Enhancing Chronic Stroke Lesion Detection and Segmentation through nnU-net and Multi-Modal MRI Analysis

Lounès Meddahi*, Arthur Masson†, Elise Bannier†‡, Stéphanie Leplaideur§†, Francesca Galassi†
*Ecole Normale Supérieure de Rennes, Rennes, France
†Inria, CRNS, Inserm, IRISA UMR 6074, Empenn U1228, Univ Rennes, Rennes, France
‡CHU Rennes, Department of Radiology, Rennes, France
§Physical and Rehabilitation Medicine Department, University Hospital of Rennes, Rennes, France

***Abstract*—Stroke lesion segmentation plays a crucial role in diagnosing and treating patients who have experienced a stroke. Manual segmentation, however, is labor-intensive and requires significant time and effort from experts. To address this challenge, current methods employ Deep Learning techniques to automate the process. Nonetheless, these approaches have limitations, as they do not fully exploit the 3D spatial coherence, complementary nature of MRI modalities and optimized approach. In this study, we introduce a novel method, which integrates T1-weighted and FLAIR MRI modalities through a custom algorithm to adapt a neural network for stroke lesion segmentation. Our approach initially pretrains the model on an extensive public database, followed by fine-tuning using in-house data with a combination of T1 and FLAIR modalities. The resulting model outperforms existing techniques, delivering enhanced accuracy and significantly reducing the time needed for segmentation, thus offering a more efficient solution for stroke lesion identification.**

***Index Terms*—nnU-Net, MRI modalities, Stroke lesion segmentation, Deep learning, medical images segmentation**

I. Introduction

Stroke is a leading cause of death and disability worldwide, with millions of people affected each year [1]. Early and accurate detection of stroke lesions is crucial for determining the appropriate treatment and improving patient outcomes [2]. However, manual segmentation of lesions by experts is a time-consuming and challenging task. Moreover, manual segmentation is prone to human error and variability, which can result in suboptimal treatment decisions.

The complexity of brain anatomy, the heterogeneity of stroke lesions, and variations in MRI image quality make manual lesion detection and segmentation challenging for experts. While strokes can be classified into multiple phases, the two phases of particular interest are acute (early phase, within 24 hours) and chronic (late phase, after several weeks). Segmentation of chronic strokes is particularly difficult due to factors such as lesion fogging and tissue changes over time. The industry has primarily focused on developing tools for acute stroke segmentation due to its urgency in clinical settings. However, accurate and reliable tools for chronic stroke segmentation are equally important for medical purposes, such as monitoring disease progression. As a result, there is a growing need to improve automated methods to accurately identify and segment both acute and chronic stroke lesions, with a current shortage of tools available for chronic stroke segmentation.

In recent years, advancements in artificial intelligence and Deep Learning (DL) have led to the development of sophisticated models capable of performing complex tasks, such as medical image segmentation, with remarkable accuracy. These models have the potential to revolutionize stroke lesion detection and segmentation, significantly reducing the time and effort required by experts, while also improving the overall accuracy and reliability of the process.

In this research paper, we aim to propose and evaluate the potential of DL techniques, specifically nnU-net and fine-tuning, for the automated segmentation of chronic stroke lesions from multimodal MRI data. We will examine the benefits of incorporating both T1-weighted and FLAIR (Fluid-Attenuated Inversion Recovery) MRI modalities into our model to enhance segmentation accuracy and robustness. Additionally, we will investigate the use of pretrained models and finetuning techniques to optimize our model's performance and comparing its results to those achieved through manual segmentation by experts.

II. Related Works

Before delving into the details of our proposed approach, we first present a literature review of existing works related to DL applied to MRI segmentation tasks.

Since the introduction of the U-net architecture for medical image segmentation in 2015 [3], numerous papers have presented U-net based architectures for brain disease segmentation [4]. These architectures aim to improve image segmentation robustness and generalization through the use of deeper and more complex U-Net architectures [5]–[8], attention mechanisms [9], [10], data augmentation techniques [7], [11], and multi-task learning [12], [13]. Selecting the appropriate architecture depends on various factors, such as the specific task, dataset characteristics, and performance requirements. Recently, some U-net based architectures [8], [10], [13], [14] have been specifically designed to enhance MRI segmentation accuracy, emphasizing the growing interest

in optimizing models for this critical medical application. Among these, the nnU-net model has shown the most promise for medical image segmentation tasks. A key strength of nnU-Net is its automated optimization process, which can be both time-consuming and challenging. nnU-Net achieves this by employing a series of optimization strategies, including hyperparameter tuning, data augmentation, and model ensembling, tailored to the specific medical image segmentation task.

