SEMARAK ILMU PUBLISHING
202103268166(003316878-P)

Journal of Advanced Research in Applied Sciences and Engineering Technology

Journal homepage:
https://semarakilmu.com.my/journals/index.php/applied_sciences_eng_tech/index
ISSN: 2462-1943

NIVE: NeuroImaging Volumetric Extractor, a High-Performance Skull-Stripping Tool

Khuhed Memon[1], Norashikin Yahya[1,*], Shahabuddin Siddiqui[2], Hilwati Hashim[3], Mohd Zuki Yusoff[1], Syed Saad Azhar Ali[4]

[1] Centre for Intelligent Signal and Imaging Research (HICoE-CISIR), Department of Electrical and Electronic Engineering, Universiti Teknologi PETRONAS, Malaysia, 32610 Seri Iskandar, Perak, Malaysia
[2] Department of Radiology, Pakistan Institute of Medical Sciences (PIMS) Islamabad. Advanced Diagnostic Center, Advanced International Hospital Islamabad, Pakistan
[3] Department of Radiology, Faculty of Medicine, Universiti Teknologi MARA, 47000 Sungai Buloh, Selangor, Malaysia
[4] Aerospace Engineering Department & Center for Smart Mobility and Logistics, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia

ARTICLE INFO

Article history:
Received 13 December 2023
Received in revised form 22 May 2024
Accepted 10 June 2024
Available online 20 August 2024

Keywords:

Brain extraction; deep learning; NIVE; skull stripping; DeepLabV3+

ABSTRACT

Brain extraction is an important preprocessing step in Computer Aided Diagnosis (CAD) from brain MRI. It facilitates stripping off irrelevant extra-cranial tissues including skull, eyes, and neck muscles, thereby enhancing accuracy of inference of an AI CAD system. Numerous studies have been conducted to achieve this task ranging from old-school image processing techniques to more sophisticated deep learning approaches. This research investigates the performance of two deep learning-based semantic segmentation techniques, DeepLabV3+ and U-Net for brain extraction of brain MRI scanning images. The two networks were trained using the largest human brain MRI images dataset having more than 160 thousand training images and the performance is compared with SynthStrip from MIT, as the current state of the art system. Besides, the trained networks were also tested using newly collected MRI images from a private hospital, AIH, Islamabad. The result indicates that DeepLabV3+ outperforms SynthStrip and U-Net in all public datasets and produces comparable results from the new dataset with a mean Dice score of 0.98. Using the trained DeepLabV3+, an app called NIVE is developed for public use. Since the DeepLabV3+ was trained using the most comprehensive human brain MRI dataset to date, NIVE is essentially the most versatile brain extraction app capable of handling MRI images in all types of file formats, sequences, orientations, acquisition hardware, and subject age.

1. Introduction

Skull stripping, the process of extracting brain region from magnetic resonance imaging (MRI) scanning images, is the foremost step of any CAD regimen [1-5]. Its accuracy is of utmost significance for all subsequent operations leading to diagnosis [6-11]. Optimum pathology detection is only

* *Corresponding author.*
E-mail address: norashikin_yahya@utp.edu.my

https://doi.org/10.37934/araset.50.2.228245

possible after this process [11] as the presence of the skull can cause misjudgements, especially in cases of brain lesions [12] and tumours [6,13]. Manual segmentation by experts (radiologists/ neurologists) is considered as the “gold standard”, but it is a very tedious and time-consuming job, given the size of massive MRI datasets [1,14]. In order for the CAD system to be fast, accurate and human intervention-free, a robust skull stripping tool is mandatory. Extensive research has been carried out in this direction resulting in development of algorithms ranging from old-school image processing and morphological operations [15] to the more recent artificial intelligence (AI) based systems [16].

Literature suggests that deep learning (DL) based approaches outperform conventional methods to carry out this task [2,17]. Deep Convolutional Neural Networks (CNN) have exhibited outstanding performance [18], with U-Net architecture being widely used [1,19]. Among the studies conducted and systems developed so far, few shortcomings have been observed periodically. Majority of the systems work with only T1W MRI [7,10,16,20] for healthy adult brains [6], rendering them unreliable for T2W, fluid attenuated inversion recovery (FLAIR) and other sequences and their performance drops in the presence of pathologies [16]. Rehman *et al.,* [16] suggest that fast, accurate, user-friendly, sequence orientation and pathology-agnostic skull stripping tools would be an absolute requirement in times to come.

Addressing this requirement, we propose a software tool, known as NeuroImaging Volumetric Extractor (NIVE) for removing the skull from brain MRI. Notably, NIVE has been trained with the most comprehensive human brain MRI dataset encompassing normal/pathological and adult/infant brains with T1, T2, FLAIR and proton density (PD) sequences in axial, sagittal and coronal orientations. In addition, NIVE comes with a vast input file type support and a very simple user interface (UI) to assist radiologists and neurologists who might not be comfortable with command line operations. The performance of the proposed system has been compared with SynthStrip [10], the current state of the art system, in the light of the literature.

The organization of this paper is such that prior art in this field is given in Section 2, Section 3 presents the hardware, software, brain MRI dataset and DL architectures used, whereas the results are presented in section 4. Section 5 and 6 describe the availability of NIVE and the online resources, and the paper is concluded in Section 7.

2. Prior Art

In this section we briefly review the research pertaining to skull stripping, with focus on SynthStrip [10] which we have used as the benchmark for performance comparison. A summary of the developed algorithms along with their performance metrics and datasets used is given in Table 1. Many algorithms have been used in developing skull strippers, but DL based approaches stay dominant, with adequate performance for medical image segmentation applications. The use of limited datasets for training, relying on a single sequence, T1W predominantly and catering only normal adult human brain MR images appear as the major shortcomings in most of the research. This results in lower performance accuracy for other sequences and for pathological brains, rendering such systems incapable of acting as robust assistive tools.

Very few systems [10] cater infant MRI due to their small size and dynamic intensity changes [19]. Salehi *et al.,* [21] cater fetal brain MRI. The research in [22] uses morphological operations by extracting the largest connected component and it sometimes fails, in which case they use information from adjacent slices in an MRI volume. This approach is not suitable for individual MRI slices. The accuracy in [23] drops in regions around the eyes and below. If extra-cranial content is retained, it might hamper optimum pathology detection [11] at later stages especially in cases of brain tumours or lesions [6]. Kaliyugarasan *et al.,* [24] suggest that similar performance accuracy is observed using both 2D and 3D U-Nets, with 2D being slower due to slice-by-slice processing approach. Others state that in order to improve efficiency, 3D MRI must be broken down into 2D sequences and stripped layer by layer [1]. Hence this approach is adopted by NIVE.

Regarding "Gold Standard" ground truth labelling by experts, Lucena *et al.,* in [4] suggest that single-rater based manual annotation may be biased and hence a consensus approach among multiple raters may be used. In this research, the gold standard ground truths for testing have been obtained from one consultant radiologist only, which stays as a limitation at the moment, and may be addressed later. SynthStrip [10] considered as the benchmark tool developed in 2022 by MIT uses U-Net as the baseline architecture and outperforms multiple skull strippers including ROBEX [7], BET [25], 3DSS, BEaST [9], FSW and DMBE. NIVE developed in this research is compared in performance with SynthStrip using online datasets and those acquired from a hospital. The materials and methods used in the development of NIVE are presented in the subsequent sections.

Table 1
Related works on Brain Extraction summarized in terms of techniques, performance; Jaccard index and Dice score and MRI dataset

Paper	Technique	Performance	Dataset
[26]	Watershed algorithms and deformable surface models	Jaccard up to 0.885	T1W MRI
[5]	Meta Algorithm, Brain Surface Extractor (BSE) from BrainSuite, the Brain Extraction Tool (BET) from FSL, 3DIntracranial from AFNI, and MRI Watershed from FreeSurfer	Dice up to 0.980 ± 0.00374	275 subjects - BEMA trained on 25 scans from the ICBM data set, 27 from the IPDH, 48 from LIJMC, 30 from ZENIT, and 10 from NEUROVIA
[27]	2D/ 3D brain extraction algorithms (BEA). LPF and morphological operations to find largest connected component (LCC)	--	20 normal/ abnormal MRI - T2W
[28]	Morphological operations- double and Otsu's thresholding	Up to 96.67% acceptance	90 T1W, T2W and FLAIR MRI
[29]	Region labelling and morphological operations	Dice 0.938	61 scans - IBSR, KGS Advanced MR and CT scans, Madurai, Tamil Nadu, India, T1W only
[8]	Comparison - BET, BSE, HWA and MAPS	Jaccard up to 0.98	ADNI - 682 1.5 T and 157 3 T T1W MRI
[7]	ROBEX - discriminative model (Random Forest classifier), the generative model (point distribution model)	Dice up to 96.6% ± 0.3	Trained using 92 scans from a proprietary dataset – tested on IBSR, LPBA40, and OASIS, 137 scans in total
[3]	Simplex Mesh and Histogram Analysis Skull Stripping (SMHASS)	Dice up to 0.972	20 simulated T1W MRI images from the BrainWeb website, 18 real T1W MRI images from the IBSR, 40 real T1W MRI images from the SVE
[9]	BEaST - nonlocal segmentation embedded in a multi-resolution framework	Dice 0.9781 ± 0.0047	NIH Paediatric Database (NIHPD), ICBM, ADNI
[11]	Adaptive intensity thresholding followed by morphological operations	Dice up to 0.99	Simulated MR images obtained from "BrainWeb: Simulated Brain Database" and real MR images of IBSR, T1W only
[30]	3D-CNN	Dice 0.9519	IBSR, LPBA40 and OASIS, totalling 135 volumes

