Grenze International Journal of Engineering and Technology, January Issue

Inter-modality MRI Image Translation using Deep Learning

Prachee Patel[1], Dr. Pariza Kamboj[2] and Dr. Nirali Nanavati[3]
[1-3]Sarvajanik College of Engineering and Technology, Surat, India
Email: pracheepatel@hostname1.org, parizakamboj@hostname2.org, niralinanavati@hostname2.org

***Abstract*— Magnetic resonance imaging (MRI) is one of the most predominantly used Medical Imaging in neurology and neurosurgery. In medical image translation, one medical image modality is translated into another medical image modality. Image translation and deep learning techniques in the medical domain reduce workload and total acquisition timing by removing the need to take some imagining sequences such are those taken in MRI called T1WI, T2WI, and Flair (most common sequence). Acquisition timing for Flair image modality is much longer than T1WI and T2WI. The objective of this work is to minimize the total acquisition time of MRIs by doing image translation on different MRI modalities called T1WI, T2WI, and Flair with the help of deep learning models. The proposed U-net Based model generates Flair image modality from T1WI and T2WI from the Deep Learning model in standard medical image format Digital Imaging and Communications in Medicine (DICOM) which significantly reduces the total acquisition timing of MRIs and thus reduces motion artefacts.**

***Index Terms*— Medical image translation; Inter-modality MRI translation; Deep learning; U-net model.**

I. INTRODUCTION

Medical imaging has become an indispensable cornerstone of modern healthcare, playing a critical role in the diagnosis, monitoring, and treatment planning of a wide range of medical conditions. Various imaging modalities, including X-rays, computed tomography (CT), ultrasound, and magnetic resonance imaging (MRI), provide unique advantages tailored to specific clinical scenarios. Among these, MRI stands out as a highly versatile and informative tool for brain imaging due to its superior soft tissue contrast, high spatial resolution, and ability to capture detailed anatomical and functional information [1–3]. Its non-invasive nature allows clinicians to examine internal structures without the risks associated with surgical procedures, which is particularly important for neurological assessments where invasive interventions are often impractical or unsafe [2].

When comparing MRI to other imaging techniques for the brain, several key benefits are evident. MRI provides unparalleled soft tissue contrast, enabling precise differentiation of gray matter, white matter, cerebrospinal fluid (CSF), and pathological lesions. This capability is essential for the detection and characterization of neurological conditions, including tumors, multiple sclerosis, stroke, and vascular abnormalities [4]. Additionally, MRI allows for the acquisition of multiple imaging sequences, each highlighting different tissue properties, which collectively support comprehensive diagnosis and treatment planning.

Grenze ID: 01.GIJET.12.1.7_1

Among the MRI sequences commonly used for brain imaging, T1-weighted (T1WI), T2-weighted (T2WI), and Fluid-Attenuated Inversion Recovery (FLAIR) images are particularly significant. T1WI provides excellent anatomical detail and is useful for delineating gray and white matter structures, while T2WI highlights pathological changes such as edema and tumors. FLAIR sequences suppress the signal from CSF, thereby improving the visualization of lesions near fluid spaces, such as periventricular or cortical abnormalities [5]. Figure 1 illustrates examples of these three modalities, demonstrating how each provides a complementary perspective on brain tissue characteristics. Considering all modalities during analysis and diagnosis enables a more robust and accurate clinical interpretation.

Inter-modality MRI image translation has emerged as a powerful approach for combining the strengths of different MRI sequences. This technique involves converting MRI data from one modality into another, thereby enabling the integration of complementary information from multiple sequences. By facilitating such integration, inter-modality translation enhances the diagnostic utility of MRI and supports the development of personalized treatment strategies. The clinical applications are broad, ranging from improving image quality—by generating higher-resolution scans from low-resolution acquisitions—to synthesizing functional MRI (fMRI) data from structural scans, or even translating MRI information into other imaging modalities such as CT or PET. This capability effectively bridges gaps between different imaging methods, enabling novel diagnostic and research opportunities that were previously unattainable.

Recent advances in deep learning have propelled the field of medical image translation forward, enabling the development of models that can learn complex mappings between different imaging modalities. Techniques based on Convolutional Neural Networks (CNNs) and Generative Adversarial Networks (GANs) have been successfully applied to transform one type of medical image into another while preserving critical anatomical and clinical information [6]. Deep learning–based image translation offers substantial benefits in terms of diagnostic workflow efficiency, patient comfort, and overall healthcare delivery. For instance, MRI sequences often differ in acquisition time due to variations in pulse sequences and imaging parameters. FLAIR imaging, while clinically valuable, requires substantially longer scan times compared to T1WI or T2WI, which increases the likelihood of patient discomfort, motion artifacts, and higher operational costs [7,8]. By enabling the synthesis of FLAIR images from faster-acquired sequences, image-to-image translation can reduce scan times, minimize patient burden, and improve the safety and efficiency of radiological procedures.

Among the architectures employed for image translation tasks, the U-Net model has garnered significant attention for its ability to capture fine-grained details while preserving global spatial relationships [9,10]. Its characteristic "U"-shaped structure, featuring contracting and expanding pathways, allows it to extract hierarchical features and reconstruct high-fidelity images. The U-Net architecture has been successfully adapted to medical image translation, demonstrating superior performance in generating realistic and clinically useful outputs.

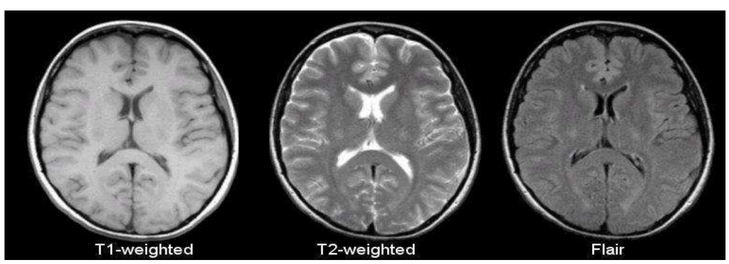

Fig. 1. Comparison between T1WI, T2WI, and Flair image modality [6].

In this study, we present a novel application of the U-Net architecture tailored specifically for inter-modality MRI translation, focusing on the generation of FLAIR images from T1WI and T2WI sequences. This approach leverages the U-Net's feature extraction and reconstruction capabilities to produce high-quality synthesized images, enhancing diagnostic precision and supporting radiologists in clinical decision-making. By generating FLAIR images from faster-acquired sequences, our method addresses the practical limitations of prolonged scan times while preserving critical anatomical structures and minimizing artifacts.

Our proposed methodology represents a significant step forward in tackling the challenges of inter-modality MRI translation. The U-Net model's capacity to extract, preserve, and reconstruct features ensures that the synthesized FLAIR images faithfully represent underlying anatomical structures, thereby improving clinical interpretability. In addition to enhancing the accuracy of diagnoses, this approach has the potential to streamline radiological workflows, reduce patient exposure to lengthy scanning procedures, and facilitate more personalized and effective treatment planning. By integrating deep learning–based image translation into clinical practice, this work contributes to advancing both the technological capabilities and practical applications of MRI in modern healthcare.

FLAIR imaging is a critical MRI technique, valued for its ability to suppress cerebrospinal fluid (CSF) and enhance the visualization of pathological brain structures [8]. Despite its clinical significance, most existing research in inter-modality MRI image translation has largely concentrated on converting between T1-weighted (T1WI) and T2-weighted (T2WI) images.

