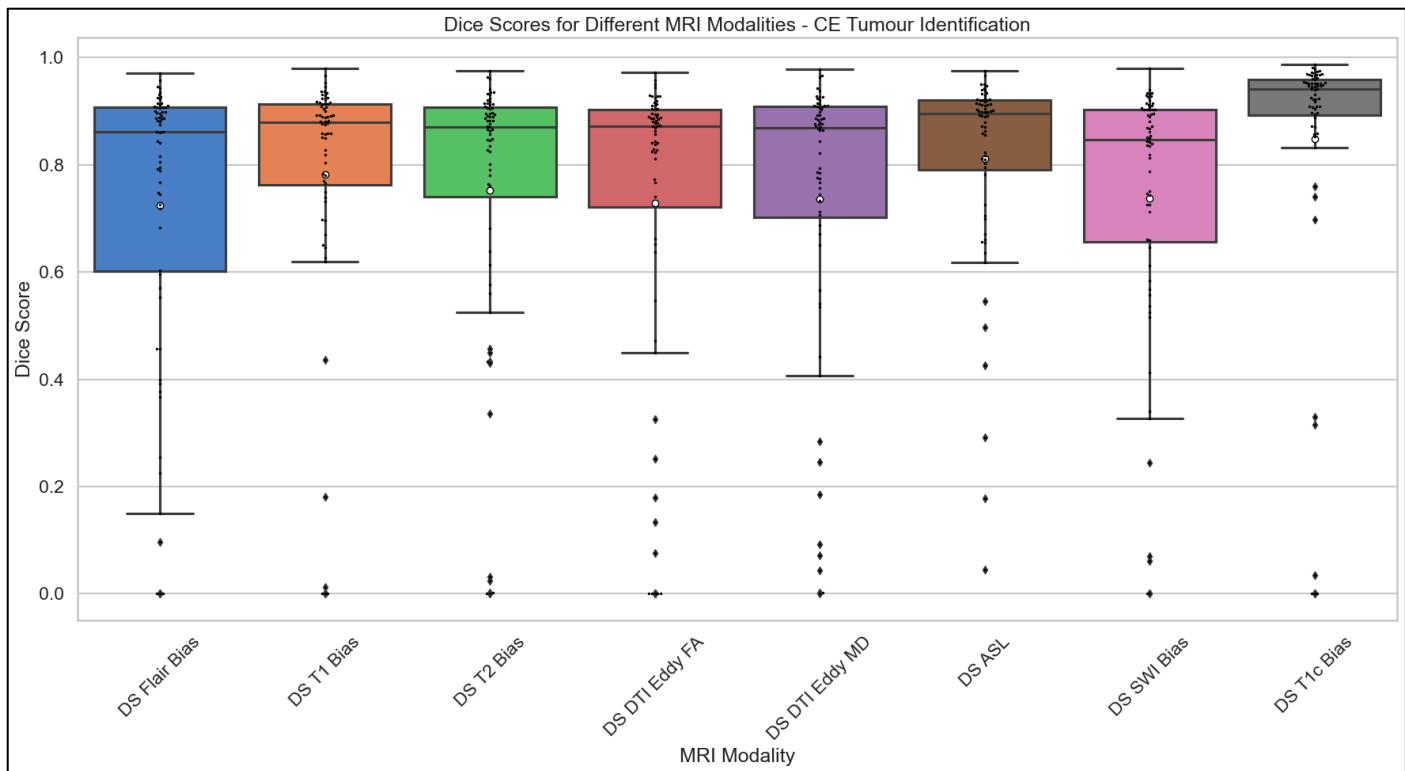


Figure 4.22 Plotting the data distribution of dice score for different MRI modalities while identifying CE tumours.