[31]	MONSTR - registration and patch matching	Dice up to 0.9833	6 datasets, ADNI-29 normal, NAMIC-10 normal controls and 10 patients with schizophrenia, MRBrainS-5 normal controls
[21]	Auto-context CNN, U-Net	Dice up to 97.73%	LPBA40, 40 T1W MRI scans of healthy subjects, OASI) 70 T1W MRI scans of healthy subjects
[14]	8 publicly available skull stripping techniques	Dice up to 97.9637 ± 0.631	Calgary-Campinas-359 (CC-359)
[1]	Multi-view pyramid skull stripping network (PSSNet)	Dice 95.44%–97.33%	70 neonates from the local hospital and 7 infants from the publicly available dataset NeuroBrainS12 (MICCAI 2012). 51 and 26 subjects used for training and validation, T1W only
[4]	2D U-Net, CONSNet	Dice up to 97.353 ± 0.003	Calgary-Campinas-359 (CC-359), LPBA40, OASIS
[18]	3D U-Net	Mean Dice 0.9903	NFBS, T1W only
[19]	Flattened Residual Network	Dice 0.986	343 subjects, covering newborns to 48-month-olds. UNC/UMN Baby Connectome Project (BCP) dataset, T1W sagittal only
[32]	HD-BET, U-Net	Dice up to 98%	LPBA40, NFBS, Calgary Campinas-359 [CC-359]
[2]	2D U-Net, 3D U-Net, FCN, DeepMedic, 3D-ResUNet	Dice up to 0.98	3340 mpMRI brain tumour scans, private and public, public including TCGA-GBM, n = 328, TCGA-LGG, n = 372 BraTS challenge
[17]	Dense-VNet	Dice score of 94.5% for tumour brains, 96.2% for healthy brains	70,000 serial structural MR studies of 2,500+ unique brain tumour patients acquired across 20+ institutions, T1Gd and FLAIR
[24]	2D and 3D U-Net	2D U-Net Dice 0.9778, 3D U-Net Dice 0.9781	ADNI, AIBL, IXI, PPMI, SLIM, Calgary-Campinas and SALD, T1W
[22]	Helmholtz free energy principle and morphological operations	Dice up to 0.968	38 MRI from IBSR, T1W
[33]	PARIETAL, modified U-Net	Dice 97.2%	21 subjects from hospitals, T1W. Calgary-Campinas dataset 359 comprising MR images of healthy adults
[10]	SynthStrip – 3D U-Net	Dice up to 97.8 ± 0.3	80 training subjects from three cohorts: 40 adult subjects from the Buckner40 dataset, 30 locally scanned adult subjects from HCP-A, and 10 infant subjects born full-term, scanned at Boston Children's Hospital at ages between 0 and 18 months
[23]	k-strip based on U-Net	Dice scores of 92%-98%	36,900 MRI slices, University Hospital Essen, consisting of 30000 T1 brain MRI 2D-slices from 207 patients. The other dataset is obtained from the NYU fastMRI initiative database (fastmri.med.nyu.edu), containing 6900 fully sampled brain MRI scans
[34]	Largest connected component extraction using morphological ops	Dice up to 80%	OASIS, T1W
[6]	Ensemble neural network (EnNet), 3DCNN	Dice 0.9850 ± 0.0171	815 cases with or without glioblastoma multiforme (GBM) at the University of Pittsburgh Medical Centre (UPMC) and TCIA
[35]	Hyperconnectivity and viscous lattices	Dice up to 0.951	38 MR images obtained from IBSR, T1W

3. Material and Method

This section states the hardware and software used in the development of NIVE. It also provides details about the datasets used for training, validation and testing, encompassing both online sources and the data acquired from a hospital in Islamabad, Pakistan. The end-to-end design and development methodology is also provided in this section.

3.1 Hardware and Software

Firstly, the training and testing of the NIVE was conducted on multiple computers. This includes Lenovo Legion Y545 using Intel(R) Core(TM) i7-9750H CPU @ 2.60GHz 2.59GHz with 16GB RAM and 6GB NVIDIA GeForce GTX 1660 Ti. Another Linux based machine with 12GB NVIDIA GeForce RTX 2080 Ti at the Center for Intelligent Signal and Imaging Research (CISIR), Universiti Teknologi Petronas (UTP) was also used for training. MATLAB R2022b and Python 3.10.7 were used for coding. TensorFlow 2.10.0, TensorFlow-GPU 2.10.0 with Cuda 11.2.2 and cudnn 8.1.1 were used for training and testing, in addition to MATLAB.

Secondly, for testing SynthStrip, a Windows Subsystem for Linux (WSL) and Ubuntu version 22.04.1 LTS were used. The SynthStrip is embedded in FreeSurfer version freesurfer ubuntu22-7.3.2 amd64.deb. In addition, DICOM to NIfTI conversion, DICOM and NIfTI tools, NIfTI visualization (version 2023.03.16) by Xiangrui Li (2023) was used to convert Dicoms to NIfTI files (to be fed as an input to SynthStrip).

Lastly, for the ground truth labelling of newly acquired images from a hospital, the Label Studio (version 1.7.3) was used with python. This was done by a consultant radiologist via remote access of the software using AnyDesk version 7.1.11. Besides, RadiAnt Dicom Viewer (version 2023.1) was also used to inspect Dicoms acquired from Advanced International Hospital, Islamabad, Pakistan.

3.2 MRI Brain Dataset

The list of 5 different MRI datasets utilized in the development and evaluation of the NIVE are summarized in Table 2. The training images are from three sources; NFBS, SynthStrip and MICCAI 2016 while the testing images come from all five datasets. Detailed description of the datasets is covered in the subsequent sections.

Table 2
Summary of the five brain MRI datasets utilized in the development and evaluation of NIVE

Dataset	Subject Type		Type of MRI				Condition	Mask		No. of images or patients	No. of slices	Utilization for NIVE	
	Adult	Infant	FLAIR	T1-W	T2-W	PD		Lesion	Brain			Training + Validation	Testing
NFBS	✓	-	-	✓	-	-	Multiple	-	✓	125 volumetric data	6000	✓	✓
SynthStrip	✓	✓	✓	✓	✓	✓	Not specified	-	✓	582 volumetric data	118606	✓	✓
MICCAI 2016	✓	-	✓	✓	✓	-	Multiple Sclerosis	✓	✓	53-patients	34082	✓	✓
Baghdad	✓	-	✓	✓	✓	-	Multiple Sclerosis	✓	-	60-patients	12 / sequence	-	✓
AIH Islamabad	✓	-	✓	✓	✓	-	Multiple	-	✓	3-patients	Varies	-	✓

3.2.1 Neurofeedback skull-stripped (NFBS) repository

The NFBS dataset [36] contains 125 T1-weighted anatomical brain MRI NIfTI volumetric data, along with brain masks and segmented brain images. Brain slices were extracted from all three orientations. 8 slices from deep brain were extracted from axial orientation which had a total of 192 slices with a resolution of (240 × 320)-pixel, making a total of 1000 MRIs and 1000 masks. Similarly, out of 256 coronal slices with (320 × 240)-pixel resolution, 8 deep brain slices per subject were extracted, and the same from 192 sagittal slices per subject with (320 × 320)-pixel resolution. This constituted a total of 3000 MRIs and their corresponding masks, later to be used for training and validation. Data cleaning and augmentation was carried out using MATLAB with removal of border artifacts in both MRI and mask images. In addition, all images were rotated by 90 degrees to increase the training data along with providing the DL network an ability to handle different MRI orientations. This was followed by zero-padding to achieve uniform image resolution of (256 × 256)-pixel. A total of 6000 MRIs and masks were eventually used from this dataset.