In this study, we present a novel approach that extends the scope of inter-modality MRI translation to include the generation of FLAIR images. Our methodology employs two distinct strategies: first, using T1WI images as input to deep learning models, and second, using T2WI images as input. This dual-input framework represents a departure from the conventional focus on T1WI–T2WI translation, thereby broadening the applicability and clinical relevance of inter-modality image synthesis.

A key aspect of our contribution lies in generating FLAIR images in the DICOM format, ensuring compatibility with established medical imaging systems and workflows. This design choice facilitates seamless integration into clinical practice, highlighting the practical utility of our approach. By delivering FLAIR images in a widely accepted standard, our work not only aligns with the operational needs of healthcare institutions and radiology departments but also supports improved diagnostic accuracy and patient care.

II. Related Work

Image translation entails the complex task of uncovering the underlying patterns and relationships between input and output images. By learning these intricate mapping functions, researchers aim to bridge gaps between different visual modalities, thereby expanding the utility and interpretability of digital imagery. Applications include converting night-time to daytime photographs, seasonal transformations (e.g., winter to summer), and mapping satellite imagery into cartographic representations [11]. In the medical domain, deep learning–based image translation holds significant potential to reduce annotation efforts and imaging workload by obviating the need to acquire certain MRI sequences, such as T1-weighted (T1WI), T2-weighted (T2WI), and FLAIR scans.

A wide range of algorithms has been proposed for medical image translation, encompassing both traditional and deep learning–based methods. Classical approaches often employ mathematical algorithms and signal processing techniques to convert images from one modality to another. One well-established technique is the atlas-based method [12], which relies on pairs of registered images sampled at the same spatial voxel locations but with differing tissue contrasts. To generate high-resolution brain MR images, methods such as Markov random fields [13] are employed to select nearest neighbours with similar characteristics from low-resolution images.

Despite their historical relevance, classical methods exhibit notable limitations. They depend heavily on labour-intensive and error-prone manual image registration, struggle to capture complex inter-modality relationships and subtle anatomical features, and are constrained by reliance on handcrafted features. Consequently, their adaptability across diverse datasets and clinical contexts is limited, and they fail to leverage the capabilities of modern computational approaches.

In contrast, deep learning approaches—notably variants of Convolutional Neural Networks (CNNs) and Generative Adversarial Networks (GANs)—have demonstrated superior performance in medical image translation. These models automatically learn and capture intricate features and complex mappings between different MRI modalities, enabling more accurate and efficient image synthesis. The flexibility and scalability of deep learning models underscore their transformative potential for healthcare applications, including improved diagnostic precision and reduced imaging burden.

The advantage of deep learning models lies in their ability to handle the inherent complexity of medical images, including variations in anatomy and image quality. One notable deep learning-based method [14], utilizes a CNN architecture. Hien Van Nguyen, Kevin Zhou et have described a location-sensitive deep network (LSDN) intended to methodically integrate spatial location and image intensity data in a well-founded manner for cross-modality generation. This CNN is designed to learn the intricate relationships between source and target modalities. By training on a large dataset of paired images from different modalities, the model can effectively generate high-quality translations.

[15–17] [8] [18] employ GAN for MRI image translation. A GAN comprises a discriminator network and a generator network trained in an adversarial fashion. The discriminator's role is to distinguish between authentic and synthesized images, while the generator strives to produce target modality images from source modality images. Through this adversarial training process, the generator is incentivized to create highly authentic outputs.

In recent years advancements in deep learning have seen the emergence of architectures like U-Net [19], which is particularly well-suited for medical image translation tasks. U-Net and its variants [20] employ a U-shaped architecture with skip connections to preserve spatial information and generate accurate translations.

[21] employed a U-Net-based architecture to address the task of translating medical images between different MRI modalities like T1WI and T2WI, specifically focusing on brain images. This research marks the first successful attempt at performing medical image translation for brain MRI images in the widely recognized DICOM format. An advanced U-Net model with deep convolutional layers was developed in [22]. This model was designed to carry out image-to-image translation tasks, specifically across T1WI, T2WI, and FLAIR MRI sequences for brain images in Nifti format.

The proposed neural network [8] comprises a domain adaptation module (DAM) and a reconstruction module (RM) for T2WI reconstruction from T1WI. It incorporates multi-channel input, RM regularization, sharp bottle-neck modules (SBM) for feature extraction, and modular training to enhance image quality, leveraging prior information and optimizing the learning process efficiently. This holistic approach improves transformation learning in medical image processing, ensuring robustness and accuracy in T2WI reconstruction.

An innovative generative technique called generative adversarial U-net, which combines generative adversarial network and U-net, is suggested in [23] to generate diverse medical images. For translation from CT to MR and MR to CT, modified pix2pix models are developed in [24]. These models are trained with unpaired CT and MR data as well as MR-CAT paired data, which is produced from the MR scans. The suggested pix2pix variations provide more realistic-looking CT and MR translations and outperform baseline pix2pix, pix2pixHD, and CycleGAN in terms of frechet inception distance (FID) and kernel inception distance (KID) score [24].

In [24], a Dense-U-Net model is utilized to reconstruct T2WI from both provided T1WI and under-sampled k-space T2WI data. Their model outperforms the conventional U-Net model, as evidenced by higher peak signal-to-noise ratio (PSNR) values. The higher the PSNR, the better the quality of the compressed, or reconstructed image. This research underscores the importance of leveraging complementary k-space information across different imaging modalities. Significantly, it introduces a novel approach by incorporating a limited set of T2WI k-space samples alongside complete T1WI data for T2WI reconstruction, utilizing the DenseNet architecture. In [25], Lei Xiang Chen, Yun Li at el harnessed conditional generative adversarial networks (cGAN) to achieve precise contrast synthesis. These cGANs underwent assessment in the context of tasks such as cross-modality registration, MRI segmentation, and the conversion of MRI scans across diverse imaging modalities.

Deep learning for inter-modality medical image translation has also explored the use of Siamese Autoencoders to improve the explainability of generated images. For instance, in [26], a framework was proposed that leverages a shared latent space to align features between two imaging modalities. Unlike conventional black-box models, the Siamese Autoencoder provides interpretable representations, enabling clinicians to understand which anatomical and textural features are being transferred across modalities. This emphasis on explainability improves trust and clinical adoption, particularly when translating between diverse modalities such as MRI, CT, and PET.

Recent work in multi-modal MR-to-CT translation have focused on reducing patient radiation exposure while exploiting the complementary strengths of different MRI sequences. Gong et at el [27], proposed a novel Mamba-based architecture that fuses T1- and T2-weighted MRI inputs to generate high-quality CT images. Unlike traditional CNN-based methods with limited receptive fields or Transformer-based models with high computational demands, the Mamba state space model achieves an effective trade-off between long-range dependency modelling and computational efficiency. To further enhance translation quality, MM2CT integrates a dynamic local convolution module and a dynamic enhancement mechanism, which significantly improve both anatomical fidelity and textural consistency. Evaluation on a pelvic dataset demonstrated that MM2CT consistently outperformed state-of-the-art baselines, including GAN-based approaches (CycleGAN, MaskGAN) and diffusion-based models (Syndiff), achieving superior PSNR and SSIM values. These findings underscore the potential of multi-modal fusion with Mamba for advancing cross-modality medical image synthesis.

Diffusion models have recently gained attention for their ability to produce high-fidelity medical images, yet their clinical utility is limited by slow inference. To address this limitation, Hongxu Jiang et al. [28],introduced an accelerated diffusion framework that reduces the number of sampling steps required for image synthesis. By employing a progressive noise scheduling scheme and an optimized sampling strategy, Fast-DDPM achieves 10–20× faster inference while maintaining competitive image quality in terms of PSNR, SSIM, and FID. The approach has been successfully applied to MRI-to-CT translation and low-dose CT enhancement, demonstrating its effectiveness in preserving both structural integrity and fine textural details. This line of research highlights how accelerating generative models is critical for making diffusion-based medical image translation methods more practical in real-time clinical workflows.