- In order to capitalize on the exceptional performance of the nnU-net model, [14] utilized an nnU-net adapted model [15] for stroke lesion segmentation based on T1-weighted MRIs. They employed T1-weighted MRIs provided by ATLAS v2.0 [16], a large dataset consisting of 655 stroke patients with T1-weighted MRIs and corresponding lesion masks. Their approach achieved an average Dice score of 0.650, which is currently the best score on ATLAS v2.0. However, in the context of stroke lesion segmentation, relying solely on T1-weighted MRIs is insufficient for achieving precise segmentation. To address this limitation, neuroscientists and experts typically use a combination of T1-weighted and FLAIR images for brain imaging studies. While T1-weighted images provide high contrast between various brain tissues and are valuable for visualizing brain anatomy, FLAIR images are especially sensitive to pathological changes in the brain, such as edema, inflammation, and demyelination—changes that are difficult to detect in T1-weighted images alone. Integrating both T1-weighted and FLAIR images enables experts to accurately detect and characterize stroke-related brain lesions, facilitating effective treatment planning and disease progression monitoring.
- For comparison, [18], [19] utilized multiple imaging modalities to train their models, demonstrating the relevance of this approach. [18] trained a VGG-SegNet model based on VGG16 to produce ischemic stroke lesion segmentation from 2D slices of Flair/DW/T1-weighted MRIs. As for [19], they used FLAIR MRIs to produce a segmentation of brain tumours using their own DL model. Although nnU-net demonstrated better results than their models for medical image segmentation tasks, their article proved that training a model with multiple modalities to predict a single segmentation is a more effective approach than using only one modality. Moreover, [18] showed that using a pre-trained model, even if it was not initially trained on the stroke segmentation task, leads to better performance once finetuned. However, unlike [14], they did not use 3D images to leverage the spatial coherence of lesions, which are continuous in the brain. Exploiting 3D coherence is important because it helps the model better capture the spatial relationships between neighboring voxels, enabling more accurate segmentation of lesion boundaries and improved detection of subtle morphological changes within the brain tissue. As for [19], they demonstrated that using FLAIR MRIs for segmentation in the case of brain tumors yields good results. On the BRATS 2019 database, their model achieved a DICE score of 0.841, outperforming the top 8 other models in comparison.

In light of these articles and the literature, we aim to develop a better segmentation tool for brain stroke lesions by leveraging large datasets, 3D images, and the complementarity of different modalities. Building on the work by [14], [18], we have decided not to use only 3D T1-weighted MRIs for segmentation, but to use T1-weighted MRIs from ATLAS v2.0 to pretrain a model and then combine T1-weighted and FLAIR MRIs, as [19], to finetune this model to get better results.

III. Preliminaries

In this section, we provide an overview of the essential concepts and techniques employed in our approach.

A. nnU-net

Medical image segmentation has witnessed the rise of the nnU-Net [20] as a leading approach for various tasks, including stroke segmentation in FLAIR MRI [12], [13], [19]. Introduced by Isensee in 2021, the nnU-Net (no-new U-Net) is a robust, self-adapting framework designed to eliminate the need for manual architecture engineering. The primary goal behind nnU-Net's development was to establish a framework capable of consistently achieving state-of-the-art performance across diverse segmentation tasks, without necessitating task-specific modifications or hyperparameter tuning.

The 3D nnU-Net, depicted in Fig.1, is a U-Net architecture incorporating various design principles and best practices from DL literature. These enhancements include an automated architecture search, which adapts the model's architecture based on input data, and advanced training techniques such as focal loss, deep supervision, and data augmentation strategies. Furthermore, the nnU-Net framework offers a preprocessing pipeline that automatically handles the standardization, resampling, and normalization of input data, ensuring optimal performance across varied datasets.

In our research, we employ the nnU-Net framework for segmenting stroke lesions from T1 and FLAIR MRI scans. By harnessing the strengths of the nnU-Net, we aim to provide an accurate and robust stroke segmentation solution that can be easily generalized to other datasets and tasks, ultimately advancing the development of automated diagnostic tools for stroke patients.

B. LongiSeg4MS

The Longitudinal Segmentation for Multiple Sclerosis (LongiSeg4MS) model [21] is a specialized method for detecting new MS lesions using the nnU-Net framework. LongiSeg employs two sets of MRI scans to compute a segmentation map of the lesions. These scans can be a combination of FLAIR, T1, or T2 MRI modalities, or they can comprise probability maps obtained from cross-sectional predictions. LongiSeg executes the following three steps:

- Preprocess the data by applying multiple preprocessing steps (e.g., reorient, register, crop, normalize, adjust the images),

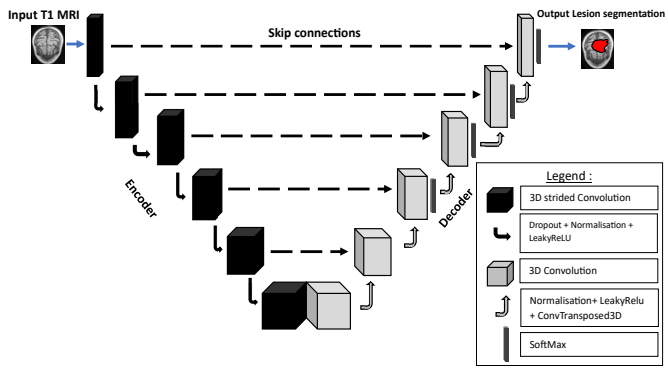

Fig. 1. Simplified nnU-net architecture

- Compute predictions, which are probability maps containing the likelihood of identifying new lesions (this step uses the trained model),
- Postprocess these maps by thresholding them to obtain the final segmentations, which are then transformed back to the original space (removing lesions that are too small).