3.2.2 The SynthStrip dataset

The SynthStrip dataset [10] is a publicly available collection of MR images with corresponding ground-truth brain masks from 622 MRI, CT, and PET scans. This diverse data collection includes images acquired with various MRI sequences, resolutions, and in subject populations ranging from infants to patients with glioblastoma. Since our study deals with MRI modality only, 20 CT and 20 PET scans available in this dataset were not used. In total, 118606 T1, T2, FLAIR, PD and infant MRI slices from various orientations were extracted and retained from NIfTI volumes after discarding 131 slices due to presence of noisy artifacts or too dark/ blank images. A sample set of the images discarded is given in Figure 1. The dataset was also cleaned for border artifacts and resized to (256 × 256)-pixel for uniformity.

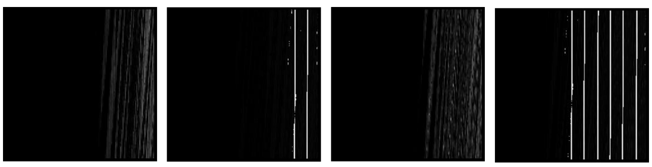

Fig. 1. Discarded noisy images from SynthStrip dataset

3.2.3 MICCAI 2016 challenge dataset

The MICCAI 2016 [37] database of images is composed of 53 Multiple Sclerosis patients. The dataset contains T1, T2 and FLAIR images among others, both raw and pre-processed versions. It also contains whole brain masks and lesion masks, but this dataset is only used for skull stripping. Only raw FLAIR images were considered. T1-weighted images were not considered to avoid dataset bias since NFBS had all T1-weighted images previously acquired. In addition, T2-weighted MR images and their corresponding brain masks had resolution mismatch issues, hence they were also discarded. A total of 34082 axial, sagittal and coronal MR slices and their corresponding non-zero masks were extracted from this dataset to be used for training and testing the skull stripping model. These images were also checked for border artifacts and resized to a resolution of (256 × 256)-pixel. Out of the 102246 images, 68164 were discarded due to resolution issues. Including brain MR images of MS patients for training would empower the system to effectively extract the brain from pathological scans with lesions in addition to healthy brain MRI.

3.2.4 Baghdad teaching hospital dataset for multiple sclerosis [*38*]

This dataset contains NIfTI volumes of FLAIR, T1 and T2-weighted MRIs and their corresponding lesion masks from 60 patients with Multiple Sclerosis. In this research we have used all three sequences from patient 1 to compare the performance of SynthStrip and NIVE.

3.2.5 Advanced International Hospital (AIH) Islamabad dataset

MRI scans from three subjects were acquired from Advanced Diagnostic Centre, Advanced International Hospital Islamabad. This dataset consists of Dicom sequences and was subsequently converted to NIfTI volumes to be fed into SynthStrip for performance analysis. The details of the subjects are given in Table 3. The ground truth brain masks were labelled by a consultant radiologist using Label Studio GUI in python. This dataset was later used for testing, performance evaluation and comparison of SynthStrip and NIVE using Dice similarity coefficient as the metric.

Table 3
Detailed information on the MRI from Advanced International Hospital (AIH) Islamabad

Subject	Gender	Age	Orientation	Sequence	Slices
1	Male	27	Axial	T1W	19
Clinical Information: Diagnosed case of HIV. Presented with fits, desaturation and hypotension					
2	Female	26.2	Sagittal	T2W	20
Clinical Information: Seizure disorder					
3	Male	36.1	Coronal	FLAIR	35
Clinical Information: Fits					

3.3 Brain Extraction Technique

Skull stripping was performed using SynthStrip from FreeSurfer and our custom trained models, and the results were compared. SynthStrip embedded in FreeSurfer version freesurfer ubuntu22-7.3.2 amd64.deb was installed on Windows 11 OS using Windows Subsystem for Linux (WSL) and Ubuntu version 22.04.1 LTS.

For the custom trained models, two separate skull stripping models were trained, one using DeepLabV3+ architecture and the other one using U-Net, with a 90-10 split for training and validation.

MATLAB Computer Vision Toolbox and Deep Learning Toolbox were used to achieve this. GPU always proves to be a lifesaver in such operations since it drastically reduces the training time depending on the hardware used. DeepLabV3+ uses Atrous Spatial Pyramid Pooling (dilated convolutions) and transposed convolutions for down-sampling and up-sampling respectively, whereas U-Net uses 2D Max-pooling and 2D-transposed convolutions. The U-Net model has 58 layers with a total of 31M learnable, the learning rate was set to 0.001. On the other hand, the DeepLabV3+ model has a total of 43.9M learnable, 206 layers, input layer of 256 × 256 and pixel classification as its output.

The learning rate was set at 0.01 and the ResNet50 backbone was used. Both the networks were trained using 158688 images. This dataset was a merger of 3 publicly available datasets and is the most diverse and massive dataset used to train a skull stripper so far, as evident from the literature. The models were trained with a minibatch size of 8 for 2 epochs, and validated with 10% of the dataset. In addition, testing was conducted using 2 subjects from each of the three online datasets, along with the three cases acquired from hospital with the ground truth data provided by a consultant radiologist. The system design and development methodology for NIVE is given in Figure 2.

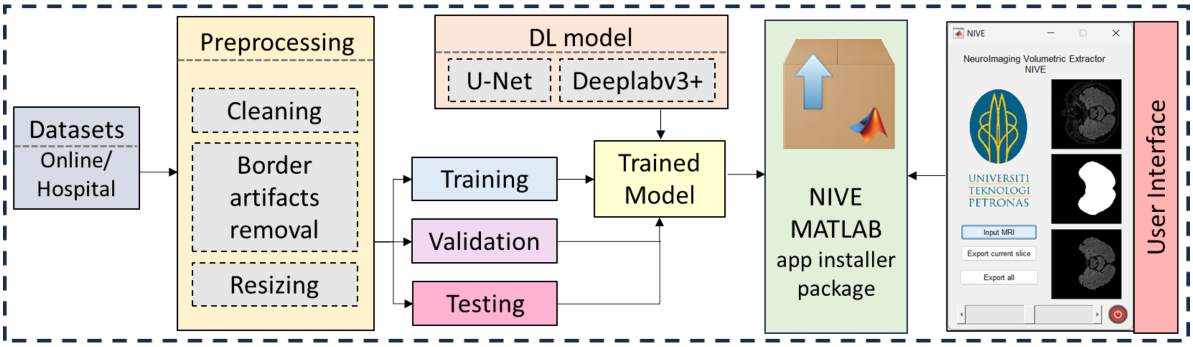

Fig. 2. Design and development flow of NIVE

3.4 Performance Metric

Since this is a semantic segmentation problem, in which each pixel is assigned a class, pixel accuracy can prove to be misleading if there exists a class imbalance. In that case, the Dice similarity coefficient can prove to be a meaningful performance metric. The performance of the segmentation method is measured using the normalized confusion matrix and Dice similarity coefficient (Dice). For the confusion matrix, TP, TN, FP and FN represent the true positive, true negative, false positive, and false negative, respectively. The Dice computes the similarity of elements between predicted output, *prediction* and original label, *target* sets,

$$Dice = 2\frac{|target \cap prediction|}{|target|+|prediction|} \times 100 \tag{1}$$

where $|target|$ represents the cardinal of the set target. Dice is similar to BF (Boundary F1) contour matching score between the predicted segmentation and the true segmentation (Ground Truth). Intersection over Union (IoU) is also used to evaluate performance of object detection systems, as shown in Eq. (2) and Figure 3. The performance comparison results are provided in the following section.

$$IoU = \frac{Area\ of\ Overlap}{Area\ of\ Union} \tag{2}$$

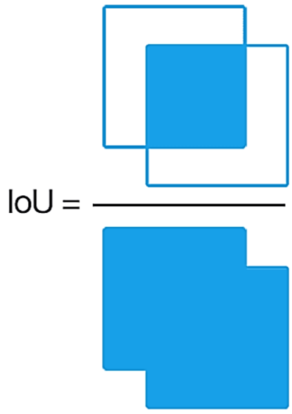

Fig. 3. Illustration on the area of overlap and area of union for computation of IoU

4. Result and Discussion

4.1 Training of U-Net and DeepLabV3+

The training and validation process for U-Net and DeepLabV3+ is given in Figures 4 and 5 respectively. The models were trained with a training/validation split of 90/10. The models were trained for a maximum of 2 epochs and each one took a little over 32 hours using NVIDIA GeForce RTX 2080 Ti.