U-Net and CycleGAN were both used to the dataset having T2WI of MRI and CT scan images. Unsupervised deep learning was accomplished using CycleGAN, whereas supervised deep learning was accomplished using U-Net. When compared to the unsupervised approach, the supervised deep learning method generated MR/CT synthesis with higher accuracy [22].

In medical image-to-image translation research, NIfTI images have garnered significant attention [5][8] [17-18] [21] [26-28]. NIfTI's flexibility in accommodating diverse neuroimaging data types has made it a prime choice for experimentation and algorithm development. On the other hand, within the clinical domain of MRI facilities, DICOM [29] images reign supreme. These images are the standard due to their widespread adoption and compatibility with various medical imaging equipment. DICOM ensures seamless data interchange and is tailored to meet the stringent requirements of healthcare settings.

Additionally, NIfTI serves research exploration, while DICOM caters to the practical demands of everyday clinical MRI operations. Bridging these formats is crucial for translating research advancements into real-world medical practice. The majority of research efforts in medical image translation have concentrated on transforming MRI into CT scans. Conversely, there has been a noticeable scarcity of work pertaining to inter-modality translation for brain MRIs, encompassing T1WI, T2WI, and FLAIR modalities. Furthermore, studies have predominantly focused on converting T1WI to T2WI, despite the substantial clinical importance of incorporating Flair imaging, leaving this crucial aspect of MRI translation relatively unexplored.

III. Proposed Solution

A. Motivation

MRI is not painful, but the patient must remain still in an enclosed machine, which may be a problem for claustrophobic patients, child patients and old age patients. The magnetic fields that change with time create loud knocking noises which may be harmful for ears. The radiofrequency energy used during the MRI scan could lead to heating of the body [30]. The potential for heating is greater during long MRI examinations. An MRI scan can harm implanted cardiac devices, such as pacemakers and defibrillators. Some patients, such as those who are totally reliant on their pacemaker, may be at danger since the strong magnets might trigger disruptions in a pacemaker's settings [31].

Compared to T1WI, T2WI and Flair images take much longer to obtain, extending the acquisition process and increasing the risk of motion artifacts while also raising the cost. The diagnosis procedure can be sped up and made safer if we are able to transform a medical image from one modality to another. The majority of image translation effort is restricted to T1WI and T2WI, despite Flair's significance in MRI scanning, according to a literature review. We are proposing to generate Flair image modality in two ways, first using T1WI and second using T2WI as input in deep learning model. The production of Fair image modalities in DICOM format is something we are the first to suggest.

B. Methodology

Real-time data will be used for our suggested solution. Pre-processing and cleaning will be necessary in order to get the genuine data from dispersed resources and various formats. Nevertheless, it is essential to give raw data to the Deep Learning model in order to maintain its true characteristics. During the pre-processing stage, we have changed image resolution of images to make the same aspect ratio for paired data and label, sort images to make pairs of data and label, resize the images and normalise the pixel data value of the images. After that in the training phase Dataset was divided into training, validation and testing dataset. Deep learning models were trained to generate flair images from a given pair of data and labels. Finally, the analysis of the results is done.

By comparing the model-generated flair image to the actual scene, we want to evaluate the model. The analysis for a better method of producing flair images will be completed, and the results will be provided. Selecting an inaccurate metric will result in poor training since loss metrics show how inaccurate the model's prediction is during training. Loss metrics for image translation must evaluate three key elements: brightness, contrast, and structure. Luminance in a medical image refers to the brightness of the pixels, contrast to the detail, and structure to the strength of the various textures. Together, these characteristics describe the pixel values of tissue and organs in a medical picture.

For evaluation of model PSNR (Peak Signal to Noise Ratio), MSE (Mean Squared Error) and MAE (Mean Absolute Error) will be calculated for generated Flair images against ground truth. Figure below provides the formula for the three applied loss measures. During training, certain measurements might be given more weight or focus than others by combining these indicators in a proportionate way. In order to improve performance, the model must learn, and this can be conceptually determined [18].

C. Deep Learning Model

By exploring various architecture of CNN at the basic level we are planning to use U-net architecture for our image-to-image translation task. U-net is developed for image segmentation in the biomedical domain. Image segmentation is also image translation from image to binary image in case of black & white segmentation. It is pixel-to-pixel image translation [5].

Working of U-net

The U-Net architecture is a type of artificial neural network that employs an autoencoder structure enhanced with skip connections, designed to efficiently capture both global context and fine-grained spatial features. The network is divided into two primary components: an encoder, responsible for feature extraction, and a decoder, tasked with reconstructing the target output, such as a segmentation map, from the learned features.

The encoder resembles the architecture of fully convolutional networks (FCNs) and consists of a series of convolutional blocks interleaved with down-sampling operations. Each block typically includes multiple convolutional layers followed by a max-pooling layer to progressively reduce the spatial dimensions while increasing the depth of the feature maps. After each down-sampling step, the number of convolutional filters is doubled, allowing the network to capture increasingly complex hierarchical features. By the end of the encoder, the input image is transformed into a compact, high-dimensional feature map that encodes essential spatial and contextual information.

The decoder is designed to perform up-sampling and reconstruct the target output from the encoded features. It begins with a deconvolutional (transposed convolution) layer that up-samples the feature map while halving the number of filters in the output. This is followed by a series of up-sampling blocks, each comprising two convolutional layers and a deconvolutional layer, which gradually restore the spatial dimensions of the feature maps. The final convolutional layer produces the output, applying a Sigmoid activation function for tasks such as binary segmentation, while all intermediate layers utilize ReLU activation to introduce non-linearity and facilitate effective feature learning.

A key innovation of the U-Net is the use of skip connections, which link corresponding encoder and decoder blocks. Specifically, the feature maps generated in the encoder prior to max-pooling are concatenated with the output of the corresponding decoder block. This operation allows the decoder to recover spatial information that may have been lost during down-sampling, enhancing the network's ability to preserve fine-grained details. By combining hierarchical features from the encoder with the up-sampled representations in the decoder, U-Net effectively balances global contextual understanding with precise local reconstruction, making it highly effective for medical image segmentation and translation tasks [8].

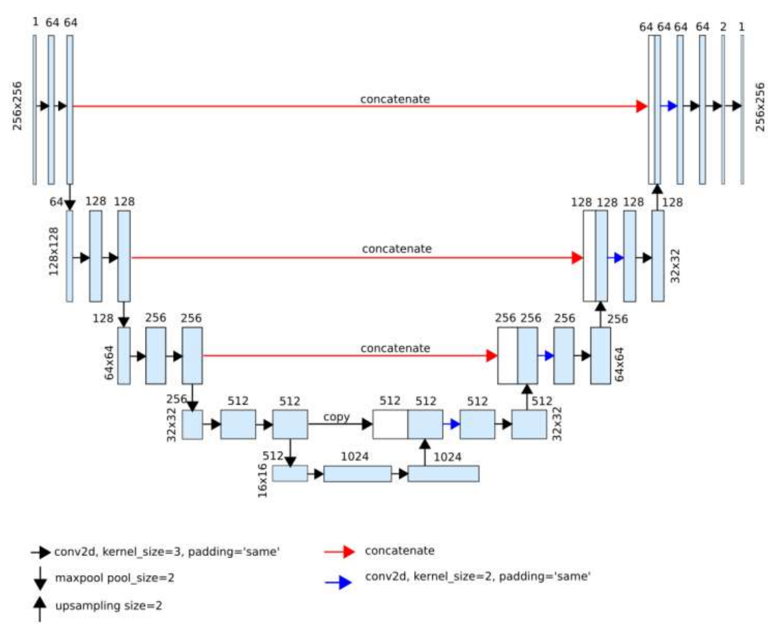

Fig. 2. U-net Architecture: Biomedical Image Segmentation (CNN)

The U-Net architecture is a popular deep learning model for image segmentation tasks, where the goal is to classify each pixel in an image into a specific class. Image translation, on the other hand, is the task of transforming an input image into an output image that has a different appearance but preserves its semantic content.

