

3D Brain Extraction from Magnetic Resonance Imaging Using Knowledge Distillation

Kali Gurkahraman
0000-0002-0697-125X

Department of Computer Engineering
Sivas Cumhuriyet University
Sivas, Turkey
kgurkahraman@cumhuriyet.edu.tr

Ahmet Firat Yelkuvan
0000-0003-4148-1923

Department of Computer Engineering
Sivas Cumhuriyet University
Sivas, Turkey
aftyelkuvan@cumhuriyet.edu.tr

Rukiye Karakis
0000-0002-1797-3461

Department of Software Engineering
Sivas Cumhuriyet University
Sivas, Turkey
rkarakis@cumhuriyet.edu.tr

Abstract—Brain extraction, or skull stripping, is a crucial preprocessing step in magnetic resonance imaging (MRI), isolating brain tissue from surrounding structures like the skull and scalp. However, existing methods have limitations, such as parameter sensitivity in traditional approaches and computational complexity in advanced deep learning architectures. This study proposes a knowledge distillation framework utilizing two UNet++ models—a high-capacity teacher network and an efficient student network—for 3D brain extraction tasks. The teacher network generates detailed grayscale brain predictions, capturing subtle intensity transitions and anatomical boundaries. The student network learns to produce precise binary segmentation masks from the teacher’s feature representations, guided by a hybrid loss function combining Dice, Structural Similarity Index Measure (SSIM), and Mean Squared Error (MSE). Evaluations conducted on T1-weighted, T2-weighted, and proton-density weighted MRI images from the IXI dataset demonstrated the student model’s superior performance, achieving a Dice coefficient of 0.97857. These findings suggest that the proposed framework may offer a practical and accurate solution for brain extraction in diverse medical imaging scenarios.

Index Terms—brain extraction, deep learning, hybrid loss, knowledge distillation, UNet++

I. INTRODUCTION

BRAIN extraction (or skull stripping) is a key Magnetic Resonance Imaging (MRI) preprocessing step that isolates brain tissue from non-brain structures like the skull and scalp. Its accuracy significantly impacts downstream tasks such as volumetric analysis, image registration, lesion segmentation, anatomical delineation, cortical thickness estimation, motor function prediction, and neurosurgical planning [1-3].

Brain extraction enhances the accuracy of downstream analyses by isolating the brain region, thereby eliminating extraneous signals and noise from surrounding non-brain tissues. This results in more precise and consistent outcomes. Although manual segmentation remains the gold standard, it is labor-intensive, time-consuming, and subject to inter-operator variability, affecting reproducibility [4]. Inadequate brain extraction, if not manually corrected, can introduce significant errors in further neuroimaging analyses. To overcome these challenges, many (semi-)automated brain extraction methods have been proposed and refined in recent years.

In the literature, many methods have been proposed to

separate the brain—considered the region of interest—from non-brain tissues. These include basic image processing steps like erosion, dilation, thresholding, and edge detection [4-7]. One well-known approach is the Brain Surface Extraction (BSE) method developed by Shattuck et al. [6], which is used in the BrainSuite software [7]. BSE combines anisotropic diffusion filtering, edge detection, and morphological operations to extract the brain. However, these types of methods often need manual tuning of parameters for each image, which makes them less practical and more time-consuming for large datasets.

To improve brain extraction, deformable models such as active contours (snakes) and level sets have been used [4]. These methods iteratively adjust a shape to match the brain’s boundaries by minimizing energy. A well-known example is the Brain Extraction Tool (BET) [8] and its improved version BET2 [9], both part of the FMRIB Software Library (FSL) [10]. They begin with a spherical model at the head’s center and deform it to fit the brain. These models handle intensity variations well and produce smooth results, but their accuracy depends on initial settings and may be limited by irregular brain shapes or pathologies [4].

Another widely used approach involves atlas-based methods, which use pre-segmented brain atlases to guide the segmentation of new MRI scans. These approaches typically involve registering the atlas to the subject’s image and transferring labels to identify brain regions. For example, Dale et al. [11] introduced a skull-stripping method within the FreeSurfer framework [12], which normalizes image intensities and deforms a tessellated ellipsoidal template to fit the inner skull surface. Another well-known method, BEaST [13], performs patch-based segmentation with linear registration to the ICBM152 template using multi-resolution images from both healthy and Alzheimer’s patients. Although atlas-based methods can achieve high accuracy—especially when the atlas closely represents the target population—their performance strongly depends on the accuracy of image registration and the quality of the chosen template.