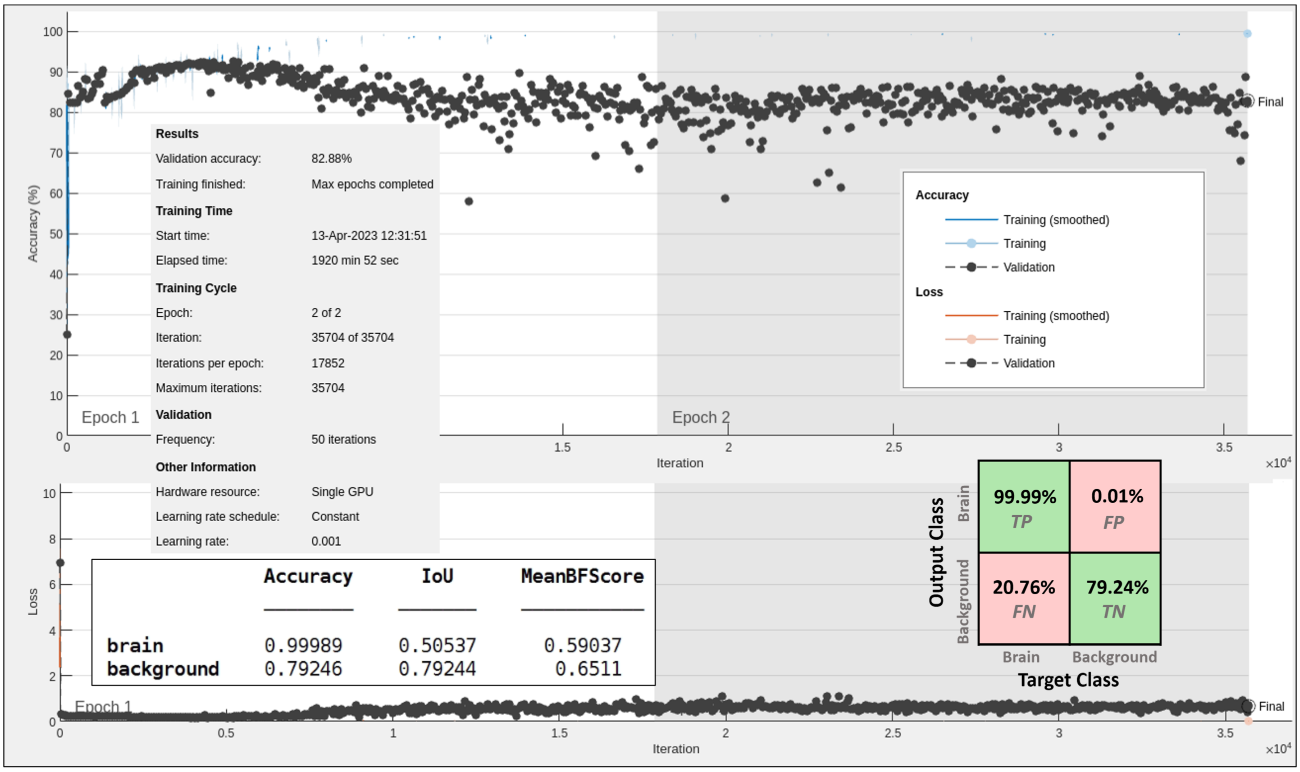

Fig. 4. Training progress, training details (top-right inset) of U-Net, performance evaluation in terms of accuracy, IoU and MeanBFScore (lower-right inset) and normalized confusion matrix (left inset)

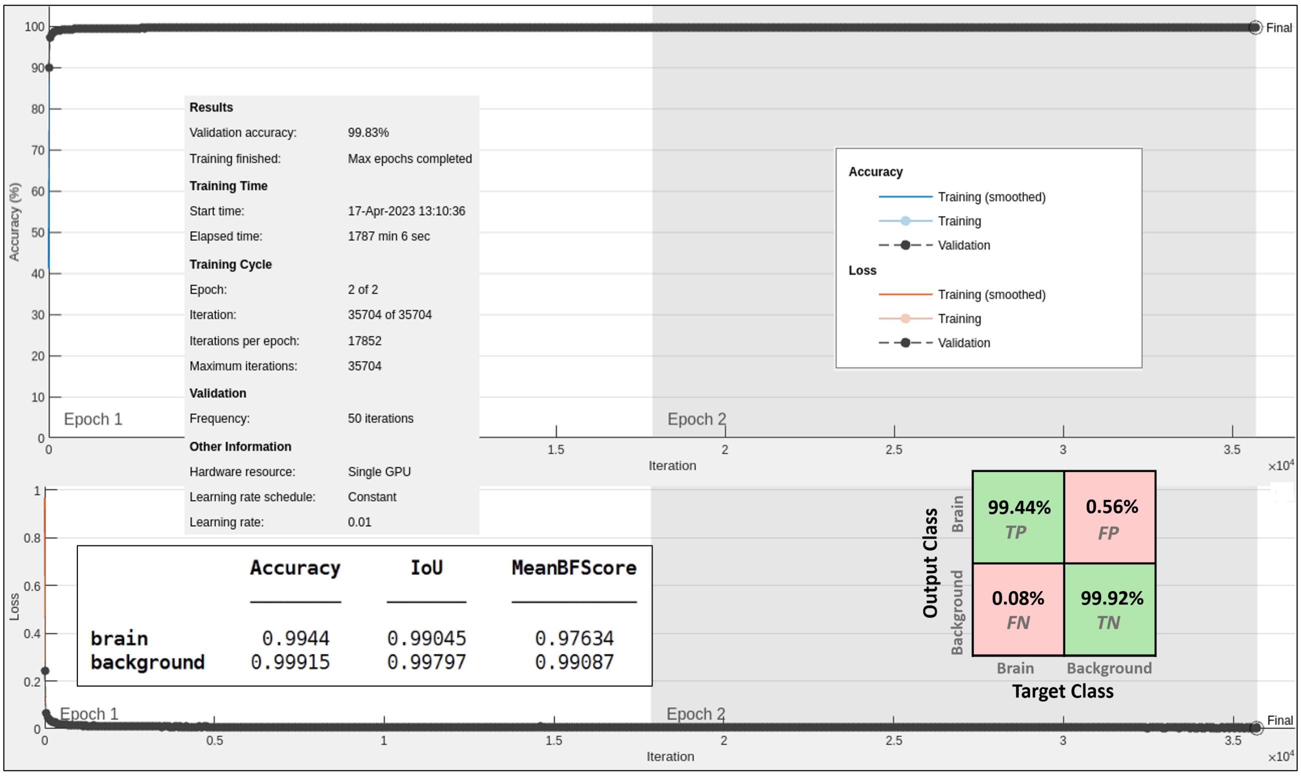

Fig. 5. Training progress, training details (top-right inset) of DeepLabV3+, performance evaluation in terms of accuracy, IoU and MeanBFScore (lower-right inset) and normalized confusion matrix (left inset)

The performance of models and the normalized confusion matrices are shown in Figures 4 and Figure 5. It can be observed that DeepLabV3+ outperforms U-Net, since U-Net falsely categorized a lot of non-brain regions as brain. Further testing of the two models trained in this research and SynthStrip was conducted using online datasets and real-time data acquired from a hospital. The real-time data was labelled by consultant radiologist using RadiAnt and Label Studio plugin for python, and the gold standard ground truth brain masks were secured.

4.2 Evaluation of Skull Stripping for Different Type of MRI Sequence – Qualitative Analysis

SynthStrip results for T1, T2 and FLAIR MRI respectively, are depicted in Figure 6, along with results from NIVE. These images have been acquired from the Brain MRI dataset of Multiple Sclerosis with consensus manual lesion segmentation and patient meta information, from Baghdad Teaching Hospital [38]. 12 slices have been chosen from a 19-slice 256 × 256 NIfTI volume in three sequences and axial orientation.