U-net model architecture used for inter-modality MRI image translation task

The U-Net architecture is a widely adopted deep learning model, originally designed for image segmentation tasks, where the objective is to assign a class label to each pixel in an image. Its distinctive “U” shaped structure, comprising contracting (encoder) and expanding (decoder) paths, enables the capture of both local and global contextual information, making it highly effective for tasks requiring precise localization.

While U-Net was initially intended for segmentation, its architecture has proven versatile and has been successfully adapted for image translation, where the goal is to transform an input image into an output image with a different appearance while preserving the underlying semantic content. This makes it an ideal candidate for medical image synthesis tasks, such as inter-modality MRI translation.

In this work, we employ two U-Net–based architectures for our proposed solution. The first model follows the architecture described in [21], which has been modified to better suit the requirements of MRI image translation. A notable enhancement in our implementation is the inclusion of dropout layers, a regularization technique designed to mitigate overfitting during training. Dropout operates by randomly deactivating a fraction of neurons in a layer during each training iteration, compelling the network to learn more robust and generalizable features. This addition ensures that the model does not become overly reliant on any specific features in the training data, thereby improving its performance on unseen inputs. The architecture used in this study introduces additional dropout layers compared to the original U-Net model described in [19]. Figure 3 illustrates the modified U-Net architecture employed in this research, highlighting the placement of convolutional, deconvolutional, and dropout layers.

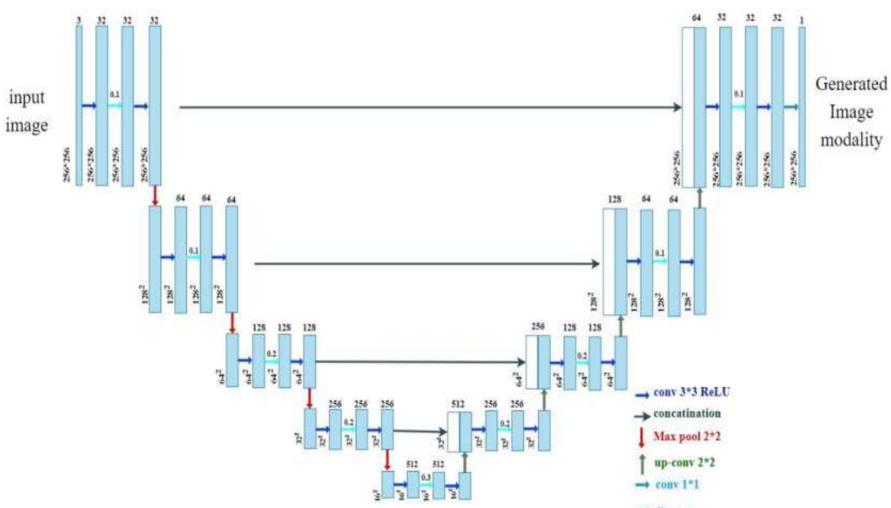

Fig. 3. U-net deep learning model proposed in paper [21].

Furthermore, we have refined the dropout rates across the convolutional and deconvolutional blocks of the network. In the original model, the dropout rates were set as (0.1, 0.1, 0.2, 0.2, 0.3, 0.2, 0.2, 0.1, 0.1), whereas in our proposed model, they have been adjusted to (0.1, 0.1, 0.3, 0.3, 0.5, 0.3, 0.3, 0.1, 0.1). This progressive increase in dropout, particularly in the deeper layers, enhances the network’s ability to generalize by reducing its sensitivity to specific features in the training set. By preventing the model from overfitting, these modifications allow it to capture more meaningful and generalizable patterns in the input images.

The combined effect of these architectural enhancements is improved image translation performance, particularly for medical imaging applications where the accurate reconstruction of anatomical structures and preservation of fine details are paramount. By balancing feature learning with regularization, the modified U-Net model is better equipped to generate high-quality, realistic outputs while maintaining the semantic integrity of the input images. Consequently, these modifications not only improve the robustness of the model but also enhance the clinical relevance and applicability of the translated MRI images.

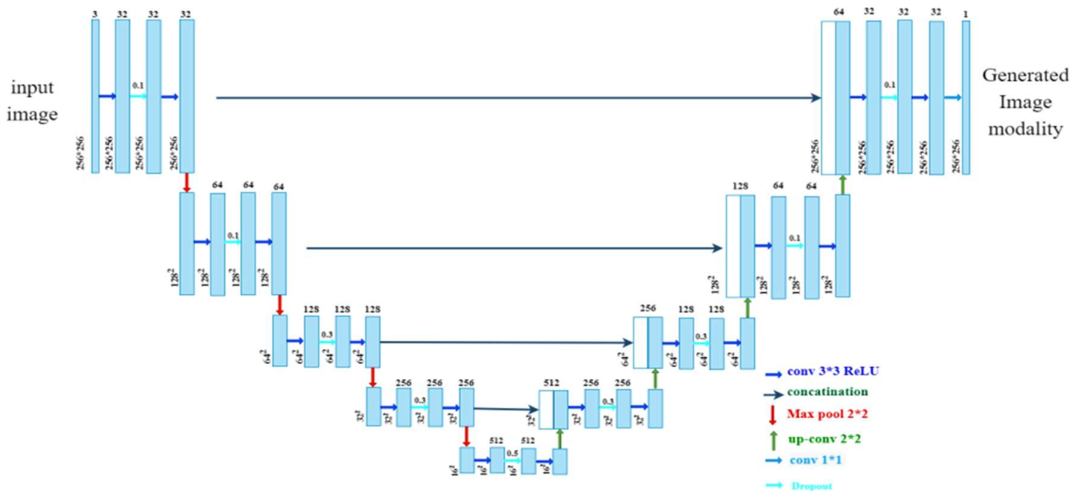

Fig. 4. Proposed U-net deep learning model

Thus, increase in dropout rate results in a greater emphasis on the remaining features, ultimately increasing the overall peak signal-to-noise ratio (PSNR) score. Fig illustrates the architecture of proposed U-net deep learning models used in the study.