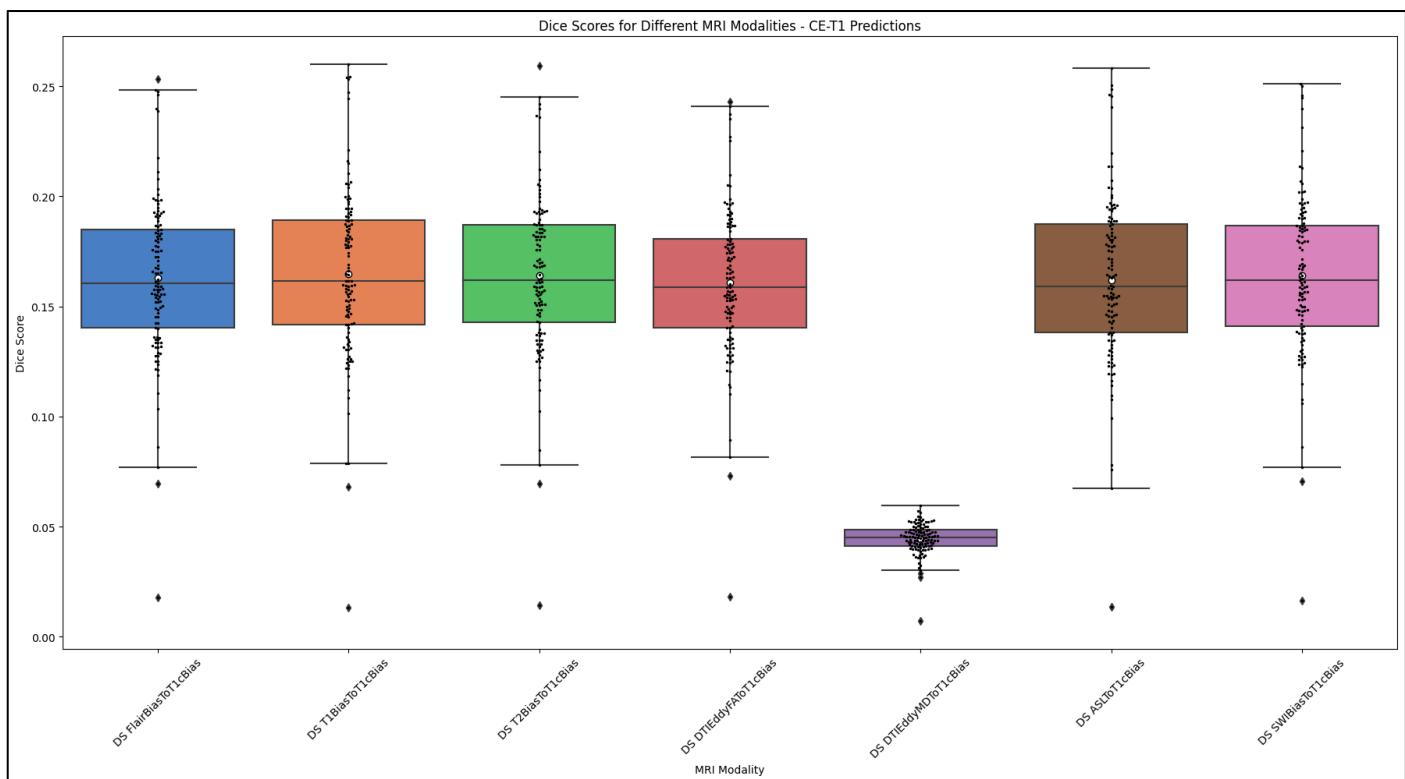


Figure 4.23 Plotting the data distribution of dice score for different MRI modalities while predicting CE-T1 MR images.

5 Chapter 5: Discussions

Our study used the UCSF-PDGM dataset of glioma patients and implemented a 2D U-Net deep learning architecture with multi-modal MRI data. Various studies have presented innovative deep-learning architectures that can efficiently produce synthetic contrast-enhanced MR images, which we have briefly discussed in this study. We analysed this vast MRI dataset - UCSF-PDGM - to assess its quality by comparing the different MRI sequences and determining their effectiveness in identifying contrast-enhanced tumours or generating contrast-enhanced T1-weighted MR images when fed into our neural network model. The objective of our study was to build the groundwork for designing a new, efficient deep learning architecture in the future that can generate contrast-enhanced MR images without the usage of Gadolinium-based contrast agents, which have a toxic side effect of gadolinium deposition in the brain. Our research addressed the primary inquiry of determining the optimal number and type of input modalities required to produce precise CE-weighted images. In this study, we evaluated the individual modalities and then combined the top modalities to see if there was an increment in the model's efficiency.