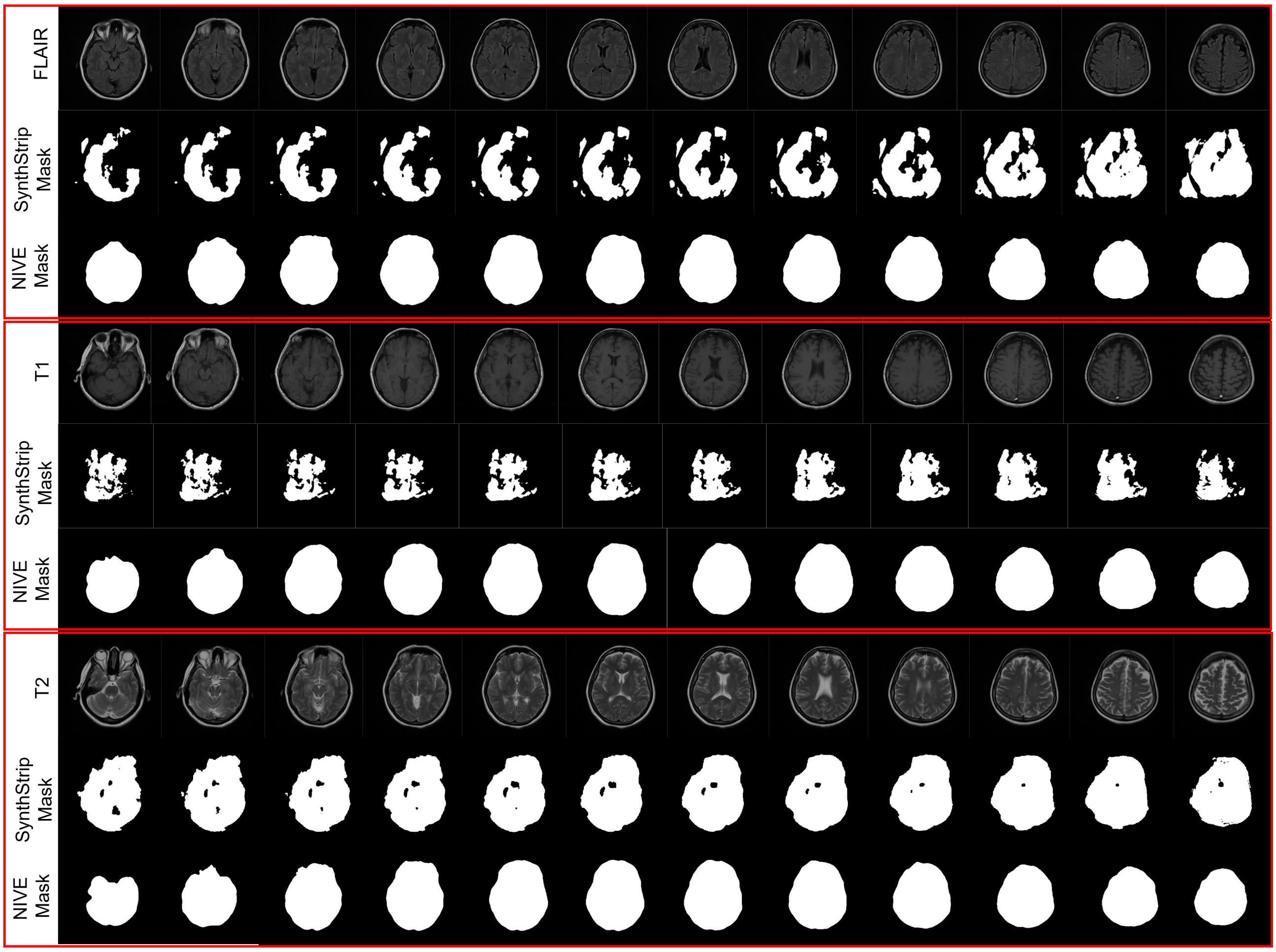

Fig. 6. T1, T2 and FLAIR image from patient 1 of Baghdad Teaching Hospital dataset stripped by SynthStrip and NIVE

It can be observed that in its default settings as shown in Figure 7, SynthStrip from FreeSurfer is incapable of adequate brain extraction. It either removes important cortical areas, or retains skull regions as well. The unsatisfactory performance of SynthStrip on Baghdad dataset, as shown in Figure 6, resulted in the motivation for this research. This section provides a performance comparison of DeepLabV3+, U-Net and SynthStrip.

```
kk@LAPTOP-O6JFKF03:~$ mri_synthstrip -i ./mri/1-T1.nii -o stripped.nii -m mask.nii
Configuring model on the CPU
Running SynthStrip model version 1
Input image read from: ./mri/1-T1.nii
Masked image saved to: stripped.nii
Binary brain mask saved to: mask.nii
If you use SynthStrip in your analysis, please cite:
----------------------------------------------------
SynthStrip: Skull-Stripping for Any Brain Image.
A Hoopes, JS Mora, AV Dalca, B Fischl, M Hoffmann.
NeuroImage 206 (2022), 119474.
```

Fig. 7. SynthStrip command line

4.3 Evaluation of Skull Stripping using MICCAI 2016, SynthStrip and NFBS Online Datasets

For semantic segmentation, Dice is considered more accurate. Figure 8 shows the Dice scores of the validation set for DeepLabV3+, with about 10% near skull images (containing minimum brain content) discarded. The mean Dice score is 0.98 which is quite comparable to SynthStrip. To further assess the current state of the art SynthStrip and the models trained in this research, two subjects were used for testing from each of the three online datasets namely the MICCAI 2016 challenge dataset, the SynthStrip Dataset and the Neurofeedback Skull-stripped (NFBS) repository. From the MICCAI Dataset, the subjects chosen for testing include FLAIR axial scans from testing centre 7, patient 9 and 10. Qin FLAIR 45 and 46 were chosen from the SynthStrip dataset in FLAIR sequence and axial orientation, whereas T1-weighted sagittal scans from subjects 63589 and 64081 were selected from NFBS dataset.

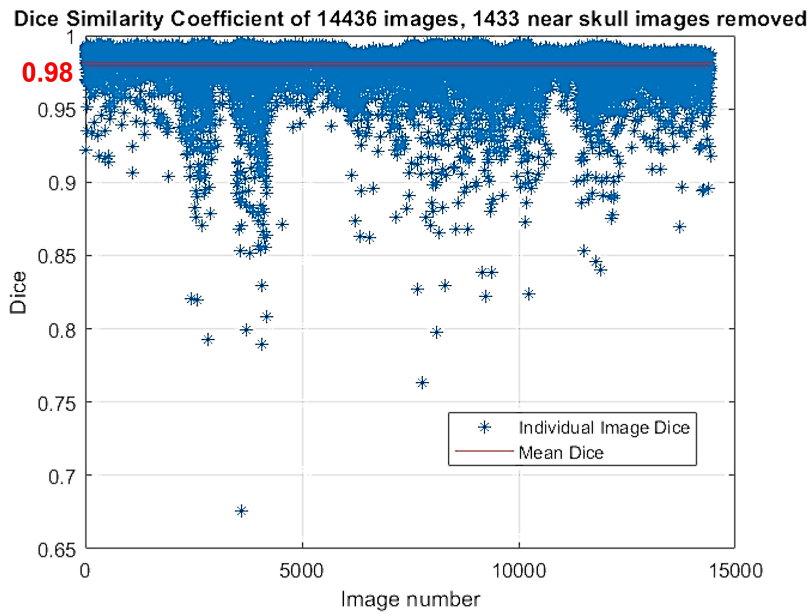

Fig. 8. Mean Dice score for DeepLabV3+ on validation set

The dataset and subject identifier, the number of slices examined, along with the Dice scores for the three models is given in Figure 9. DeepLabV3+ can be seen outperforming the other two, closely followed by SynthStrip. To further analyse the performance of DeepLabV3+ and SynthStrip, tests on real-time absolutely fresh and unseen data from a hospital were performed and the results are presented in the following section.

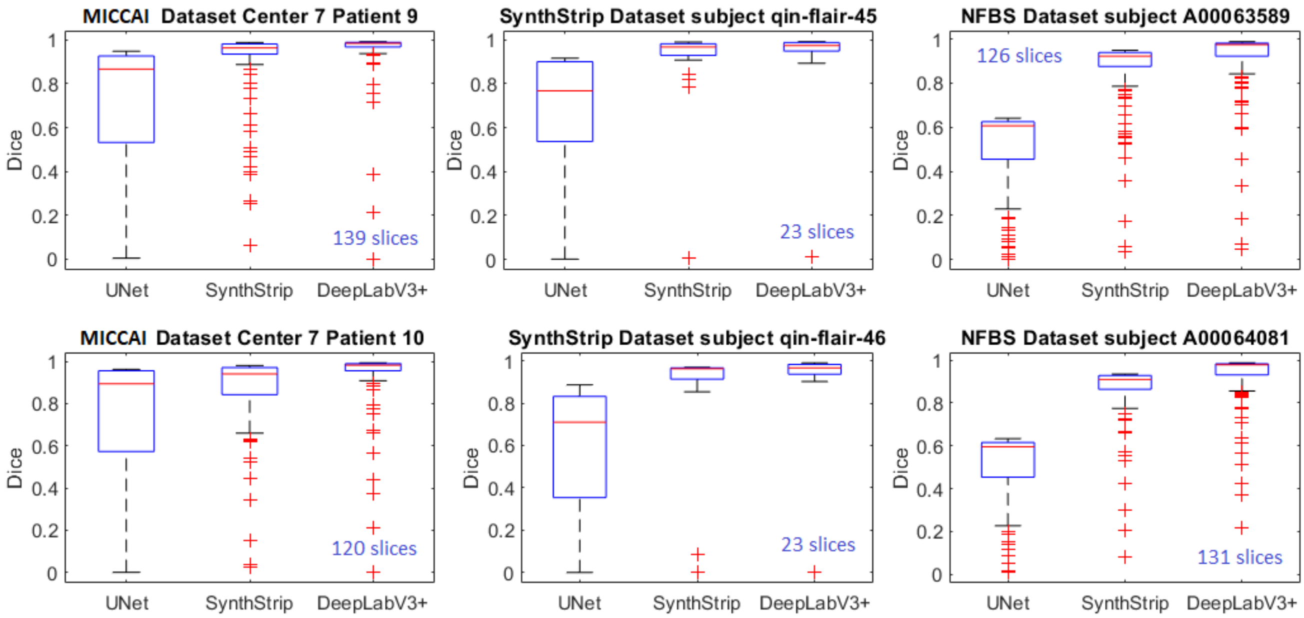

Fig. 9. Dice score of 6 different test subjects from MICCAI 2016, SynthStrip and NFBS (from left to right) datasets using the SynthStrip, and trained U-Net/DeepLabV3+. The number of MRI slices for each subject is given in the plot, with the lowest value of 23-slice and the highest value of 139-slice

4.4 Evaluation of Skull Stripping using AIH Islamabad Dataset

Three subjects were chosen by a consultant radiologist incorporating T1, T2 and FLAIR scans for all three orientations. The performance of the three models was compared with the gold standard brain masks. The Dice scores of the three models, the MRI sequence and orientation, and the number of slices analysed per subject are given in Figure 10. The mean Dice scores are given in Table 4.