In recent years, hybrid methods have been proposed to build on the strengths of atlas-based and other traditional techniques. These methods aim to improve both robustness and accuracy

by combining multiple approaches. For instance, thresholding can be enhanced with machine learning, or deformable models can be integrated with atlas-based strategies. Souza et al. [14] used eight segmentation methods to generate brain masks, which were then fused using the STAPLE algorithm [15] to produce a consensus result. Another example, the Hybrid Watershed Algorithm (HWA) [16], combines watershed segmentation with deformable surface modeling based solely on intensity information. While hybrid approaches often yield better results than individual methods, their performance still depends on factors such as parameter tuning, atlas/template selection, and registration quality.

Artificial neural networks (ANNs), especially deep learning (DL) models, have shown strong performance in medical image analysis tasks [3, 17-18]. In brain extraction, several DL-based methods have been developed to improve segmentation accuracy. One of the earliest was proposed by Kleesiek et al. [19], who introduced a convolutional neural network (CNN) for brain extraction. Their model achieved competitive Dice scores on T1-weighted, T2-weighted, and FLAIR MRI scans and showed better specificity compared to traditional methods. Building on early CNN-based work, researchers have adapted 3D-UNet architectures from their 2D versions, using encoder-decoder blocks to improve performance [4, 19-23]. Hwang et al. [20] applied a modified 3D-UNet to T1w MRIs, and Isensee et al. [4] introduced HD-BET, which performed well across different MRI sequences and scanners. Zhang et al. [21] proposed FRNET with residual connections and a boundary loss function, showing strong results on infant MRIs, though it has not been tested on adult data. Other studies have also enhanced 3D-UNet models using residual features [22], or by combining real and synthetic images through GANs, as seen in the work by Hoopes et al. [23].

The UNet model has been adapted in many ways to handle complex medical imaging tasks better, especially those involving MRI. In recent studies, several ensemble and cascaded versions of UNet have been developed to improve segmentation accuracy. Cascaded UNet architectures, in particular, have shown strong performance in capturing complex anatomical structures and resolving ambiguous boundaries. These models often work in multiple stages, where an initial network produces a rough prediction and a second network refines it. This setup leads to more stable and detailed results, especially around brain edges [24-26].

However, despite their benefits, cascaded UNet models can be computationally expensive and may overfit when trained on limited or highly variable datasets. To address these issues, this study proposes a knowledge distillation framework where a high-capacity teacher model guides a lightweight student model for 3D brain extraction from MRI scans. The teacher network produces a detailed grayscale brain prediction, which helps the student model learn to generate an accurate binary brain mask. Both models are designed for volumetric data, and a hybrid loss function—combining Dice, structural similarity index measure (SSIM), and mean squared error (MSE)—is used to improve boundary accuracy while handling

class imbalance. This approach achieves strong segmentation performance on the tested dataset and shows potential for broader application in similar medical imaging tasks.

II. METHODOLOGY

This study proposes a knowledge distillation framework that integrates two UNet++ models, each trained with a distinct objective to balance anatomical precision and segmentation efficiency. The teacher model is trained to perform grayscale brain extraction using input from three common MRI modalities: T1-weighted (T1w), which highlights anatomical structure; T2-weighted (T2w), which is sensitive to fluid and pathology; and proton density-weighted (PDw) imaging, which emphasizes tissue contrast based on hydrogen concentration. These scans were obtained from the IXI dataset and selected to ensure diversity in anatomical and contrast information. This design enables the teacher to capture fine-grained intensity patterns and structural boundaries. In contrast, the student model receives the same MRI input but is trained to produce a binary brain mask. By learning from the feature representations of the teacher, the student model acquires anatomical awareness while remaining optimized for efficient binary segmentation.

The proposed framework incorporates a modified DL architecture (UNet++), a hybrid loss function combining Dice, SSIM, and MSE, and a knowledge transfer strategy to enhance generalization. Details of the model structure, the training objectives for both teacher and student networks, the knowledge distillation process, and the evaluation metrics used for performance assessment are described in the following sections.