In this research, out of an initial pool of 501 patients, a subset of 390 patients was utilised. The exclusion of 111 patients was necessitated by data inconsistencies, specifically about identifying CE tumour masks; these patients needed more information in the dataset. From the 390 selected patients, 15 slices were chosen, concentrating on the mid-section of the brain image to maximise informational content within a limited slice range. While it would have been ideal to incorporate all 155 slices of a given image, computational constraints dictated by the resources allocated for this project made it untenable. A more extensive image set would have enhanced the model's learning capacity. Furthermore, to maintain uniformity in conditions across both tasks, the patient count was retained at 390. Otherwise, the job of synthesising synthetic CE-T1 MR images could have benefited from a larger patient sample during training. Still, our analysis deserves credit for using many images from different modalities to train the model. For one task, we used 17550 MR images as input.

In our research, we utilised the binary cross-entropy loss function to train our 2D U-Net model, which was purpose-built for identifying CE tumours. The nature of this classification problem required determining the presence or absence of a tumour, making binary cross-entropy a fitting choice. Our findings validated this decision. However, when it came to predicting CE-T1 images, binary cross-entropy proved to be less effective. The task necessitated the identification of multiple classes, rendering binary cross-entropy suboptimal. For image-to-image translation tasks with regression characteristics, alternative loss functions such as Mean Squared Error or Mean Absolute Error may better capture subtle differences in image synthesis.

Our research is centered on identifying the most fitting evaluation metrics. In our pursuit of detecting CE tumours, we opted for the Dice Score as our primary metric due to its alignment with binary classification assessments. This decision yielded highly effective results. However, as we ventured to enhance our model to predict CE-T1, the appropriateness of the Dice Score came into question. Since this task encompasses a broader range than binary classification, metrics customised for binary assessments, such as the Dice Score, may only capture some subtleties.

To enhance the accuracy of CE-T1 image predictions, it would be prudent to incorporate metrics such as Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM). PSNR thoroughly evaluates the model's efficacy in transforming one image into another. At the same time, SSIM gauges the perceived variance in structural details between the generated and original images, thereby providing a holistic assessment of the quality of the generated images.

5.1 Limitations and Suggestions

Regarding deep learning, the hardware used to train algorithms is critical to their effectiveness. Powerful computational capabilities are essential for deep learning models, so Graphics Processing Units (GPUs) are a must-have tool. GPUs are particularly well-suited for deep learning because they handle thousands of threads simultaneously. This is crucial for complex operations, such as matrix multiplications and convolutions. The system's computational power is closely tied to any deep learning algorithm's training speed, overall performance, and success, making high-powered GPUs an essential component. In this study, the computational GPU resource available was the NVIDIA Corporation's GPU with graphics processor GM200, specifically the GeForce GTX TITAN X, which has 12GB memory and 3072 cores. The current 2D U-Net model could use the entire GPU space at 100% capacity. We could not train our model with a batch size of more than 16. Therefore, computational constraints were one of the most crucial limitations in this study.

In the initial phases of our project, we aimed to educate our 2D U-Net model to recognise CE tumour masks across all modalities. We planned to subsequently employ the same pre-trained model to grasp the post-contrast synthesis of MR images. Regrettably, we had to train the model distinctly for each MRI modality, as we encountered computational constraints and needed more advantages of pre-training.

The generative Adversarial Network (GAN) framework presents a promising methodology for predicting contrast-enhanced T1-weighted images. GANs, characterised by their dual architecture of a generator and discriminator working in tandem, excel at synthesising data that mirrors accurate distributions. These GAN models have already been used in studies by (Wang et al., 2022) and (Jayachandran Preetha et al., 2021) with noticeable results. So, instead of using CNNs, we can use CGAN architecture to get better results.

5.2 Future Works

In subsequent research endeavours, the main aim is to develop a novel deep-learning architecture that can be trained on this huge multimodal UCSF-PDGM MRI dataset of glioma patients. The current research analysis will help us classify the necessary MRI modalities by shedding light on their individual and collective efficacies for generating synthetic post-contrast MR images to reduce the usage of gadolinium-based contrast agents. To achieve this feat, multiple unexplored ideas exist for future extension to our current research. The first one is instead of using a two-dimensional input image-based model; we can surely enhance the performance by converting it into a three-dimensional model. This will be computationally more expensive, but it would give us better results, like in the study by Gong et al. We can also incorporate image augmentation techniques into our machine learning pipeline of data pre-processing before feeding it into the neural network. As an extension of current research, another general idea is to integrate Radiomics to extract clinically relevant quantitative features from MR images. This technique could aid deep learning architecture in learning the data more precisely. Xie et al. conducted a similar study using a retina U-Net to extract semantic features from a dataset. Then, they combined the information into a synthetic model to generate synthesised CE MR images.