C. MRI Overview

Magnetic Resonance Imaging (MRI) is a non-invasive medical imaging technique that sets itself apart from other imaging modalities due to its capacity to capture detailed, high-resolution images of internal human body structures without exposing patients to ionizing radiation. This technique is based on nuclear magnetic resonance principles and is particularly beneficial for stroke diagnosis and monitoring, as it offers insights into affected brain tissues, enabling experts to distinguish between ischemic and hemorrhagic strokes and determine the extent of the damage.

The inherent complexity and variability of MRI data present distinct challenges for image processing and segmentation techniques. Traditional methods can be labor-intensive and require substantial expertise, which may lead to longer processing times and increased variability in results. DL approaches, such as the nnU-Net, offer an efficient and accurate alternative by automating the segmentation process, significantly reducing the manual workload for experts and providing consistent, high-quality outcomes.

IV. Materials and Methods

In this section, we describe the materials and methodologies employed to develop a robust DL model for stroke lesion segmentation.

A. Databases

We used two databases for analysis. The first database is the publicly available ATLAS v2.0 [16], an extended version of the ATLAS v1.2 database [17], specifically designed to facilitate research in stroke lesion segmentation. This database provides a diverse set of T1-weighted MRI scans from 655 subjects, sourced from multiple sites and scanners. The ATLAS v2.0 database is a valuable resource for training DL models, given its large size, high resolution, and multi-site nature, offering a wide variety of MRI images. Comprehensive metadata for the ATLAS v2.0 database is presented in Table I.

The second database is an in-house dataset comprising MRI scans from 47 patients from the IRB approved national clinical trial AVCPOSTIM (NCT01677091)[1]. This dataset includes both FLAIR and T1-weighted MRI scans, with manual segmentations performed on each patient. Although smaller in size, the in-house dataset complements the ATLAS v2.0 database by providing additional MRI modalities and annotations, which can enhance the model's generalizability and performance across various imaging contexts.

TABLE I: ATLAS V2.0 Summary

Metadata Field	Probability
Primary Stroke Hemisphere	0.45 (left), 0.44 (right), 0.11 (other)
Secondary Stroke Hemisphere	0.14 (left), 0.15 (right), 0.09 (other)
Lesion Volume (mm^3)	4000 ($\leq$50%), 28607 (avg)
Days Post Stroke (50%, avg)	507 ($\leq$50%), 953 (avg)

By combining the ATLAS v2.0 database with our in-house dataset, we aim to create a robust training environment for our DL model, ensuring improved performance and generalizability across a broad range of stroke lesion segmentation tasks. The diverse nature of the data, including different MRI modalities and multiple acquisition sites, will enable our model to adapt to various imaging conditions.

B. Preprocessing

Preprocessing plays a vital role in machine learning and data analysis, as it involves cleaning, transforming, and preparing raw data for further processing. The quality of input data significantly influences the accuracy and effectiveness of a model. In this subsection, we detail the preprocessing steps undertaken to clean and transform our raw data prior to its use in model training.

1) MR volumes are reoriented in RAS coordinates, ensuring consistent orientation across all images, which is crucial for accurate analysis and comparison.
2) Skulls and skin tissues are removed using animaAtlasBasedBrainExtraction[2]. This step eliminates noise and artifacts that may interfere with subsequent processing stages, such as segmentation and registration.
3) In the in-house database, since segmentation is performed on the FLAIR scans, T1-weighted images are resampled to match the FLAIR scans' spatial resolution. Subsequently, segmentations are realigned on the T1 scans, ensuring that both T1 and FLAIR scans have the same size and that segmentations are accurately placed over the T1 scans.
4) Images are cropped using the T1 scans as masks, improving consistency, reducing noise, enhancing accuracy, and decreasing computational requirements.

[1]https://clinicaltrials.gov/ct2/show/NCT01677091
[2]anima.irisa.fr

5) Bias arising from spatial inhomogeneity is estimated using the N4itk algorithm [22] and removed from the data (employing animaN4BiasCorrection[1]). Bias field or intensity inhomogeneity is a common issue in MR images, where signal intensity varies across the image due to factors such as magnetic field strength variations and tissue properties. This can cause inaccuracies in subsequent processing stages, including image registration, segmentation, and classification.

C. Data Augmentation

Data augmentation is a widely adopted technique in machine learning and computer vision that expands training datasets by generating additional samples from existing ones. This method is essential for enhancing the generalization and robustness of machine learning models. In this subsection, we delineate the data augmentation techniques employed in our study to boost our model's performance.

- Rotations and scaling are applied with a probability of 0.2 each. Rotation along each axis follows a uniform distribution, *U(-30,30)*. Scaling within the voxel grid entails multiplying coordinates by a scaling factor randomly sampled from *U(0.7,1.4)* for all patch types.
- Mirroring is conducted with respect to the sagittal plane, providing additional perspective on the data without altering the fundamental structure.