A. UNET++

In this study, the UNet++ architecture, shown in Figure 1, was employed for 3D brain extraction from MRI. UNet++ was selected due to its enhanced capability to capture fine structural details and improve segmentation accuracy—particularly at object boundaries—through its nested and densely connected design. As an advanced variant of the original U-Net, UNet++ replaces simple skip connections with intermediate convolutional blocks that help reduce the semantic gap between encoder and decoder features. In the architecture diagram, solid black arrows indicate down-sampling in the encoder, while dashed black arrows represent up-sampling in the decoder. Dashed green arrows show lateral dense skip connections at the same resolution, and dashed blue arrows highlight up-sampled features passed into intermediate convolution blocks. The light blue modules represent these nested blocks, which contribute to progressive feature refinement. Additionally, the architecture supports deep supervision at multiple decoder depths, facilitating efficient training and making it well-suited for complex volumetric segmentation tasks such as brain extraction [27].

B. Knowledge Distillation

Knowledge distillation (KD) is a technique where a larger teacher model guides a smaller student model by transferring

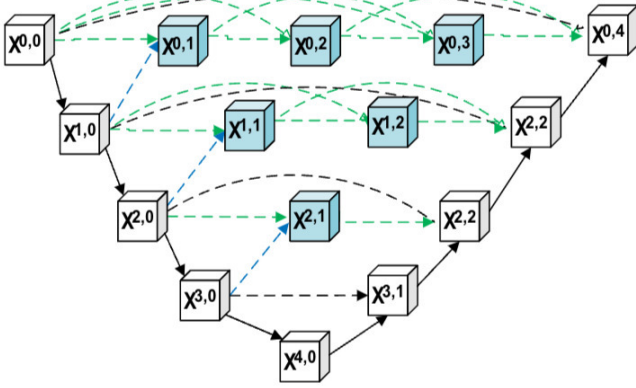


Fig. 1. UNet++ is structured with an encoder and decoder linked by a series of nested, densely connected convolutional blocks.

learned representations. Instead of learning only from ground truth labels, the student also learns from the teacher's soft outputs, capturing richer structural information. In this study, the teacher produces grayscale brain predictions, and the student learns to generate binary brain masks, supervised by both segmentation and distillation losses.

To transfer knowledge from the teacher network to the student network in a medical image segmentation context, we employed a custom distillation loss formulation that integrates both task-specific segmentation loss and feature-level guidance from the teacher model[28].

During training, the student model is supervised by two objectives:

$$L_{total} = (1 - \alpha) \times L_{seg} + \alpha \times L_{distill} \quad (1)$$

Where L_{seg} is the segmentation loss computed between the predicted mask from the student and the ground truth binary mask. This ensures that the student learns the final segmentation task correctly.

$L_{distill}$ is the distillation loss, calculated between the student's predicted mask and the gray-level soft prediction output of the teacher. This guides the student to imitate the spatial structure and internal representation captured by the teacher.

$\alpha \in [0, 1]$ is a weighting coefficient that balances the contribution of the segmentation loss and the distillation loss. In this study α , which is set as 0.3, emphasizes the importance of teacher guidance during training.

C. Loss Function

This study employed a hybrid loss denoted as L_{seg} uses a hybrid loss function combining MSE, SSIM loss, and Dice loss, while $L_{distill}$ typically employs MSE to align the student's output with the soft gray-level guidance provided by the teacher network. The overall hybrid loss function is as in (2).

$$L_{seg} = L_{mse} + L_{ssim} + L_{dice} \quad (2)$$

L_{mse} is used to penalize pixel-wise intensity differences and ensures that the predicted output closely matches the ground truth in terms of raw voxel intensities.

L_{ssim} captures perceptual differences by focusing on luminance, contrast, and structure, thus preserving anatomical consistency in the predicted images.

L_{dice} promotes spatial alignment between binary structures in the ground truth and prediction, which is crucial for accurate segmentation performance.

L_{ssim} and L_{dice} losses are calculated as in (3) and (4).

$$L_{ssim} = 1 - SSIM(P, R) \quad (3)$$

$$L_{dice} = 1 - DICE(P, R) \quad (4)$$

SSIM and DICE between predicted (P) and real (R) images are computed as in (5) and (6).

$$SSIM(P, R) = \frac{(2\mu_P\mu_R + C_1)(2\sigma_{PR} + C_2)}{(\mu_P^2 + \mu_R^2 + C_1)(\sigma_P^2 + \sigma_R^2 + C_2)} \quad (5)$$

$$DICE(P, R) = \frac{2|P \cap R|}{|P| + |R|} \quad (6)$$

The proposed KD framework employs two UNet++ models with distinct training objectives tailored to optimize both anatomical precision and segmentation efficiency. The teacher model is trained to learn a grayscale brain extraction task, where the input is the original T1w, T2w, or PDw MRI image, and the output is a grayscale brain-only image. This approach encourages the teacher network to capture subtle intensity transitions and detailed structural boundaries of the brain tissue. On the other hand, the student model is trained using the same original brain image as input, but its target output is a binary brain mask that delineates the brain region. By learning from the teacher's feature representations via knowledge distillation, the student model gains anatomical awareness while being optimized for efficient binary segmentation.