6 Chapter 6: Conclusions

The brain plays a crucial role in our cognitive abilities. It is essential to detect brain tumours early to prevent potential harm. While Magnetic Resonance Imaging (MRI) has effectively seen brain anomalies, there are concerns about the contrast-enhancing agents that can lead to gadolinium deposition. Although rare, this can pose health risks, leading to a search for alternative methods. Artificial Intelligence (AI) and intense learning methodologies have emerged as viable options to reduce the need for contrast agents. In this study, we analysed published works that have developed deep-learning architectures capable of generating synthetic contrast-enhanced MR images from non-contrast images. Our analysis also aimed to identify contrast-enhanced tumours and compare various MRI modalities to determine which ones were more informative. We combined modalities to increase diversity in our model's learning curve and determine if data from multiple modalities improved the model's accuracy. We performed the same experiments to generate synthesised CE-T1 MR images and obtained similar insights.

Our extensive research has found that the UCSF-PDGM dataset boasts an array of MRI modalities with remarkable potential in identifying CE tumours. ASL, T1, and T2 MRI sequences stand out in particular among these modalities. As for generating CE-T1 MR images, we have observed that T1 and T2 sequences have yielded the most promising outcomes. Despite these findings, we have also determined that certain adjustments in the deep learning architecture need to be implemented in our 2D U-Net model to enhance its performance in this domain.

7 Code List

For this project in Python, code was written using Jupyter Notebook. To ensure the best performance, University College London's online Linux system was configured to provide the server's GPU access. To make the notebooks easily accessible, they have been uploaded to a public repository on GitHub. If you want to view this project's code, please click the link: https://github.com/amritansh22jain/UCL_MastersThesis.

The workbooks are numbered in sequence from Workbook 1 to Workbook 8. Each notebook is designed to identify CE tumours for a specific MRI modality. Workbook 9 to Workbook 15 predicts contrast enhanced T1 MR images from pre-contrast MRI modality. Workbook 16 to Workbook 19 combines MRI modalities to perform both tasks.

8 References

Aamir, M., Rahman, Z., Dayo, Z. A., Abro, W. A., Uddin, M. I., Khan, I., Imran, A. S., Ali, Z., Ishfaq, M., Guan, Y., & Hu, Z. (2022). A deep learning approach for brain tumor classification using MRI images. Computers and Electrical Engineering, 101, 108105. <https://doi.org/10.1016/j.compeleceng.2022.108105>

Amber Roberts. January 1, 2023. Binary Cross Entropy: Where To Use Log Loss In Model Monitoring: <https://arize.com/blog-course/binary-cross-entropy-log-loss/> [Accessed on: 23/09/2023]

Baliyan, V., Das, C. J., Sharma, R., & Gupta, A. K. (2016). Diffusion-weighted imaging: Technique and applications. World Journal of Radiology, 8(9), 785. <https://doi.org/10.4329/wjr.v8.i9.785>

Bardin, T., & Richette, P. (2010). Nephrogenic systemic fibrosis. Current Opinion in Rheumatology, 22(1), 54–58. <https://doi.org/10.1097/BOR.0b013e328333bf3d>

BMJ Best Practice: <https://bestpractice.bmj.com/topics/en-gb/729>

Calabrese, E., Villanueva-Meyer, J. E., Rudie, J. D., Rauschecker, A. M., Baid, U., Bakas, S., Cha, S., Mongan, J. T., & Hess, C. P. (2022). The University of California San Francisco Preoperative Diffuse Glioma MRI Dataset. Radiology. Artificial Intelligence, 4(6), e220058. <https://doi.org/10.1148/ryai.220058>

Carol DerSarkissian, MD on November 19, 2022, WebMD – Brain Cancer and Gliomas: <https://www.webmd.com/cancer/brain-cancer/malignant-gliomas>

Chen, C., Raymond, C., Speier, W., Jin, X., Cloughesy, T. F., Enzmann, D., Ellingson, B. M., & Arnold, C. W. (2023). Synthesising MR Image Contrast Enhancement Using 3D High-Resolution ConvNets. IEEE Transactions on Biomedical Engineering, 70(2), 401–412. <https://doi.org/10.1109/TBME.2022.3192309>