D. Postprocessing

Postprocessing is another crucial step in machine learning, involving the application of additional transformations to a model's output to enhance its performance or interpretability. This step is particularly vital when working with intricate models or in scenarios where prediction accuracy is critical. In this subsection, we outline the postprocessing techniques employed to refine our model's output and boost its overall performance.

- The softmax output map is converted into a binary map by applying a threshold of 0.2, improving the interpretability of the segmentation results.
- Connected components with a 26-connectivity and a volume greater than $4mm^3$ are retained, while others are discarded, ensuring the model's output remains focused on relevant and meaningful structures.
- The segmentation mask is resampled to match the slab and resolution of the original image at the first time step, facilitating accurate comparisons and evaluations of the model's performance against the ground truth.

The chosen values of 0.2 for the threshold, 26-connectivity, and a volume greater than $4mm^3$ were determined empirically, as they yielded the best results in our study.

E. Loss function

The loss function (3) employed in nnU-Net is a combination of the Dice loss (1), also known as the Sørensen–Dice index, and the binary cross-entropy loss (2). The Dice loss measures the similarity between two sets of data. In our case, it evaluates the similarity between the ground-truth segmentation (represented by set G) and the segmentation produced by our model (represented by set S). However, the Dice loss is limited as it does not penalize false positives or false negatives.

On the other hand, the binary cross-entropy loss considers a specific voxel of a segmentation. In this context, i represents a specific voxel, $o_i \in [0, 1]$ is the probability of i being part of the lesion to be segmented or not, and $y_i \in \{0, 1\}$ is the actual class of this voxel. Although the binary cross-entropy loss penalizes false positives and false negatives and encourages the model to produce more balanced predictions, it does not directly promote accurate and precise segmentations.

By combining the Dice coefficient and the binary cross-entropy loss, the resulting loss function encourages the model to produce accurate and precise segmentations while also penalizing false positives and false negatives.

$$L_{DICE} = \frac{2|G \cap S|}{|G| + |S|} \tag{1}$$

$$L_{bce} = \sum_i y_i \log o_i + (1 - y_i) \log (1 - o_i) \tag{2}$$

$$Loss = L_{DICE} + L_{bce} \tag{3}$$

V. Proposed Approach

By combining T1-weighted and FLAIR modalities, our approach addresses the limitations of T1-weighted MRIs, such as low contrast between various tissues and acute stroke lesions appearing hypointense/dark, and those of FLAIR MRIs, including false positives and low sensitivity to small lesions. Our method aims to yield a more accurate segmentation by leveraging the strengths of both modalities. Building upon the nnU-net architecture and one of its improvements specifically designed for combination of T1-weighted, T2-weighted, and FLAIR modalities, the LongiSeg4MS [14], we propose nnU-net for stroke segmentation from T1-weighted and FLAIR MRIs.

Initially, we train our model using the ATLAS v2.0 database, which provides a rich and diverse dataset for stroke lesion segmentation. This pretraining step allows us to obtain a model that has already learned the basic features relevant to the task. Next, we employ our custom algorithm[3] to reshape the pretrained model whilst doubling the number of its input channels. This enables our model to accept both T1 and FLAIR MRIs as inputs, optimizing it for multi-modal data. Importantly, the optimized model shape for fine-tuning is determined by nnU-net itself, ensuring that the model is not just a simple replication of the pretrained one with more input channels but is specifically tailored to work efficiently with T1 and FLAIR data.

Subsequently, we fine-tune our model using T1 and FLAIR MRIs from our in-house database, enabling the model to better adapt to the specific characteristics of our dataset, such as Multi-modality MRI, MRI acquisition parameters and stroke

[3]https://github.com/LounesMD/MMStrokeNet

segmentation practices. Furthermore, we conduct tests and experiments with datasets from different sites (9 different sites on ATLAS v2.0 and 1 in-house site), specifically curated to assess the performance of the proposed approach, while considering diverse populations (brain shapes, subjectivity to segment, stroke stages, genders, lesion types, etc). Our main contributions are as follows:

- We adapt a state-of-the-art model to generate stroke segmentation from T1-weighted and FLAIR MRIs, addressing the limitations of existing approaches and improving segmentation accuracy.
- We develop an algorithm that adapts a given nnU-net generated model with one input channel to a new one, which is specially and correctly designed to handle two different input modalities.
- Our experimental testbed demonstrates that the stroke segmentation generated by our model yields globally satisfactory results, highlighting the effectiveness of our approach in providing accurate and robust segmentations.

In conclusion, our goal is to develop a more accurate and robust segmentation tool for brain stroke lesions by harnessing the power of large datasets, utilizing 3D images, and capitalizing on the complementary strengths of different MRI modalities. By doing so, we strive to make a significant contribution to the field of medical imaging and improve the diagnosis and treatment of stroke patients.

VI. Evaluation

In this section, we assess the performance and reliability of our proposed model for stroke lesion segmentation, leveraging both T1 and FLAIR MRI modalities. We initially detail the pretraining process of our model, followed by the finetuning procedure. It is important to note that for both stages of model training, nnUnet employs a 80% training and 20% validation split of the dataset. As each patient only appears once in the database, this ensures a robust and leakage-free validation strategy, enabling us to effectively evaluate our model's capabilities in generating accurate stroke lesion segmentations.