Dataset information

The dataset is given by Horizon imaging centre. The radiologist has provided real data for our research work. We have dataset having MRI of 50 subjects in DICOM format. Each subject has a different set of image modality sequences. Among them, for each modality (T1WI, T2WI and Flair) per subject 20 slices in axial view are available. Dataset contain MRIs of both type of brain, with diseases and without diseases. Image resolution of T1WI is 256*176, T2WI is 384*300 and Flair is 320*270. For each sequence scan timing is not fixed, it is up to radiologist to decide scan timing. From which imaging centre we have acquire dataset uses scan timing for acquisition as mention in Table 1. We have T1WI, T2WI and Flair images which can be view using special software called MicroDicom DICOM viewer. Image representation of that is as below:

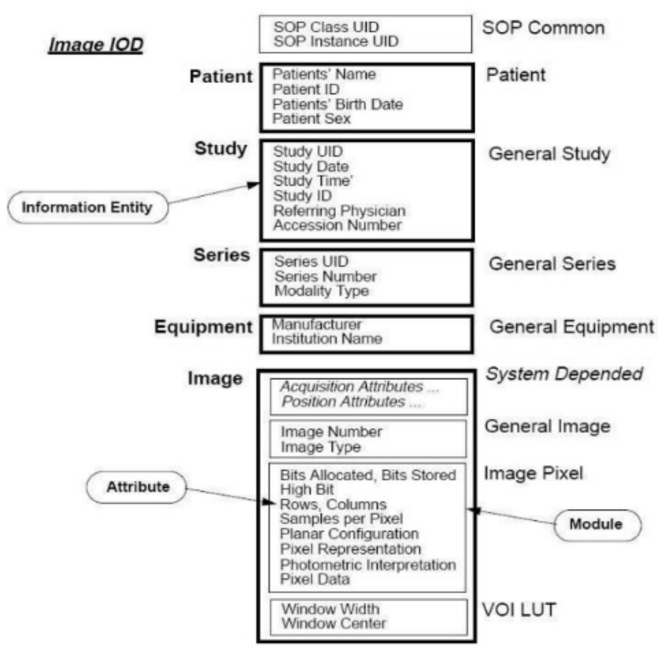

Fig. 5. Structure of DICOM study

Fig. 6. Seventh slice of T1WI, T2WI and Flair Modality in DICOM format

TABLE I. ACQUISITION TIMING USED FOR ACQUIRING BRAIN MRI

Modality	Acquisition timing for each modality in MRI facilities
T1WI	1:00 to 4:00 mins
T2WI	1:30 to 4:00 mins
Flair	2:00 to 5:00 mins
T1WI+T2WI+Flair	4:30 to 13:00 mins

Details of implementation

Since we were able to get paired DICOM studies, we chose to employ the supervised image to image translation technique on paired data since it has a higher accuracy than unsupervised image to image translation being applied to unpaired image data [11].

Preprocessing steps

Data cleaning

Data is collected from hospital so it is in row form. One of benefits of using deep learning is that is require less annotation tasks as compared to machine learning. We have done minimum pre-processing on our data to maintain the actual features of the data. The collected data was of 50 subjects among them 5 subjects has not require a modality file having '.dcm' format. So, have removed 5 subjects' data.

Out of the total of 45 subjects, the last four slices of each subject were found to be unsuitable for model training and therefore were excluded. Furthermore, an additional 2 subjects' data and 3 slices were removed from the dataset due to issues related to blurriness and unclear visual quality. After, performing cleaning of slices dataset has now 685 slices per image modality (T1WI, T2WI and Flair).

Same resolution of (T1WI, Flair) and (T2WI, Flair)

It is important to make image resolution of paired data and label same for image translation task to generate corresponding output. For the same task in built functions of python's SimpleITK library were used. By applying those functions pair of T1WI and Flair has same image resolution of 256*176 and the pair of T2WI and Flair has same image resolution of 300*384.

Sorting of slices

Available DICOM studies contains slices in random manner in each directory. For each patient, all slices per modality are not mapped to another modality. We need each paired slice for our study. For obtaining paired data slices DICOM files parameter called instancenumber for slice number is used. All slices are sorted using this parameter and now slices are in sorted manner (in accordance to slice number in series).

Image to NumPy array conversion

To simplify future computation, all DICOM encoded image modalities are transformed to NumPy arrays. The pixel_array built-in function of the Pydicom package is employed for this. Each slice for each modality is transformed into an array with a shape appropriate to its image resolution.

Crop and padding of image

Each modality has different image resolution and for our model input it was required to have all paired data images to be in same 1:1 resolution. For that we have used crop and padding concept. In each image background pixel value is 0 while brain structure has non zero-pixel value. Using this brain structure from image is cropped.

After that in new 2D array of desired size having all zeros, OR operation is performed with cropped image. In this way, we have desired sized image having brain structure at centre of image. We have employed this approach for getting all images in 256*256 sized array for all 685 slices of T1WI and corresponding Flair image and 480*480 sized array for all 685 slices of T2WI and corresponding Flair image.

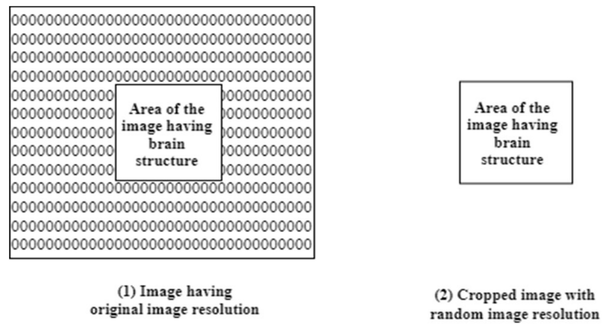

Fig. 7. (A) Diagram illustrating concept of cropping and padding with zero

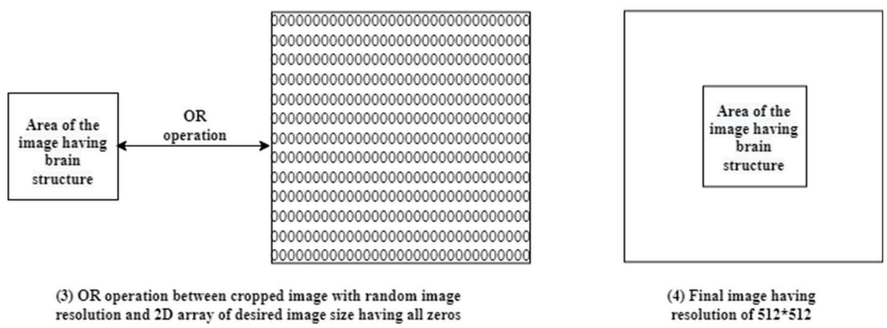

Fig. 7. (B) Diagram illustrating concept of cropping and padding with zero

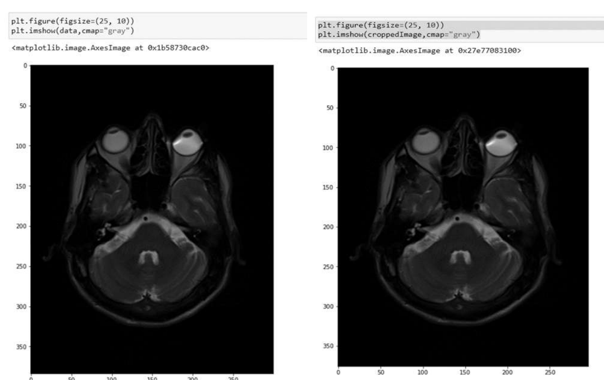

Fig. 8. (A) Plotting of Single slice of T2WI modality: Before and After cropping and padding having a resolution of 300*384 and 376*296,

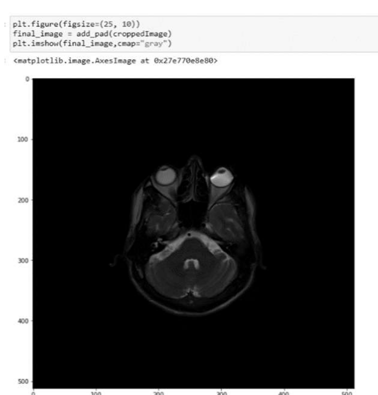

Fig. 8. (B) Plotting of Single slice of T2WI modality: After padding having final image resolution of 480*480

Normalization:

Each MRI image is normalized using min-max normalization since the intensity of the images typically varies. Due to the different scales of the intensity values collected from the pixel array, they frequently contribute equally to the model fitting and model learning functions, which causes the model to be biased. So, in order to solve this issue, we employ normalization [5].