Cowper, S. E., Robin, H. S., Steinberg, S. M., Su, L. D., Gupta, S., & LeBoit, P. E. (2000). Scleromyxoedema-like cutaneous diseases in renal-dialysis patients. The Lancet, 356(9234), 1000–1001. [https://doi.org/10.1016/S0140-6736\(00\)02694-5](https://doi.org/10.1016/S0140-6736(00)02694-5)

Daniel Godoy, 2018. Binary Cross Entropy Loss function: <https://towardsdatascience.com/understanding-binary-cross-entropy-log-loss-a-visual-explanation-a3ac6025181a>

Davenport, M. S. (2018). Choosing the Safest Gadolinium-based Contrast Medium for MR Imaging: Not So Simple after All. Radiology, 286(2), 483–485. <https://doi.org/10.1148/radiol.2017172224>

Deval Shah. May 30, 2023. <https://www.v7labs.com/blog/intersection-over-union-guide#:~:text=Calculate%20IoU%20score%3A%20Compute%20the,to%20determine%20the%20IoU%20score>

Envision Radiology: What is an MRI with Contrast? <https://www.envrad.com/what-is-an-mri-with-contrast/>

European Medicines Agency. Questions and answers on the review of gadolinium-containing contrast agents. http://www.ema.europa.eu/docs/en_GB/document_library/Referrals_document/gadolinium_31/WC500015635.pdf

FDA Requests Boxed Warning for Contrast Agents Used to Improve MRI Images: <https://www.itnonline.com/content/fda-requests-boxed-warning-contrast-agents-used-improve-mri-images>

Ferlay, J., Colombet, M., Soerjomataram, I., Parkin, D. M., Piñeros, M., Znaor, A., & Bray, F. (2021). Cancer statistics for the year 2020: An overview. International Journal of Cancer, 149(4), 778–789. <https://doi.org/10.1002/IJC.33588>

Gong, E., Pauly, J. M., Wintermark, M., & Zaharchuk, G. (2018). Deep learning enables reduced gadolinium dose for contrast-enhanced brain MRI. Journal of Magnetic Resonance Imaging, 48(2), 330–340. <https://doi.org/10.1002/jmri.25970>

Grobner, T. (2006). Gadolinium – a specific trigger for the development of nephrogenic fibrosing dermopathy and nephrogenic systemic fibrosis? Nephrology Dialysis Transplantation, 21(4), 1104–1108. <https://doi.org/10.1093/ndt/gfk062>

Guo, B. J., Yang, Z. L., & Zhang, L. J. (2018). Gadolinium Deposition in Brain: Current Scientific Evidence and Future Perspectives. Frontiers in Molecular Neuroscience, 11. <https://doi.org/10.3389/fnmol.2018.00335>

Halefoglu, A. M., & Yousem, D. M. (2018). Susceptibility weighted imaging: Clinical applications and future directions. World Journal of Radiology, 10(4), 30–45. <https://doi.org/10.4329/wjr.v10.i4.30>

Ibrahim MA, Hazhirkarzar B, Dublin AB. Gadolinium Magnetic Resonance Imaging. [Updated 2023 Jul 3]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2023 Jan-. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK482487/>

Jacobs, M. A., Ouwerkerk, R., Petrowski, K., & Macura, K. J. (2008). Diffusion-Weighted Imaging with Apparent Diffusion Coefficient Mapping and Spectroscopy in Prostate Cancer. Topics in Magnetic Resonance Imaging, 19(6), 261–272. <https://doi.org/10.1097/RMR.0b013e3181aa6b50>

Jason Brownlee. February 24, 2020. A Gentle Introduction to the Fbeta-Measure for Machine Learning: <https://machinelearningmastery.com/fbeta-measure-for-machine-learning/>