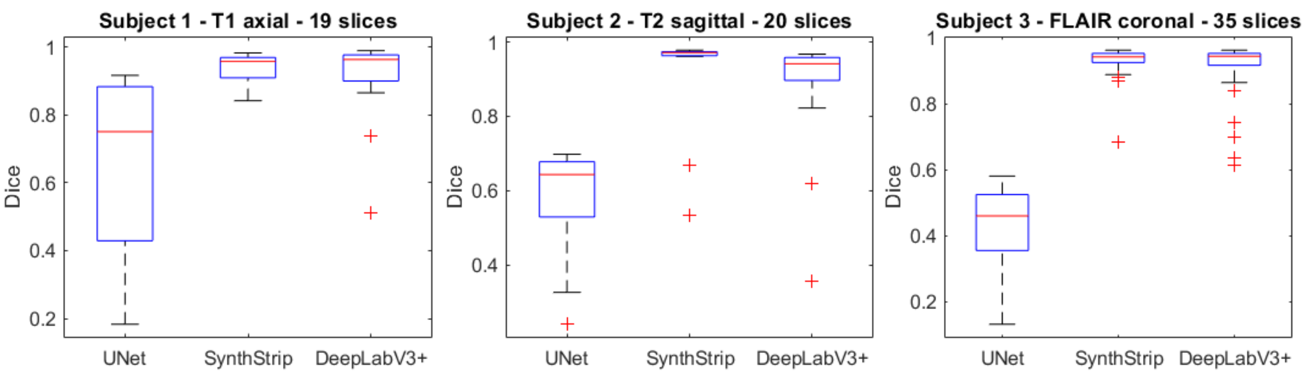

Fig. 10. Dice score of 3 test subjects from AIH Islamabad dataset evaluated using axial, sagittal and coronal MRI. The number of slices for subject 1, 2 and 3 (from left to right) is respectively, 19, 20 and 35

The brain masks generated by U-Net, SynthStrip and DeepLabV3+ from subject 1 (T1W axial) are given in Figure 11. The intersection of model mask and ground truth mask is given in green, the blue region depicts the brain region in ground truth not detected by model, whereas the red region shows the areas which were treated as brain by the models but were not actually the brain according to the ground truth data.

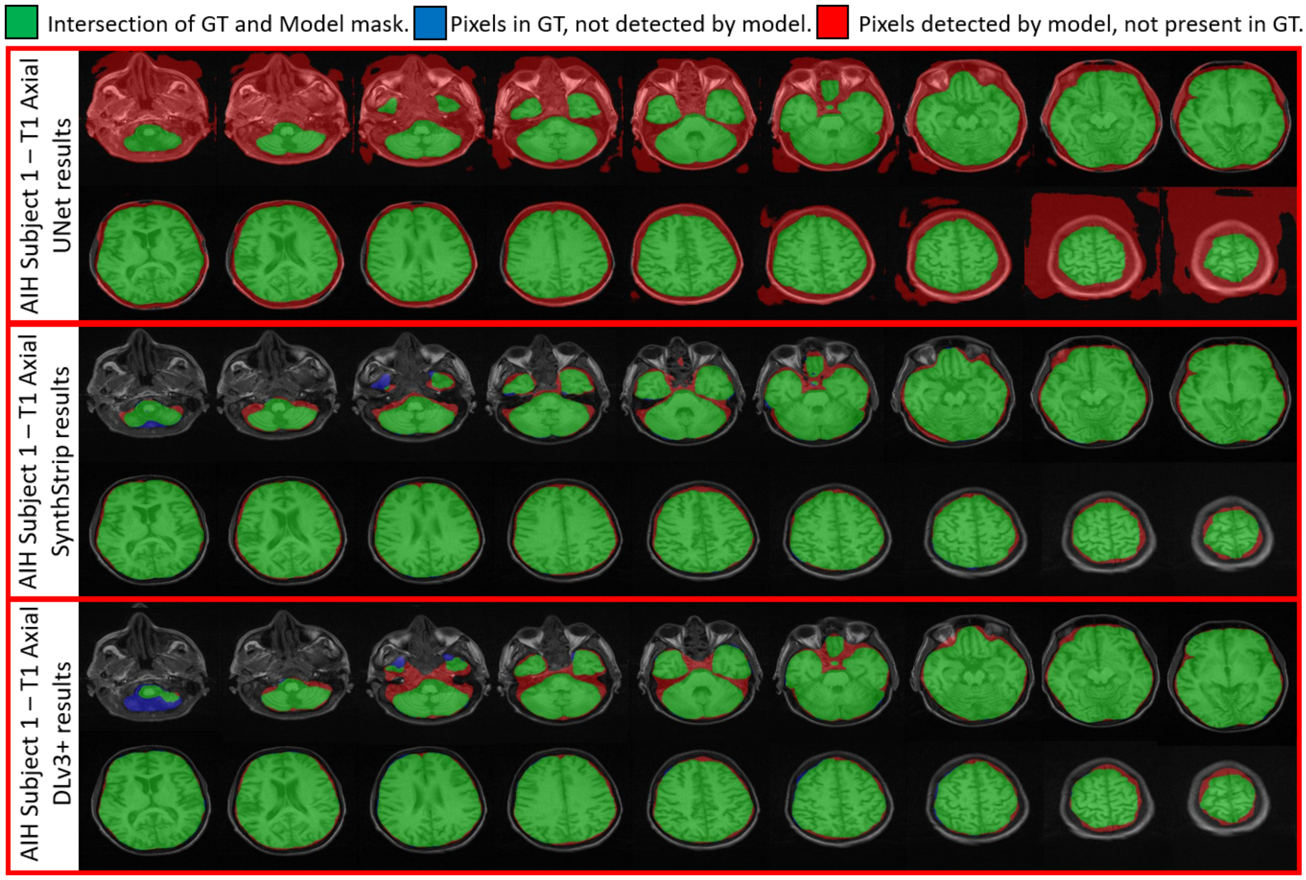

Fig. 11. Extraction of brain region of Subject 1 form AIH Islamabad dataset using U-Net (top-row), SynthStrip (mid-row) and DeepLabV3+ (bottom-row). Here, green is the intersection of the ground truth (GT) with the predicted mask, blue is the pixel in the GT but not detected by the model and red is the wrongly detected pixel

It can be observed here that U-Net did not perform the skull stripping job well, as almost the entire MRI is being treated as the brain. SynthStrip seems to be performing quite well other than slice 1 and 3, in which it misses out some important brain regions. The poor performance of DeepLabV3+ in the first slice with very low brain content may be because from the NFBS dataset, only 8 deep brain slices were taken for training, compromising the model's performance on near skull (MRI edge) images.

DeepLabV3+ can be observed outperforming SynthStrip in all online datasets. In the AIH Islamabad dataset, SynthStrip and DeepLabV3+ show comparable results for T1w axial (subject 1) and FLAIR coronal (subject 3). In case of subject 2 (T2W sagittal), the inferior performance of DeepLabV3+ may be because of the model being trained on fewer images in that orientation as compared to axial and coronal MRI. The overall satisfactory performance of DeepLabV3+ encouraged the development of NIVE (NeuroImaging Volumetric Extractor) to serve as a tool to assist neurologists, radiologists, data scientists and researchers in medical practice and CAD research.