Every MRI image is standardized using min-max normalization since the MRI images' intensities typically vary. The model is biased because the intensity values acquired from the pixel array are of different sizes and hence likely to contribute equally to model fitting and model learning. For normalization in we employed min-max normalization method among other normalization techniques. After applying min-max normalization we got all values in array having values between 0 to 1.

Training and testing phase

Division of Dataset into Training, Validation, and Testing Dataset

Dataset is divided into Training, Validation and Testing Dataset in 80%, 10%, 10% respectively for both Dataset containing pairs of (T1WI, Flair) and (T2WI, Flair) for 100% dataset and for 50% of dataset. Total of four dataset were used for translation task-1) Dataset containing paired 685 slices of T1WI and Flair image, 2) Dataset containing paired 685 slices of T2WI and Flair image, 3) Dataset containing paired 342 slices of T1WI and Flair image, 4) Dataset containing paired 342 slices of T1WI and Flair image

Training of deep learning model to generate flair image from T1WI and to generate flair image from T2WI

Deep learning model is trained to generate flair image on 100% dataset and 50% dataset for all four datasets (for translation of T1WI to Flair image and T2WI to Flair image). on two separate deep learning model. In total 8 ways flair image will get generated.

In both deep learning models T1WI/T2WI is treated as data while Flair image is treated as label. The model will try to predict pixel value for the generation of flair image from the data's pixel value, taking reference of label's pixel value making it as regression task.

Testing of Deep learning model to generate flair modality from T1WI and T2WI

Testing was done on unseen data where only data provided as input to the model and the corresponding flair image is generated. For testing of the model different type of loss function get calculated for generated flair image against the ground truth flair image. On the trained deep learning model, testing was conducted using unseen images of T1WI and T2WI as inputs for eight separate trained models (one for each dataset). The models were designed to generate a Flair image. The Flair image is produced as an output of the trained deep-learning models.

Evaluation of deep learning model using generated flair image and ground truth flair image

After the testing phase is finished evaluation is carried out for generated flair image. Analysis for generated Flair images from T1WI and from T2WI is done. Also, comparison of two U-net-based model-generated flair images been made. Result of all deep learning models with epoch of 150 is noted for both dataset (100% and 50% of the Dataset). Here, the calculation of all loss function is done with generated flair image against ground truth flair image. The formula used for calculation of loss function is mention below.

Mean Squared Error (MSE): Mean Squared Error loss, or MSE for short, is calculated as the average of the squared differences between the predicted and actual values. The result is always positive regardless of the sign of the predicted and actual values and a perfect value is 0.0 [19].

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \underline{y_i}\right)^2 \qquad (1)$$

Where, y_i is the numpy array of Flair ground truth and $\underline{y_i}$ is the numpy array of generated Flair [21].

Peak Signal-To-Noise Ratio (PSNR): PSNR (Peak Signal-to-Noise Ratio) is a commonly used metric to evaluate the quality of synthesized images [32], and it typically falls within the range of 30 to 50 dB. A higher PSNR value indicates better image quality, which means that the synthesized MRI is accurately predicting the evenly or sparsely distributed data. Generally, a PSNR value over 40 dB is considered very good, while a value below 20 dB is unacceptable [31].

$$PSNR = 10 * \frac{MAX_1^2}{MSE} \qquad (2)$$

Where, MAX_1 is the maximum possible pixel value of the generated Flair image [21].

Mean Absolute Error (MAE): Sums the absolute differences between the predictions and ground truth, and finds the average [21].

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - x_i| \quad (3)$$

Where, y_i is the numpy array of Flair ground truth and x_i is the numpy array of generated Flair [21].

Result and analysis

Following table represents results scores- PSNR, MAE, MSE for generated flair image in all experiments.

TABLE II. RESULT SCORE FOR TRANSLATION OF T1WI TO FLAIR IMAGE

T1WI to Flair Image Translation		
Epoch-150		
Dataset-342 Slices		
	Proposed U-net	U-net used in Paper [21]
	Testing Score	Testing Score
PSNR	22.444	20.731
MAE	00.032	00.038
MSE	00.005	00.008
Epoch-150		
Dataset-685 Slices		
	Proposed U-net	U-net used in Paper [21]
	Testing Score	Testing Score
PSNR	22.577	21.960
MAE	00.035	00.038
MSE	00.005	00.006

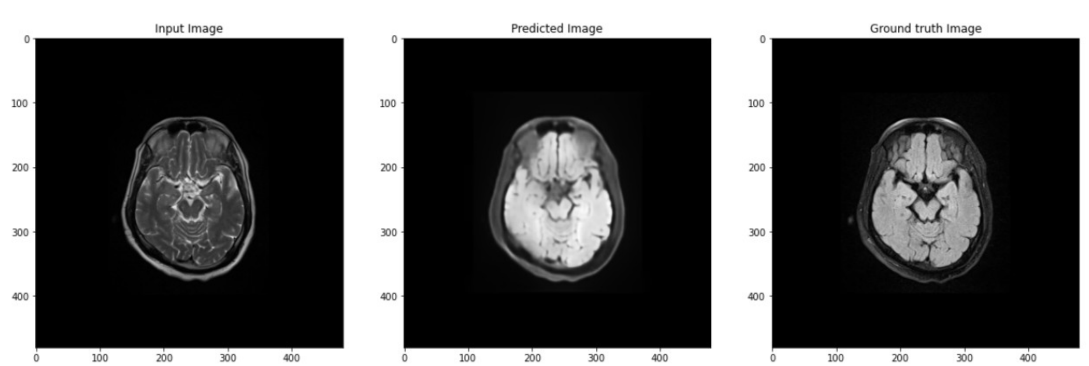

Fig. 9. (A) Representation of single slice of input image (source modality) T1WI, generated image Flair and Ground truth image (target modality) Flair image

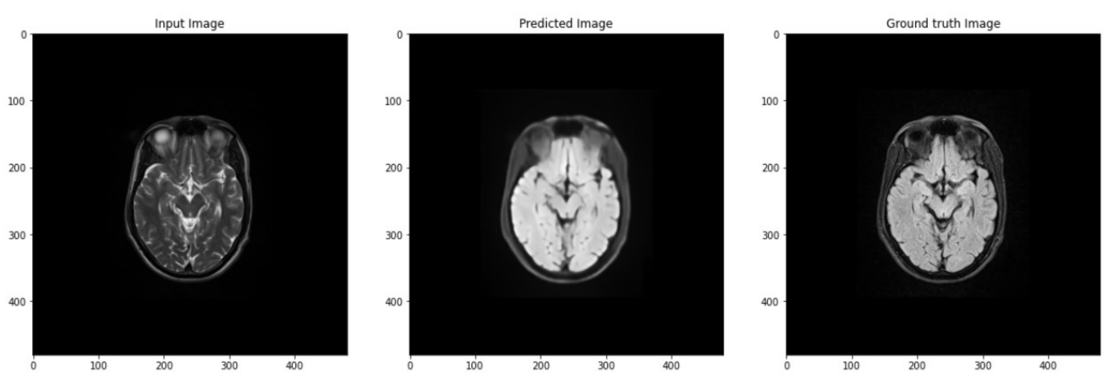

Fig. 9. (B) Representation of single slice of input image (source modality) T1WI, generated image Flair and Ground truth image (target modality) Flair image