D. Dataset

The IXI dataset includes multiple MRI modalities. In this study, T1w, T2w, and proton PDw MRI images from the IXI dataset [23] were selected to ensure sufficient anatomical detail and to introduce modality diversity in the brain extraction experiments (Table 1).

TABLE I
SUMMARY OF IXI DATASET MODALITIES, Voxel Size, AND DATASET

Modality	Voxel Size (mm ³)	Images
T1w MRI	0.9×0.9×1.2	50
T2w MRI	0.9×0.9×1.2	50
PDw MRI	0.9×0.9×1.2	50

E. Performance Metrics

The segmentation performance of the UNet++ model was assessed using three standard metrics: Dice coefficient, sensitivity, and specificity. As shown in equation (6), the Dice coefficient is computed as the ratio of twice the overlap between the predicted and ground truth masks to the total area covered by both masks.

Sensitivity, also known as recall, quantifies the model's ability to correctly identify brain tissue within the segmentation, as defined in equation (8). In contrast, specificity measures the effectiveness of the model in correctly excluding non-brain tissue and is computed using the formulation provided in equation (9).

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (7)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (8)$$

III. EXPERIMENTAL RESULTS AND DISCUSSION

The proposed DL model was implemented using the Keras library (Python 3.9) with TensorFlow. Experiments were conducted on a workstation equipped with an NVIDIA RTX A6000 GPU (48 GB), an Intel i9-12900KS processor (3.40 GHz), and 64 GB of RAM. The model was trained using the Adam optimizer with dropout and L2 regularization to reduce overfitting. Key training parameters were set as follows: a learning rate of 0.001, momentum of 0.8, and weight decay of 0.00001. Based on empirical testing, training was performed for 100 epochs with a batch size of 4 to accommodate the memory demands of 3D volumes. An 80/20 train-test split was used, and model performance was evaluated using five-fold repeated random subsampling validation (RSV).

Figure 2 presents representative axial, coronal, and sagittal slices from a T1w MRI scan in the IXI dataset. Figure 2a shows the original input image, while Figure 2b displays the corresponding grayscale brain image used as ground truth for training the teacher model. Figure 2c illustrates the output predicted by the UNet++ teacher network. The predicted brain images show strong visual similarity to the reference images, preserving anatomical structures and intensity gradients across views. Quantitatively, the teacher model achieved a Mean Absolute Error (MAE) of 0.0249, a PSNR of 51.46 decibel (dB), and an SSIM of 0.9490, demonstrating high reconstruction accuracy and perceptual quality. These results confirm the effectiveness of the proposed approach in generating anatomically faithful grayscale brain extractions.

Figure 3 shows representative binary segmentation results from axial, coronal, and sagittal MRI slices. Figure 3a displays the ground truth binary brain masks used as references. Figure 3b illustrates predictions generated by a standard UNet++ model trained without knowledge distillation, while Figure 3c presents outputs from the UNet++ student model trained with knowledge distillation. The student model predictions

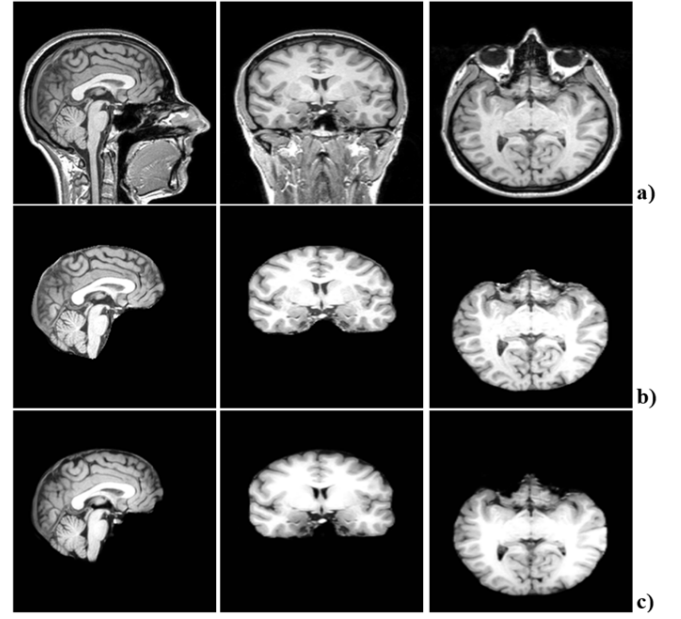


Fig. 2. Sagittal, coronal, and axial slices from a T1w MRI image in the IXI dataset. (a) Original 3D input volume, (b) ground truth grayscale brain image, (c) grayscale brain output predicted by the UNet++ teacher model.