Table 4
Comparison of Mean Dice Score between SynthStrip, and skull stripping techniques using U-Net, DeeplabV3+

Dataset	Sequence	Orientation	U-Net	SynthStrip	DLv3+
MICCAI Center7 case9	FLAIR	Axial	0.7174	0.9086	0.9094
MICCAI Center7 case10	FLAIR	Axial	0.7275	0.8669	0.9190
SynthStrip qin 45	FLAIR	Axial	0.6801	0.9046	0.9246
SynthStrip qin 46	FLAIR	Axial	0.6001	0.8658	0.8794
NFBS 63589	T1W	Sagittal	0.5064	0.8574	0.9117
NFBS 64081	T1W	Sagittal	0.5137	0.8631	0.9265
AIH Islamabad Subject 1	T1W	Axial	0.6550	0.9356	0.9134
AIH Islamabad Subject 2	T2W	Sagittal	0.5854	0.9337	0.8866
AIH Islamabad Subject 3	FLAIR	Coronal	0.4256	0.9280	0.9043

5. NIVE Graphical User Interface

MATLAB Graphical User Interface Development Environment (GUIDE) was used to design and develop NIVE. NIVE is capable of handling NIfTI, Dicom, Jpg, Png, Bmp and other image data formats as input which gives it an edge over SynthStrip which only accepts NIfTI file format as input, and generates NIfTI volumes for both brain masks and skull-stripped output. Since software used by radiologists (like RadiAnt) majorly support Dicom sequences, a diverse input acceptance feature of NIVE would prove to be helpful in medical practice. NIVE can provide individual skull-stripped slices or entire volumes depending on user input. It has a flexible and user-friendly interface, which offers skimming through slices using slider and visualizations for raw unprocessed MRI, its corresponding brain mask, and the skull-stripped version simultaneously. SynthStrip on the other hand is a command-line based tool, which some medical professionals might find tedious to use. The data import and export options in NIVE can be controlled using pushbuttons. The UI for NIVE is given in Figure 12.

6. NIVE, Model and Dataset Availability

NIVE v1.0 is available at MathWorks as a MATLAB app installer package at the link https://www.mathworks.com/matlabcentral/fileexchange/129574-nive. The online publicly available datasets can be downloaded from their respective web links whereas the data acquired from Advance International Hospital Islamabad, along with the ground truth masks is available at the link https://www.kaggle.com/datasets/khuhedkhalid/aih-skullstripping-data.

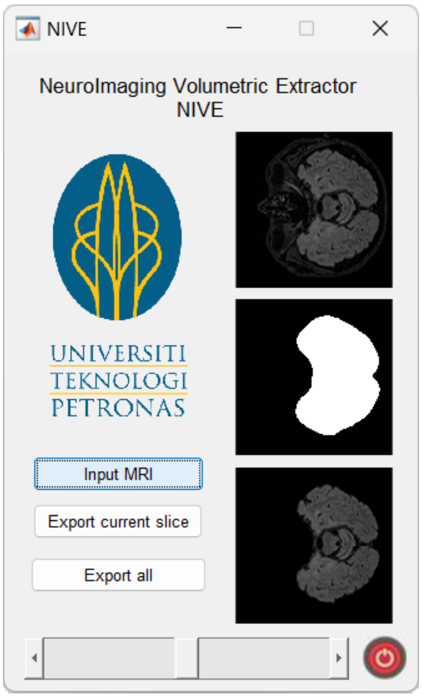

Fig. 12. NIVE GUI developed on MATLAB

7. Conclusions

Brain extraction is an important preprocessing step in CAD systems using brain MRI. Adequate performance of a skull stripping system facilitates the subsequent processes involved in reaching concrete diagnoses. DeepLabV3+ model has been trained with the most comprehensive human brain dataset and has outperformed U-Net and the current state of the art tool SynthStrip. The trained model has been embedded in NIVE to serve as a tool to assist radiologists and data scientists working on CAD systems for neurological disorders. NIVE has proven to be the best publicly available system so far which is agnostic to MRI input file type, sequence, orientation, subject age, brain pathology, and acquisition hardware variations.

Acknowledgement
This work was supported in part by the Ministry of Higher Education (MOHE), Malaysia under Fundamental Research Grant Scheme (FRGS/1/2021/TK0/UTP/02/17), in part by the Universiti Teknologi PETRONAS under Pre-Commercialization Prototype Grant (015PBA-038) and in part by Deanship of Research Oversight and Coordination (DROC) at King Fahd University of Petroleum and Minerals (KFUPM) project No. EC221016. The authors would also like to thank the Centre for Intelligent Signal & Imaging Research (CISIR) & the Centre of Graduate Studies (CGS), Universiti Teknologi PETRONAS (UTP) for providing part of the funding and the required facilities to conduct this research work.

References
[1] Gao, Yan, Jie Li, Haojun Xu, Miaomiao Wang, Congcong Liu, Yannan Cheng, Mengxuan Li, Jian Yang, and Xianjun Li. "A multi-view pyramid network for skull stripping on neonatal T1-weighted MRI." *Magnetic resonance imaging* 63 (2019): 70-79. https://doi.org/10.1016/j.mri.2019.08.025

[2] Thakur, Siddhesh, Jimit Doshi, Sarthak Pati, Saima Rathore, Chiharu Sako, Michel Bilello, Sung Min Ha *et al.,* "Brain extraction on MRI scans in presence of diffuse glioma: Multi-institutional performance evaluation of deep learning

methods and robust modality-agnostic training." *NeuroImage* 220 (2020): 117081. https://doi.org/10.1016/j.neuroimage.2020.117081
[3] Galdames, Francisco J., Fabrice Jaillet, and Claudio A. Perez. "An accurate skull stripping method based on simplex meshes and histogram analysis for magnetic resonance images." *Journal of neuroscience methods* 206, no. 2 (2012): 103-119. https://doi.org/10.1016/j.jneumeth.2012.02.017
[4] Lucena, Oeslle, Roberto Souza, Leticia Rittner, Richard Frayne, and Roberto Lotufo. "Convolutional neural networks for skull-stripping in brain MR imaging using silver standard masks." *Artificial intelligence in medicine* 98 (2019): 48-58. https://doi.org/10.1016/j.artmed.2019.06.008
[5] Rex, David E., David W. Shattuck, Roger P. Woods, Katherine L. Narr, Eileen Luders, Kelly Rehm, Sarah E. Stolzner, David A. Rottenberg, and Arthur W. Toga. "A meta-algorithm for brain extraction in MRI." *NeuroImage* 23, no. 2 (2004): 625-637. https://doi.org/10.1016/j.neuroimage.2004.06.019
[6] Pei, Linmin, Murat Ak, Nourel Hoda M. Tahon, Serafettin Zenkin, Safa Alkarawi, Abdallah Kamal, Mahir Yilmaz *et al.,* "A general skull stripping of multiparametric brain MRIs using 3D convolutional neural network." *Scientific Reports* 12, no. 1 (2022): 10826. https://doi.org/10.1038/s41598-022-14983-4
[7] Iglesias, Juan Eugenio, Cheng-Yi Liu, Paul M. Thompson, and Zhuowen Tu. "Robust brain extraction across datasets and comparison with publicly available methods." *IEEE transactions on medical imaging* 30, no. 9 (2011): 1617-1634. https://doi.org/10.1109/TMI.2011.2138152
[8] Leung, Kelvin K., Josephine Barnes, Marc Modat, Gerard R. Ridgway, Jonathan W. Bartlett, Nick C. Fox, Sébastien Ourselin, and Alzheimer's Disease Neuroimaging Initiative. "Brain MAPS: an automated, accurate and robust brain extraction technique using a template library." *Neuroimage* 55, no. 3 (2011): 1091-1108. https://doi.org/10.1016/j.neuroimage.2010.12.067
[9] Eskildsen, Simon F., Pierrick Coupé, Vladimir Fonov, José V. Manjón, Kelvin K. Leung, Nicolas Guizard, Shafik N. Wassef, Lasse Riis Østergaard, D. Louis Collins, and Alzheimer's Disease Neuroimaging Initiative. "BEaST: brain extraction based on nonlocal segmentation technique." *NeuroImage* 59, no. 3 (2012): 2362-2373. https://doi.org/10.1016/j.neuroimage.2011.09.012
[10] Hoopes, Andrew, Jocelyn S. Mora, Adrian V. Dalca, Bruce Fischl, and Malte Hoffmann. "SynthStrip: skull-stripping for any brain image." *NeuroImage* 260 (2022): 119474. https://doi.org/10.1016/j.neuroimage.2022.119474
[11] Roy, Shaswati, and Pradipta Maji. "A simple skull stripping algorithm for brain MRI." In *2015 Eighth International Conference on Advances in Pattern Recognition (ICAPR)*, pp. 1-6. IEEE, 2015. https://doi.org/10.1109/ICAPR.2015.7050671
[12] Kandaya, Shaarmila, Abdul Rahim Abdullah, Norhashimah Mohd Saad, Izzatul Husna Azman, Ezreen Farina Shair, and Nur Hasanah Ali. "Analysis of Early Stroke Diagnosis Based on Brain Magnetic Resonance Imaging using Machine Learning." *Journal of Advanced Research in Applied Sciences and Engineering Technology* 32, no. 3 (2023): 241-255. https://doi.org/10.37934/araset.32.3.241255
[13] Anita, John Nisha, and Sujatha Kumaran. "Detection and Segmentation of Meningioma Tumors Using the Proposed MENCNN Model." *Journal of Advanced Research in Applied Sciences and Engineering Technology* 32, no. 2 (2023): 1-13. https://doi.org/10.37934/araset.32.2.113
[14] Souza, Roberto, Oeslle Lucena, Julia Garrafa, David Gobbi, Marina Saluzzi, Simone Appenzeller, Letícia Rittner, Richard Frayne, and Roberto Lotufo. "An open, multi-vendor, multi-field-strength brain MR dataset and analysis of publicly available skull stripping methods agreement." *NeuroImage* 170 (2018): 482-494. https://doi.org/10.1016/j.neuroimage.2017.08.021
[15] Lim, Sheh Hong, Mohd Azrul Hisham Mohd Adib, Mohd Shafie Abdullah, Nur Hartini Mohd Taib, Radhiana Hassan, and Azian Abd Aziz. "Study of extracted geometry effect on patient-specific cerebral aneurysm model with different threshold coefficient (Cthres)." *CFD Letters* 12, no. 10 (2020): 1-14. https://doi.org/10.37934/cfdl.12.10.114
[16] Rehman, Hafiz Zia Ur, Hyunho Hwang, and Sungon Lee. "Conventional and deep learning methods for skull stripping in brain MRI." *Applied Sciences* 10, no. 5 (2020): 1773. https://doi.org/10.3390/app10051773
[17] Ranjbar, Sara, Kyle W. Singleton, Lee Curtin, Cassandra R. Rickertsen, Lisa E. Paulson, Leland S. Hu, J. Ross Mitchell, and Kristin R. Swanson. "Robust automatic whole brain extraction on magnetic resonance imaging of brain tumor patients using dense-Vnet." *arXiv preprint arXiv:2006.02627* (2020).
[18] Hwang, Hyunho, Hafiz Zia Ur Rehman, and Sungon Lee. "3D U-Net for skull stripping in brain MRI." *Applied Sciences* 9, no. 3 (2019): 569. https://doi.org/10.3390/app9030569
[19] Zhang, Qian, Li Wang, Xiaopeng Zong, Weili Lin, Gang Li, and Dinggang Shen. "Frnet: Flattened residual network for infant MRI skull stripping." In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pp. 999-1002. IEEE, 2019. https://doi.org/10.1109/ISBI.2019.8759167
[20] Lutkenhoff, Evan S., Matthew Rosenberg, Jeffrey Chiang, Kunyu Zhang, John D. Pickard, Adrian M. Owen, and Martin M. Monti. "Optimized brain extraction for pathological brains (optiBET)." *PloS one* 9, no. 12 (2014): e115551. https://doi.org/10.1371/journal.pone.0115551