Jayachandran Preetha, C., Meredig, H., Brugnara, G., Mahmutoglu, M. A., Foltyn, M., Isensee, F., Kessler, T., Pflüger, I., Schell, M., Neuberger, U., Petersen, J., Wick, A., Heiland, S., Debus, J., Platten, M., Idbaih, A., Brandes, A. A., Winkler, F., van den Bent, M. J., ... Vollmuth, P. (2021). Deep-learning-based synthesis of post-contrast T1-weighted MRI for tumour response assessment in neuro-oncology: a multicenter, retrospective cohort study. *The Lancet Digital Health*, 3(12), e784–e794. [https://doi.org/10.1016/S2589-7500\(21\)00205-3](https://doi.org/10.1016/S2589-7500(21)00205-3)

Kanal, E., Maravilla, K., & Rowley, H. A. (2014). Gadolinium Contrast Agents for CNS Imaging: Current Concepts and Clinical Evidence. *American Journal of Neuroradiology*, 35(12), 2215–2226. <https://doi.org/10.3174/ajnr.A3917>

Kanda, T., Ishii, K., Kawaguchi, H., Kitajima, K., & Takenaka, D. (2014). High Signal Intensity in the Dentate Nucleus and Globus Pallidus on Unenhanced T1-weighted MR Images: Relationship with Increasing Cumulative Dose of a Gadolinium-based Contrast Material. *Radiology*, 270(3), 834–841. <https://doi.org/10.1148/radiol.13131669>

Kiviniemi, A., Gardberg, M., Ek, P., Frantzén, J., Bobacka, J., & Minn, H. (2019). Gadolinium retention in gliomas and adjacent normal brain tissue: association with tumor contrast enhancement and linear/macrocyclic agents. *Neuroradiology*, 61(5), 535–544. <https://doi.org/10.1007/s00234-019-02172-6>

Kleesiek, J., Morshuis, J. N., Isensee, F., Deike-Hofmann, K., Paech, D., Kickingereder, P., Köthe, U., Rother, C., Forsting, M., Wick, W., Bendszus, M., Schlemmer, H.-P., & Radbruch, A. (2019). Can Virtual Contrast Enhancement in Brain MRI Replace Gadolinium? *Investigative Radiology*, 54(10), 653–660. <https://doi.org/10.1097/RLI.0000000000000583>

Lohrke, J., Frenzel, T., Endrikat, J., Alves, F. C., Grist, T. M., Law, M., Lee, J. M., Leiner, T., Li, K.-C., Nikolaou, K., Prince, M. R., Schild, H. H., Weinreb, J. C., Yoshikawa, K., & Pietsch, H. (2016). 25 Years of Contrast-Enhanced MRI: Developments, Current Challenges and Future Perspectives. *Advances in Therapy*, 33(1), 1–28. <https://doi.org/10.1007/s12325-015-0275-4>

M. Gab Allah, A., M. Sarhan, A., & M. Elshennawy, N. (2023). Edge U-Net: Brain tumor segmentation using MRI based on deep U-Net model with boundary information. *Expert Systems with Applications*, 213, 118833. <https://doi.org/10.1016/j.eswa.2022.118833>

Magnetic Resonance Imaging (MRI) of the Brain and Spine: Basics - <https://case.edu/med/neurology/NR/MRI%20Basics.htm>

McKinney, P. A. (2004). Brain tumours: incidence, survival, and aetiology. *Journal of Neurology, Neurosurgery, and Psychiatry*, 75 Suppl 2(Suppl 2), ii12-7. <https://doi.org/10.1136/jnnp.2004.040741>

Mohammed, F. M., Essa, M. M., & Maseer, A. W. (2019). Comparison between MRI And CT-Scan In Diagnosis The Brain Tumor Images. *International Journal of Medical Science*, 6(5), 1–4. <https://doi.org/10.14445/23939117/IJMS-V6I5P101>

Moya-Sáez, E., de Luis-García, R., & Alberola-López, C. (2023). Toward deep learning replacement of gadolinium in neuro-oncology: A review of contrast-enhanced synthetic MRI. *Frontiers in Neuroimaging*, 2. <https://doi.org/10.3389/fnimg.2023.1055463>

Murata, N., Gonzalez-Cuyar, L. F., Murata, K., Fligner, C., Dills, R., Hippe, D., & Maravilla, K. R. (2016). Macrocyclic and Other non-group 1 Gadolinium Contrast Agents Deposit Low Levels of Gadolinium in Brain and Bone Tissue. *Investigative Radiology*, 51(7), 447–453. <https://doi.org/10.1097/RLI.0000000000000252>