A. Pretrained Model

In this subsection, we discuss the pretrained model used for finetuning and elaborate on our decision to train it for more than the recommended 1000 epochs with only one fold.

As demonstrated in Fig.2, we initially trained our model for 1000 epochs as recommended, but observed that further training could potentially yield a lower loss value (less than -0.6), albeit at a slower rate and over a greater number of epochs. Consequently, we trained a new model for 3000 epochs to achieve a better loss and a higher average DICE score, as illustrated in Fig.3.

However, we found that additional training did not yield improved accuracy, as evidenced by the clear plateau in the loss validation. Furthermore, we noticed that beyond 2500 epochs, even though the validation loss did not decrease, the training loss began to decrease. To avoid potential overfitting, we utilized the model trained over the first 2500 epochs. For these trainings, we employed the polynomial recommended learning rate by nnUnet, which is equal to $0.01 * (1 - \frac{epoch}{epoch_{max}})^{0.9}$. With this training, the model achieved an average DICE score of 0.633, which is very similar to the dice score of 0.650 reported by [14].

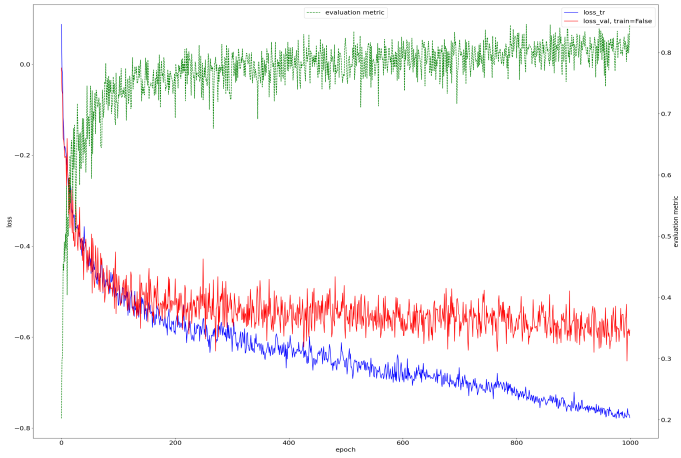

Fig. 2. nnU-net trained over 1000 epochs

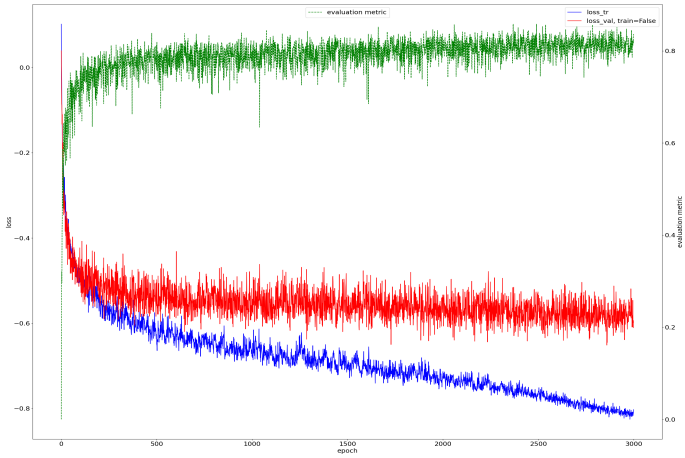

Fig. 3. nnU-net trained over 3000 epochs

Having obtained a reliable pretrained model with consistent results on the ATLAS v2.0 dataset, in the next section we will discuss the finetuning process and the results obtained from our model.

B. Finetuned Model

With the pretrained model ready for use, our objective is to finetune it to accept both T1-weighted and FLAIR scans as input for predicting segmentations. However, the pretrained model cannot be used directly since it has only one input channel for a single MRI modality. Furthermore, nnU-net is designed to build optimized models depending on the dataset. To finetune an optimized model using the weights of the