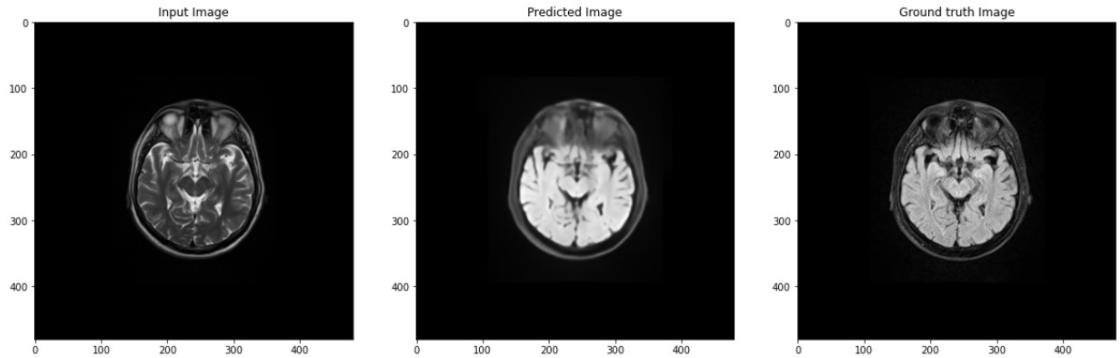

Fig. 9. (C) Representation of single slice of input image (source modality) T1WI, generated image Flair and Ground truth image (target modality) Flair image

It has been observed that for Proposed U-net model translation task T2WI to Flair image result score is higher than other experiments. Additionally, between Proposed U-net and U-net used in paper [21], proposed U-net yield higher result score with higher generated image quality. Increasing training data lead to higher result scores for both translation- T1WI to Flair image and T2WI to Flair image. In future, for increased number of epoch and for more training data result score can significantly improve. Below is a representation of some of the generated flair image from T2WI of training data of 685 slices using the proposed U-net.

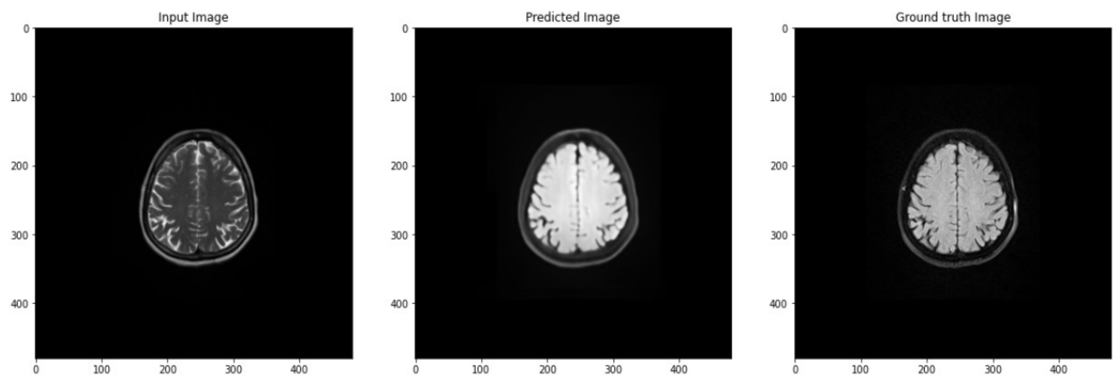

Fig. 10. (A) Representation of single slice of input image (source modality) T2WI, generated image Flair and Ground truth image (target modality) Flair image

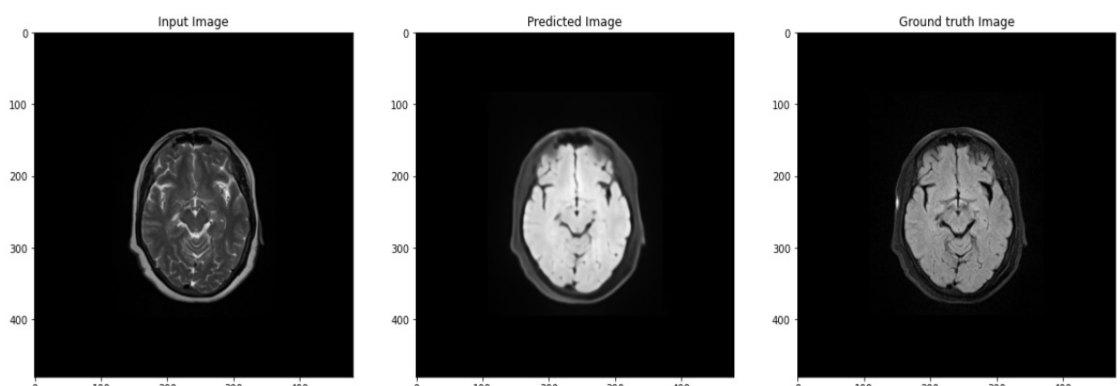

Fig. 10. (B) Representation of single slice of input image (source modality) T2WI, generated image Flair and Ground truth image (target modality) Flair image

TABLE III. RESULT SCORE FOR TRANSLATION OF T2WI TO FLAIR IMAGE

T2WI to Flair Image Translation		
Epoch- 150		
Dataset 342		
	Proposed U-net	U-net used in Paper [21]
	Testing	Testing
PSNR	25.3080	25.2670
MAE	0.0180	0.0200
MSE	0.0020	0.0030
Epoch- 150		
Dataset 685		
	Proposed U-net	U-net used in Paper [21]
	Testing	Testing
PSNR	26.3570	26.2710
MAE	0.0140	0.0150
MSE	0.0020	0.0020

IV. FUTURE SCOPE AND CONCLUSION

In this work, we reviewed deep learning models for image-to-image translation in the medical domain, including various types of medical image translation and prior related research. Using an image-to-image translation approach, we implemented a solution to generate FLAIR image modalities from T1WI and T2WI scans. For the inter-modality translation task, DICOM data was pre-processed, and FLAIR images were generated using both our proposed U-Net model and the U-Net architecture described in [21]. We analyzed the quality of generated FLAIR images from both models and compared results obtained from T1WI-based versus T2WI-based generation.

The proposed solution has the potential to reduce MRI acquisition time and minimize motion artifacts, benefiting patients who are claustrophobic, pediatric, or elderly by providing faster, safer, and more comfortable MRI scans. It also optimizes medical resources and reduces scan time for healthcare providers. In future work, we plan to leverage our DICOM dataset with the proposed U-Net model to generate FLAIR modalities from the synthesized T2WI images, further enhancing the efficiency of MRI workflows.

ACKNOWLEDGMENTS

The authors would like to express their gratitude to the Horizon Imaging Centre for granting access to their advanced imaging facilities, which were essential for conducting this research. Additionally, they appreciate the provision of brain MRI data, which was crucial for the successful execution of the study.

REFERENCES

[1] Marjan Laal., “Innovation Process in Medical Imaging”, Procedia - Social and Behavioral Sciences Vol. 81, pp.60-64 2013.

[2] Hussain S, Mubeen I, Ullah N, Shah SSUD, Khan BA, Zahoor M, Ullah R, Khan FA, Sultan MA. “Modern Diagnostic Imaging Technique Applications and Risk Factors in the Medical Field: A Review”, BioMed research international, Jun 2022.