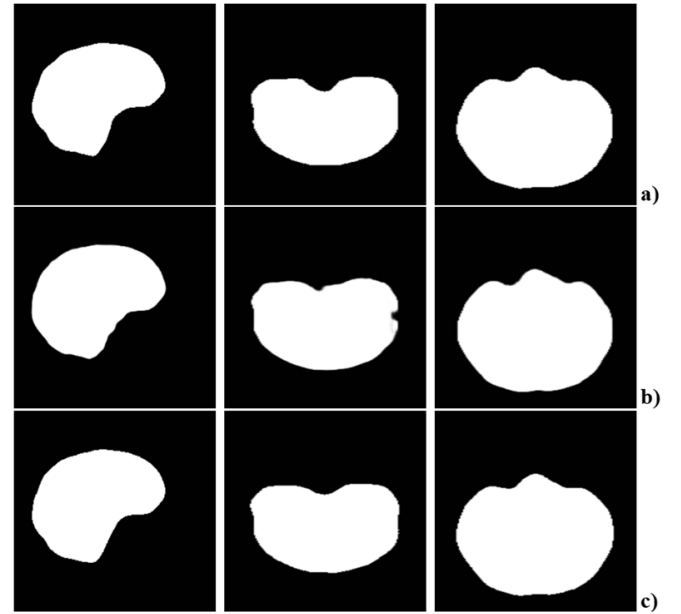


Fig. 3. Binary brain masks from sagittal, coronal, and axial views: (a) Ground truth brain masks, (b) UNet++ model predictions without knowledge distillation, (c) UNet++ student model predictions trained via knowledge distillation.

(Figure 3c) closely match the ground truth masks, demonstrating smoother boundaries and fewer segmentation inaccuracies compared to the standard UNet++ outputs (Figure 3b).

Table 2 summarizes these observations quantitatively by comparing the segmentation performance of the standard UNet++ model, the proposed UNet++ student model, and a

TABLE II
SEGMENTATION PERFORMANCE COMPARISON OF UNET++ MODELS.

DL Model	Dice	Sensitivity	Specificity
UNet++	0.95997	0.95766	0.99827
UNet++ Student	0.97857	0.97478	0.99899
Hoopes et al.[23]	0.96700	-	-

recent method in terms of Dice, sensitivity, and specificity metrics [23]. The UNet++ student model achieved superior performance, with a Dice score of 0.97857, sensitivity of 0.97478, and specificity of 0.99899. These qualitative and quantitative results collectively confirm that the knowledge distillation approach significantly enhances segmentation accuracy, especially in improving boundary delineation and detection sensitivity for 3D brain extraction tasks.

This study demonstrates strong segmentation performance using a knowledge distillation framework with UNet++ models trained on T1w, T2w, and PDw MRI from the IXI dataset. Accurate brain extraction is critical in MRI, enabling both automated analysis and clinical interpretation to focus on relevant regions, particularly in noisy images or when detecting subtle lesions. Unlike traditional methods that require manual parameter tuning and are sensitive to variability, the proposed approach is robust and parameter-free. Knowledge distillation further allows a lightweight student model to achieve high accuracy at a lower computational cost, making it well-suited for practical use. Nevertheless, certain limitations should be considered, as discussed below.

Although multiple MRI contrasts were utilized, experiments were confined to a single dataset; therefore, further validation across different datasets is necessary to establish broader applicability and robustness. Additionally, the relatively large 3D input dimensions (256×288×288) used in this study might present challenges in terms of computational resource demands and practical deployment, particularly in clinical scenarios. The relatively small size of the dataset may also limit the statistical power and generalizability of the results. Future studies may address this by augmenting the dataset, for example, through synthetic image generation or advanced data augmentation techniques, to improve robustness. Furthermore, the proposed framework is inherently flexible and could be adapted to other MRI sequences, such as diffusion-weighted or FLAIR imaging, as well as to different neuroimaging modalities, including PET and functional MRI, by adjusting to their specific characteristics and spatial resolutions (e.g., 1 mm or 2 mm isotropic), thereby extending its clinical applicability. Nevertheless, given the modular nature of the proposed knowledge distillation framework, it is reasonable to suggest that extending this approach to other UNet-based architectures may also yield improved segmentation performance.