Nadipally, M. (2019). Optimization of Methods for Image-Texture Segmentation Using Ant Colony Optimization. In *Intelligent Data Analysis for Biomedical Applications* (pp. 21–47). Elsevier. <https://doi.org/10.1016/B978-0-12-815553-0.00002-1>

O'Donnell, L. J., & Westin, C.-F. (2011). An Introduction to Diffusion Tensor Image Analysis. *Neurosurgery Clinics of North America*, 22(2), 185–196. <https://doi.org/10.1016/j.nec.2010.12.004>

Petcharunpaisan, S. (2010). Arterial spin labelling in neuroimaging. *World Journal of Radiology*, 2(10), 384. <https://doi.org/10.4329/wjr.v2.i10.384>

Pullicino, R., & Das, K. (2017). Is it Safe to Use Gadolinium-Based Contrast Agents in MRI? *Journal of the Royal College of Physicians of Edinburgh*, 47(3), 243–246. <https://doi.org/10.4997/jrcpe.2017.306>

Radbruch, A., Haase, R., Kieslich, P. J., Weberling, L. D., Kickingereder, P., Wick, W., Schlemmer, H.-P., & Bendszus, M. (2017). No Signal Intensity Increase in the Dentate Nucleus on Unenhanced T1-weighted MR Images after More than 20 Serial Injections of Macroyclic Gadolinium-based Contrast Agents. *Radiology*, 282(3), 699–707. <https://doi.org/10.1148/radiol.2016162241>

Rahman, T., & Islam, M. S. (2023). MRI brain tumor detection and classification using parallel deep convolutional neural networks. *Measurement: Sensors*, 26, 100694. <https://doi.org/10.1016/j.measen.2023.100694>

Ranjbarzadeh, R., Caputo, A., Tirkolaee, E. B., Jafarzadeh Ghoushchi, S., & Bendechache, M. (2023). Brain tumor segmentation of MRI images: A comprehensive review on the application of artificial intelligence tools. *Computers in Biology and Medicine*, 152, 106405. <https://doi.org/10.1016/j.combiomed.2022.106405>

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. Subramanian S, Ahmad T. Childhood Brain Tumours. [Updated 2023 Aug 8]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2023 Jan-. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK535415/>

Shipra Saxena. March 3, 2021. Binary Cross Entropy/Log Loss for Binary Classification: <https://www.analyticsvidhya.com/blog/2021/03/binary-cross-entropy-log-loss-for-binary-classification> [Last Modified on: 13/09/2023, Accessed on: 23/09/2023]

Tracy Nolan: The Cancer Imaging Archive. 2023 Jul 14. UCSF-PDGM Dataset available at: <https://doi.org/10.7937/tcia.bdgf-8v37>

UK Biobank Brain Imaging Documentation - <http://www.ukbiobank.ac.uk>

Vankdothu, R., & Hameed, M. A. (2022). Brain tumor MRI image identification and classification based on the recurrent convolutional neural network. *Measurement: Sensors*, 24, 100412. <https://doi.org/10.1016/j.measen.2022.100412>

Vitaly Bushaev. October 22, 2018. Adam – latest trends in deep learning optimisation: <https://towardsdatascience.com/adam-latest-trends-in-deep-learning-optimization-6be9a291375c>

Xie, H., Lei, Y., Wang, T., Roper, J., Axente, M., Bradley, J. D., Liu, T., & Yang, X. (2022). Magnetic resonance imaging contrast enhancement synthesis using cascade networks with local supervision. *Medical Physics*, 49(5), 3278–3287. <https://doi.org/10.1002/mp.15578>

Zou, K. H., Warfield, S. K., Bharatha, A., Tempany, C. M. C., Kaus, M. R., Haker, S. J., Wells, W. M., Jolesz, F. A., & Kikinis, R. (2004). Statistical validation of image segmentation quality based on a spatial overlap index1. *Academic Radiology*, 11(2), 178–189. [https://doi.org/10.1016/S1076-6332\(03\)00671-8](https://doi.org/10.1016/S1076-6332(03)00671-8)

Wang, Y., Wu, W., Yang, Y., Hu, H., Yu, S., Dong, X., et al. (2022). Deep learning-based 3D MRI contrast-enhanced synthesis from a 2D noncontrast T2Flair sequence. *Med. Phys.* 49, 4478–4493. doi: 10.1002/mp.15636