[3] van Beek, E. J. R., Kuhl, C., Anzai, Y., Desmond, P., Ehman, R. L., Gong, Q., Gold, G., Gulani, V., Hall-Craggs, M., Leiner, T., Lim, C. C. T., Pipe, J. G., Reeder, S., Reinhold, C., Smits, M., Sodickson, D. K., Tempany, C., Vargas, H. A., & Wang, M. , “Value of MRI in medicine: More than just another test?”, Journal of magnetic resonance imaging: JMRI, vol. 49 e14-e25. Jun 2019.

[4] Plantone, D., Renna, R., Sbardella, E., Koudriavtseva, T., “Concurrence of Multiple Sclerosis and Brain Tumors”, Frontiers in Neurology, Vol 6:40, March 2015.

[5] Yang Q, Li N, Zhao Z, Fan X, Chang EI, Xu Y , “MRI Cross-Modality Image-to-Image Translation”, Scientific reports, Vol 10(1), pp.1-18, Feb 2020.

[6] Sanuwani Dayarathna, Kh Tohidul Islam, Sergio Uribe, Guang Yang, Munawar Hayat, Zhaolin Chen, “Deep learning based synthesis of MRI, CT and PET: Review and analysis”, Medical Image Analysis, Vol 92(103046), Feb 2024.

[7] Kabasawa, Hiroyuki, and Shigeru Kiryu.“Pulse sequences and reconstruction in Fast MR imaging of the liver”, Magnetic Resonance in Medical Sciences Vol 22(2), pp. 176-190. Apr 2023.

[8] P. Srinivasan, P. Kaur, A. Nigam and A. Bhavsar, “Semantic Features Aided Multi-scale Reconstruction of Inter-Modality Magnetic Resonance Images,” 2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS), Rochester, MN, USA, pp. 110-113, 2020.

[9] Xu, Y., Li, Y., Shin, BS., "Medical image processing with contextual style transfer", Human-centric Computing and Information Sciences, Vol 10, pp. 1-16, 2020.

[10] Waqar Ahmad,Hazrat Ali,Zubair Shah, Shoaib Azmat , "A new generative adversarial network for medical images super resolution", Scientific Reports, Vol 12(9533), 2020.

[11] Pauliina Paavilainen, Saad Ullah Akram, Juho Kannala, "Bridging the gap between paired and unpaired medical image translation" , MICCAI Workshop on Deep Generative Models, Springer, pp. 35-44, 2021.

[12] Tanishq Abraham, Austin Todd, Daniel A. Orringer, Richard Levenson, "Chapter 7 Applications of artificial intelligence for image enhancement in pathology." Artificial Intelligence and Deep Learning in Pathology, pp. 119-148. , 2021.

[13] Miller MI, Christensen GE, Amit Y, Grenander U., "Mathematical textbook of deformable neuroanatomies." Proceedings of the National Academy of Sciences Vol 90(24) , pp. 11944-11948, Dec 1993.

[14] Hien Van Nguyen, Kevin Zhou, Raviteja Vemulapalli., "Cross-domain synthesis of medical images using efficient location-sensitive deep network", Medical Image Computing and Computer-Assisted Intervention--MICCAI 2015: 18th International Conference, Munich, Germany, Part I 18, pp. 677-684,Oct 2015.

[15] Uh J, Merchant TE, Li Y, Li X, Hua C. "MRI-based treatment planning with pseudo CT generated through atlas registration.", Medical physics Vol 41(5):051711, 2014.

[16] Catana C, van der Kouwe A, Benner T, Michel CJ, Hamm M, Fenchel M, Fischl B, Rosen B, Schmand M, Sorensen AG. "Toward implementing an MRI-based PET attenuation-correction method for neurologic studies on the MR-PET brain prototype." Journal of nuclear medicine Vol 51(9), pp. 1431-1438, 2010.

[17] Rousseau, François. "Brain hallucination", Computer Vision–ECCV 2008: 10th European Conference on Computer Vision, Proceedings, Marseille, France, 2008, Part I 10: pp. 497-508., 2008.

[18] Amini Amirkolaee, Hamed, and Hamid Amini Amirkolaee. "Medical image translation using an edge-guided generative adversarial network with global-to-local feature fusion.", Journal of Biomedical Research Vol 36(6), pp 409, 2022.

[19] Olaf Ronneberger, Philipp Fischer, Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, part III 18, pp. 234-241, 2015.

[20] Kalantar R, Messiou C, Winfield JM, Renn A, Latifoltojar A, Downey K, Sohaib A, Lalondrelle S, Koh DM, Blackledge MD. "CT-based pelvic T1-weighted MR image synthesis using UNet, UNet++ and cycle-consistent generative adversarial network (Cycle-GAN)", Frontiers in Oncology, Vol 11: 665807, 2021.

[21] Ninad Anklesaria , Yashvi Malu , Dhyey Nikalwala , Urmi Pathak , Jinal Patel , Nirali Nanavati , Preethi Srinivasan and Arnav Bhavsar "Multi Modality Medical Image Translation for Dicom Brain Images." BIOIMAGING, pp. 168-173, 2022.

[22] Osman AFI, Tamam NM. "Deep learning-based convolutional neural network for intramodality brain MRI synthesis.", Journal of Applied Clinical Medical Physics Vol 23(4): e13530, 2022.

[23] Xiaocong Chen, Yun Li, Lina Yao, Ehsan Adeli, Yu Zhang, Xianzhi Wang., "Generative adversarial u-net for domain-free few-shot medical diagnosis", Pattern Recognition Letters Vol 157, pp 112-118, 2022.

[24] Li Y, Li W, Xiong J, Xia J, Xie Y. "Comparison of supervised and unsupervised deep learning methods for medical image synthesis between computed tomography and magnetic resonance images." BioMed Research International , Vol 2020: 5193707, Nov 2020.

[25] Xiang L, Chen Y, Chang W, Zhan Y, Lin W, Wang Q, Shen D. "Ultra-fast t2-weighted mr reconstruction using complementary t1-weighted information." Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, Part I, pp.215-223, 2018.

[26] Arco, Juan E. and Jiménez-Mesa, Carmen and Ortiz, Andrés and Ramírez, Javier and Levin, Johannes and Górriz, Juan M, "Explainable Inter-Modality Medical Information Transfer Using Siamese Autoencoders", IEEE Transactions on Radiation and Plasma Medical Sciences, pp. 1-1, 2025.

[27] Chaohui Gong and Zhiying Wu and Zisheng Huang and Gaofeng Meng and Zhen Lei and Hongbin Liu, "MM2CT: MR-to-CT translation for multi-modal image fusion with mamba", (yet to publish)

[28] Hongxu Jiang, Muhammad Imran, Teng Zhang, Yuyin Zhou, Muxuan Liang, Kuang Gong, and Wei Shao. "Fast-DDPM: Fast Denoising Diffusion Probabilistic Models for Medical Image-to-Image Generation" IEEE Journal of Biomedical and Health Informatics, pp. 1-11,April 2025.

[29] Bidgood WD Jr, Horii SC, Prior FW, Van Syckle DE. "Understanding and using DICOM, the data interchange standard for biomedical imaging." Journal of the American Medical Informatics Association, Vol 4(3),pp. 199-212, 1997.

[30] Todd A. Gould, RT-(R) (MR)(ARRT) & Molly Edmonds, "How MRI Works",, https://science.howstuffworks.com/mri.htm last accessed on 2022/12/05.

[31] National Institutes of Health, "Magnetic Resonance Imaging (MRI)", https://www.nibib.nih.gov/science-education/science-topics/magnetic-resonance-imaging-mri last accessed on 2022/12/05.

[32] "PSNR score in image translation", https://in.mathworks.com/help/vision/ref/psnr.html, last accessed on 2023/11/24.