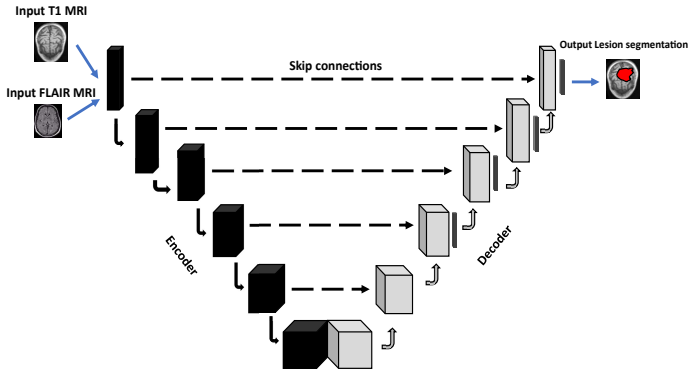

Fig. 4. Simplified custom nnU-net architecture

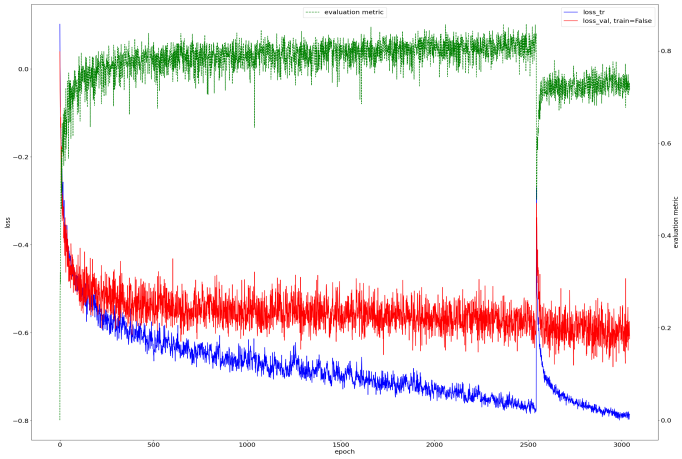

Fig. 5. Model pretrained over 2500 epochs and finetuned over 500 epochs

pretrained model, we follow these steps to create an optimized finetuned model for both T1 and FLAIR modalities:

1) Preprocess our in-house database (T1-weighted + FLAIR MRIs)
2) Train a new model using nnU-net/Longiseg4MS for only 1 epoch, generating a skeleton model with features optimized for our in-house database (i.e. optimised for T1+FLAIR modalities). This skeleton model will be used to reshape our pretrained model for optimization with the in-house database.
3) Utilize our custom algorithm[3] to adapt the shape of the first pretrained model to that of the model generated in step 2. This approach doubles the number of input channels and resizes the convolution layers as necessary.
4) Replace the skeleton model files of step 2 with the model generated in step 3.
5) Adjust the learning rate to 10^{-5} to adjusts its weights to the new model.
6) Add 500 more epochs of training and resume training the model.

Figure 4 illustrates the functionality of the model produced by our custom algorithm. It is similar to the model in Fig. 1, but with twice the number of input channels, allowing for the input of both FLAIR and T1-weighted images to produce a single segmentation. Figure 5 displays the progression of loss over time. The first 2500 epochs correspond to the loss of the pretrained model, while the last 500 represent the finetuned model[4]. As the finetuned model loss is not as poor as the initial loss of the pretrained model, it indicates that the pretraining provides consistent results even when using MRIs and segmentations from a different database. This is further evidenced by the quick convergence toward a loss comparable to the pretrained model with fewer epochs. However, it is important to note that this database is smaller and different from ATLAS v2.0. To observe and justify the improvement using the second model, it is necessary to test the first model (trained on ATLAS v2.0) on the in-house database validation data and compare the results. These results are presented in Table II and Fig. 6.

[4]https://github.com/LounesMD/MMStrokeNet

Table II provides a comprehensive comparison of the performance (DICE scores) of each model on each dataset.

Firstly, the pretrained model was evaluated on the in-house dataset, comparing the results with only the data used to evaluate the finetuned model (Pretrained Model 1) and the entire dataset (Pretrained Model 2). To emphasize the fact that segmenting FLAIR MRIs is a fundamentally different task than segmenting T1-weighted MRIs, we assessed the pretrained model on FLAIR MRIs. The average DICE scores were a mere 0.210 / 0.158 when using FLAIR MRIs, compared to 0.692 / 0.655 when using T1-weighted MRIs. This result clearly demonstrates the necessity of adapting the model to handle both modalities effectively.

Next, we compared the performance of the finetuned model using both T1-weighted and FLAIR MRIs. The average DICE score for the finetuned model was 0.707, outperforming the pretrained model's performance on both individual modalities. This superior performance highlights the effectiveness of our approach in leveraging the complementarity of T1-weighted and FLAIR MRIs, providing more accurate and robust segmentation.

By employing a cutting-edge approach that combines a pretrained model with a custom algorithm, we have successfully developed a model capable of handling both T1-weighted and FLAIR MRIs, delivering high performance in stroke lesion segmentation (mean DICE ≥ 0.7). The finetuned model's performance, demonstrated by its superior DICE score, positions it as an innovative solution with the potential to significantly impact the quality of stroke lesion segmentation.

The box plot statistics presented by the Fig. 6 offer a deeper insight into the performance of the finetuned model compared to the pretrained model 1. This analysis is important because it brings a better understanding of the performance difference of those 2 models tested with the same data.

- Minimum DICE score: The finetuned model (0.552) performs slightly worse than the pretrained model 1 (0.608). However, this difference is relatively small, considering the broader context of the other metrics. This can be

TABLE II
DICE SCORE SUMMARY

Dataset	Pretrained Model 1	Pretrained Model 2	Finetuned Model
ATLAS v2.0	0.633	0.633	x
In-house FLAIR	0.210	0.158	x
In-house T1	0.692	0.655	x
In-house T1+FLAIR	x	x	0.707

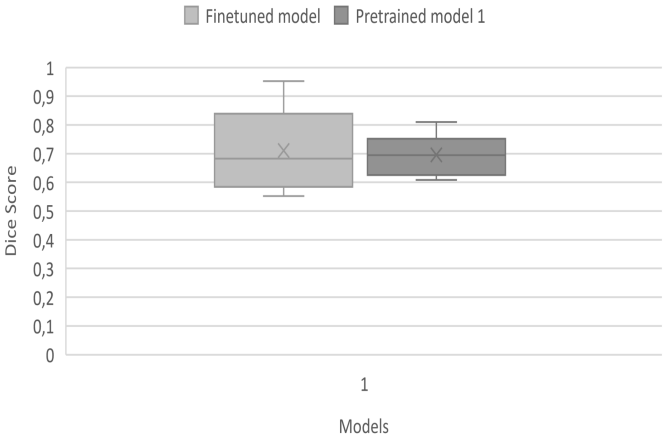

Fig. 6. Dice comparison

justified by the fact that if the T1-weighted MRI is clear enough and their is no need of the FLAIR to segment the lesions, adding the FLAIR modality possibly introduced noises.