[21] Salehi, Seyed Sadegh Mohseni, Deniz Erdogmus, and Ali Gholipour. "Auto-context convolutional neural network (auto-net) for brain extraction in magnetic resonance imaging." *IEEE transactions on medical imaging* 36, no. 11 (2017): 2319-2330. https://doi.org/10.1109/TMI.2017.2721362

[22] Ezhilarasan, K., S. Praveenkumar, K. Somasundaram, T. Kalaiselvi, S. Magesh, S. Kiruthika, and A. Jeevarekha. "Automatic brain extraction from MRI of human head scans using Helmholtz free energy principle and morphological operations." *Biomedical Signal Processing and Control* 64 (2021): 102270. https://doi.org/10.1016/j.bspc.2020.102270

[23] Rempe, Moritz, Florian Mentzel, Kelsey L. Pomykala, Johannes Haubold, Felix Nensa, Kevin Kröninger, Jan Egger, and Jens Kleesiek. "k-strip: A novel segmentation algorithm in k-space for the application of skull stripping." *Computer Methods and Programs in Biomedicine* 243 (2024): 107912. https://doi.org/10.1016/j.cmpb.2023.107912

[24] Kaliyugarasan, Satheshkumar, Marek Kocinski, Arvid Lundervold, and Alexander Selvikvåg Lundervold. "2D and 3D U-Nets for skull stripping in a large and heterogeneous set of head MRI using fastai." (2020).

[25] Smith, Stephen M. "Fast robust automated brain extraction." *Human brain mapping* 17, no. 3 (2002): 143-155. https://doi.org/10.1002/hbm.10062

[26] Ségonne, Florent, Anders M. Dale, Evelina Busa, Maureen Glessner, David Salat, Horst Karl Hahn, and Bruce Fischl. "A hybrid approach to the skull stripping problem in MRI." *Neuroimage* 22, no. 3 (2004): 1060-1075. https://doi.org/10.1016/j.neuroimage.2004.03.032

[27] Somasundaram, K., and T. Kalaiselvi. "Fully automatic brain extraction algorithm for axial T2-weighted magnetic resonance images." *Computers in biology and medicine* 40, no. 10 (2010): 811-822. https://doi.org/10.1016/j.compbiomed.2010.08.004

[28] Roslan, Rosniza, Nursuriati Jamil, and Rozi Mahmud. "Skull stripping of MRI brain images using mathematical morphology." In *2010 IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES)*, pp. 26-31. IEEE, 2010. https://doi.org/10.1109/IECBES.2010.5742193

[29] Somasundaram, K., and T. Kalaiselvi. "Automatic brain extraction methods for T1 magnetic resonance images using region labeling and morphological operations." *Computers in biology and medicine* 41, no. 8 (2011): 716-725. https://doi.org/10.1016/j.compbiomed.2011.06.008

[30] Kleesiek, Jens, Gregor Urban, Alexander Hubert, Daniel Schwarz, Klaus Maier-Hein, Martin Bendszus, and Armin Biller. "Deep MRI brain extraction: A 3D convolutional neural network for skull stripping." *NeuroImage* 129 (2016): 460-469. https://doi.org/10.1016/j.neuroimage.2016.01.024

[31] Roy, Snehashis, John A. Butman, Dzung L. Pham, and Alzheimers Disease Neuroimaging Initiative. "Robust skull stripping using multiple MR image contrasts insensitive to pathology." *Neuroimage* 146 (2017): 132-147. https://doi.org/10.1016/j.neuroimage.2016.11.017

[32] Isensee, Fabian, Marianne Schell, Irada Pflueger, Gianluca Brugnara, David Bonekamp, Ulf Neuberger, Antje Wick *et al.,* "Automated brain extraction of multisequence MRI using artificial neural networks." *Human brain mapping* 40, no. 17 (2019): 4952-4964. https://doi.org/10.1002/hbm.24750

[33] Valverde Valverde, Sergi, Llucia Coll, Liliana Valencia, Albert Clèrigues, Arnau Oliver i Malagelada, Joan Carles Vilanova, Lluís Ramió i Torrentà, Àlex Rovira, and Xavier Lladó Bardera. "Assessing the Accuracy and Reproducibility of PARIETAL: A Deep Learning Brain Extraction Algorithm." *Journal of Magnetic Resonance Imaging, 2021, Online Version of Record before inclusion in an issue* (2021). https://doi.org/10.1002/jmri.27776

[34] Duarte, Kauê Tartarotti Nepomuceno, Marinara Andrade Nascimento Moura, Paulo Sergio Martins, and Marco Antonio Garcia de Carvalho. "Brain Extraction in Multiple T1-weighted Magnetic Resonance Imaging slices using Digital Image Processing techniques." *IEEE Latin America Transactions* 20, no. 5 (2022): 831-838. https://doi.org/10.1109/TLA.2022.9693568

[35] Paredes-Orta, Carlos, Jorge Domingo Mendiola-Santibañez, Danjela Ibrahimi, Juvenal Rodríguez-Reséndiz, Germán Díaz-Florez, and Carlos Alberto Olvera-Olvera. "Hyperconnected openings codified in a max tree structure: an application for skull-stripping in brain MRI T1." *Sensors* 22, no. 4 (2022): 1378. https://doi.org/10.3390/s22041378

[36] Puccio, Benjamin, James P. Pooley, John S. Pellman, Elise C. Taverna, and R. Cameron Craddock. "The preprocessed connectomes project repository of manually corrected skull-stripped T1-weighted anatomical MRI data." *Gigascience* 5, no. 1 (2016): s13742-016. https://doi.org/10.1186/s13742-016-0150-5

[37] Commowick, Olivier, Michaël Kain, Romain Casey, Roxana Ameli, Jean-Christophe Ferré, Anne Kerbrat, Thomas Tourdias *et al.,* "Multiple sclerosis lesions segmentation from multiple experts: The MICCAI 2016 challenge dataset." *Neuroimage* 244 (2021): 118589. https://doi.org/10.1016/j.neuroimage.2021.118589

[38] Muslim, Ali M., Syamsiah Mashohor, Gheyath Al Gawwam, Rozi Mahmud, Marsyita binti Hanafi, Osama Alnuaimi, Raad Josephine, and Abdullah Dhaifallah Almutairi. "Brain MRI dataset of multiple sclerosis with consensus manual lesion segmentation and patient meta information." *Data in Brief* 42 (2022): 108139. https://doi.org/10.1016/j.dib.2022.108139