- Quartile 1 (Q1): The finetuned model (0.5945) performs slightly worse than pretrained model 1 (0.6315), but the difference is low. As mentionned for the minimun DICE, it comes from the noise introduced by the FLAIR when the T1 is sufficient in itself.
- Quartile 2 (Q2, Median): The finetuned model (0.657) underperforms pretrained model 1 (0.698).
- Quartile 3 (Q3): The finetuned model (0.801) significantly surpasses pretrained model 1 (0.732), demonstrating the model's improved ability to produce accurate segmentation. And the reason is the opposite from the previous one. Here, because the T1 MRI is not high quality enough, using the FLAIR brings a lot of valuable information for detecting and segmenting the lesions.
- Quartile 4 (Q4, Maximum): The finetuned model (0.952) outperforms pretrained model 1 (0.810) by a substantial margin, showcasing the finetuned model's ability to deliver exceptional results.
- Average: The finetuned model (0.707) exceeds pretrained model 1 (0.692), further corroborating its superior performance.

In summary, these results can be explained by the observation that utilizing FLAIR MRI when the T1-weighted MRI is already sufficient can introduce noise during segmentation prediction, leading to poorer results in certain cases. However, for the majority of cases and on average, incorporating the FLAIR modality with an optimized finetuned model enhances the quality and accuracy of lesion segmentation.

VII. Discussion and Future Research Directions

Our research demonstrates the effectiveness of our custom algorithm and finetuned model for stroke lesion segmentation using T1-weighted and FLAIR MRI modalities. To further advance this field, we propose and explain the following future research directions:

A. Multimodal Integration

Incorporating further imaging modalities, like T2-weighted, diffusion-weighted, or susceptibility-weighted MRI, may enhance segmentation outcomes. Future studies should explore the advantages of multimodal fusion in stroke lesion segmentation and develop a method to determine the relevance of utilizing all modalities or just the T1-weighted MRI to predict the segmentation of a patient.

B. Larger and Diverse Datasets

Evaluating our finetuned model on larger, more diverse datasets will help assess its generalizability and robustness across different clinical contexts and patient populations.

C. Automated Segmentation Refinement with RLHF

To enhance the accuracy of our stroke lesion segmentation model, we propose incorporating Reinforcement Learning from Human Feedback (RLHF). This approach uses expert feedback from experts to guide model learning, ensuring predictions align with real-world clinical standards. By iterating through feedback and model adjustments, our model can continually improve its segmentation capabilities. Integrating RLHF enables a more accurate and reliable tool for diagnosing and treating stroke patients.

D. Longitudinal Analysis

Applying our model to longitudinal data could provide valuable insights into stroke lesion progression and inform personalized treatment strategies for stroke patients.

E. Two-stage Finetuning

To maximize applicability in real-world settings, future research could involve a three-stage finetuning process. First, train the model using the ATLAS dataset, followed by finetuning the model using only T1-weighted MRI scans from the in-house database to capture the specific segmentation techniques employed by the experts in that database. Finally, leverage the learned features from the in-house database to further finetune the model using both T1-weighted and

FLAIR scans when available, thus enhancing its overall performance.

VIII. Conclusion

In this research paper, we have presented a novel approach to stroke lesion segmentation, leveraging both T1-weighted and FLAIR MRI modalities and an optimized neural network architecture. Our main contributions lie in the development of a custom algorithm that adapts a pretrained model to accommodate both modalities, resulting in a finetuned model capable of delivering more accurate and robust segmentation, and the idea of pretraining a model on the specific task of stroke lesion segmentation using a massive dataset, followed by finetuning with T1-weighted and FLAIR MRI data.

Our results demonstrate the effectiveness of combining T1-weighted and FLAIR MRI modalities for improved segmentation. The finetuned model outperformed the pretrained model on several key metrics, including the average DICE score, showcasing its superior performance in stroke lesion segmentation. In cases where T1-weighted MRI quality was insufficient, the addition of the FLAIR modality provided valuable information for lesion detection and segmentation, leading to enhanced accuracy.

Furthermore, the use of a pretrained model and finetuning techniques allowed for faster convergence and more consistent results when working with different databases, highlighting the potential of our approach in various clinical settings. Our finetuned model not only reduces the time and effort required for experts to manually segment lesions but also minimizes the risk of human error, ultimately contributing to better treatment decisions and patient outcomes.

In conclusion, our research demonstrates the potential of using multimodal MRI data and advanced DL techniques to revolutionize stroke lesion segmentation. The development and application of our custom algorithm, combined with the innovative idea of pretraining and finetuning, have shown promising results, emphasizing the value of combining T1-weighted and FLAIR modalities in an optimized neural network. As the field of artificial intelligence continues to evolve, we believe that our approach can serve as a foundation for future advancements in stroke lesion detection and segmentation, ultimately leading to improved patient care and outcomes.

References

[1] GBD 2019 Stroke Collaborators. Global, regional, and national burden of stroke and its risk factors, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet Neurol. 2021 Oct;20(10):795-820. doi: 10.1016/S1474-4422(21)00252-0. Epub 2021 Sep 3. PMID: 34487721; PMCID: PMC8443449.

[2] Chauhdry H. Understanding the importance of recognising, treating and preventing stroke. Nurs Stand. 2022 Jan 5;37(1):77-82. doi: 10.7748/ns.2021.e11596. Epub 2021 Oct 11. PMID: 34632749.

[3] O. Ronneberger et al. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W., Frangi, A. (eds) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science(), vol 9351. Springer, Cham. https://doi.org/10.1007/978-3-319-24574-4_28

[4] X. Yin et al. "U-Net-Based Medical Image Segmentation". J Healthc Eng. 2022 Apr 15;2022:4189781. doi: 10.1155/2022/4189781. PMID: 35463660; PMCID: PMC9033381.

[5] Zhou Z et al. UNet++: A Nested U-Net Architecture for Medical Image Segmentation. Deep Learn Med Image Anal Multimodal Learn Clin Decis Support (2018). 2018 Sep;11045:3-11. doi: 10.1007/978-3-030-00889-5_1. Epub 2018 Sep 20. PMID: 32613207; PMCID: PMC7329239.

[6] Q. Zuo et al. R2AU-Net: Attention Recurrent Residual Convolutional Neural Network for Multimodal Medical Image Segmentation, Security and Communication Networks 2021.

[7] F. Milletari et al. "V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation," 2016 Fourth International Conference on 3D Vision (3DV), Stanford, CA, USA, 2016, pp. 565-571, doi: 10.1109/3DV.2016.79.

[8] H. Shin et al. Automated segmentation of chronic stroke lesion using efficient U-Net architecture, Biocybernetics and Biomedical Engineering, Volume 42, Issue 1, 2022.

[9] O. Oktey et al. "Attention U-Net: Learning Where to Look for the Pancreas". 2018

[10] S. Manjin et al. "Cross-Attention and Deep Supervision UNet for Lesion Segmentation of Chronic Stroke". Frontiers in Neuroscience 2022.

[11] A. Zhao et al., Data augmentation using learned transformations for one-shot medical image segmentation, Computer Vision and Pattern Recognition Conference, 2019.

[12] Li et al., "H-DenseUNet: Hybrid Densely Connected UNet for Liver and Tumor Segmentation From CT Volumes," in IEEE Transactions on Medical Imaging, vol. 37, no. 12, pp. 2663-2674, Dec. 2018, doi: 10.1109/TMI.2018.2845918.

[13] W. Yu et al. SAN-Net: Learning generalization to unseen sites for stroke lesion segmentation with self-adaptive normalization, Computers in Biology and Medicine, Volume 156, 2023.

[14] K. Verma et al. Automatic Segmentation and Quantitative Assessment of Stroke Lesions on MR Images. Diagnostics (Basel). 2022 Aug 24;12(9):2055. doi: 10.3390/diagnostics12092055. PMID: 36140457; PMCID: PMC9497525.

[15] K. Verma et al. Automatic Segmentation of Stroke Lesions on MR Images for Voxel-Based Lesion-Symptom Mapping. In Proceedings of the Annual Meeting of the Society for Neuroscience, Society for Neuroscience, Virtual, 8 November 2021.

[16] SL. Liew et al. A large, curated, open-source stroke neuroimaging dataset to improve lesion segmentation algorithms. Sci Data 9, 320 (2022). https://doi.org/10.1038/s41597-022-01401-7.

[17] SL. Liew et al. A large, open source dataset of stroke anatomical brain images and manual lesion segmentations. Sci Data 5, 180011 (2018). https://doi.org/10.1038/sdata.2018.11.

[18] V. Rajinikanth et al. Deep Learning Framework to Detect Ischemic Stroke Lesion in Brain MRI Slices of Flair/DW/T1 Modalities. Symmetry 2021, 13, 2080. https://doi.org/10.3390/sym13112080.

[19] R.A. Zeineldin et al. DeepSeg: deep neural network framework for automatic brain tumor segmentation using magnetic resonance FLAIR images. Int J CARS 15, 909–920 (2020). https://doi.org/10.1007/s11548-020-02186-z.

[20] F. Isensee al. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat Methods 18, 203–211 (2021). https://doi.org/10.1038/s41592-020-01008-z.

[21] A. Masson et al. A nnUnet implementation of new lesions segmentation from serial FLAIR images of MS patients, MSSEG-2, MICCAI 2021.

[22] N.J. Tustison et al.: N4itk: improved n3 bias correction with robust b spline approximation. IEEE Trans Med Imaging 29(6), 708–711 (2010).