IV. CONCLUSION

In this study, we proposed a knowledge distillation-based DL model using UNet++ models for 3D brain extraction from MRI scans. Our results demonstrate that knowledge distillation effectively enhances segmentation accuracy, particularly at

brain boundaries, by transferring detailed anatomical knowledge from a high-capacity teacher model to a more efficient student model. The student model achieved superior Dice, sensitivity, and specificity scores compared to both a standard UNet++ model and recent literature results. These findings highlight the capability of knowledge distillation to improve segmentation performance while maintaining computational efficiency.

Despite the promising outcomes, further validation using additional MRI datasets is necessary to confirm the generalizability of the proposed approach. Future studies could explore integrating attention mechanisms or evaluating other UNet-based architectures within the proposed knowledge distillation framework to further enhance segmentation performance.

ACKNOWLEDGMENT

This work is supported by the Scientific Research Project Fund of Sivas Cumhuriyet University.

REFERENCES

- [1] P. Kalavathi and V. S. Prasath, "Methods on skull stripping of MRI head scan images—a review," *J. Digit. Imaging*, vol. 29, pp. 365–379, 2016.
- [2] R. de Boer *et al.*, "Accuracy and reproducibility study of automatic MRI brain tissue segmentation methods," *NeuroImage*, vol. 51, no. 3, pp. 1047–1056, 2010.
- [3] R. Karakis, K. Gurkahrman, G. D. Mitsis, and M. H. Boudrias, "Deep learning prediction of motor performance in stroke individuals using neuroimaging data," *J. Biomed. Inform.*, vol. 141, Art. no. 104357, 2023.
- [4] F. Isensee *et al.*, "Automated brain extraction of multisequence MRI using artificial neural networks," *Hum. Brain Mapp.*, vol. 40, no. 17, pp. 4952–4964, 2019.
- [5] M. E. Brummer *et al.*, "Automatic detection of brain contours in MRI datasets," *IEEE Trans. Image Process.*, vol. 12, no. 2, pp. 153–166, 1993.
- [6] D. W. Shattuck, S. R. Sandor-Leahy, K. A. Schaper, D. A. Rottenberg, and R. M. Leahy, "Magnetic resonance image tissue classification using a partial volume model," *NeuroImage*, vol. 13, no. 5, pp. 856–876, 2001.
- [7] D. W. Shattuck and R. M. Leahy, "BrainSuite: An automated cortical surface identification tool," *Med. Image Anal.*, vol. 6, no. 2, pp. 129–142, 2002.
- [8] S. M. Smith, "Fast robust automated brain extraction," *Hum. Brain Mapp.*, vol. 17, pp. 143–155, 2002.
- [9] M. Jenkinson, M. Pechaud, and S. Smith, "BET2 - MR-based estimation of brain, skull and scalp surfaces," Oxford Centre for Functional Magnetic Resonance Imaging of the Brain (FMRIB), Oxford, 2005.
- [10] M. Jenkinson, C. F. Beckmann, T. E. Behrens, M. W. Woolrich, and S. M. Smith, "FSL," *NeuroImage*, vol. 62, no. 2, pp. 782–790, 2012.
- [11] A. M. Dale, B. Fischl, and M. I. Sereno, "Cortical surface-based analysis: I. Segmentation and surface reconstruction," *NeuroImage*, vol. 9, no. 2, pp. 179–194, 1999.
- [12] B. Fischl, "Freesurfer," *NeuroImage*, vol. 62, no. 2, pp. 774–781, 2012.
- [13] S. F. Eskildsen *et al.*, "BEaST: Brain extraction based on non-local segmentation technique," *NeuroImage*, vol. 59, no. 3, pp. 2362–2373, 2012.
- [14] R. Souza *et al.*, "An open, multi-vendor, multi-field-strength brain MR dataset and analysis of publicly available skull stripping methods agreement," *NeuroImage*, vol. 170, pp. 482–494, 2018.
- [15] S. K. Warfield, K. H. Zou, and W. M. Wells, "Simultaneous truth and performance level estimation (STAPLE): An algorithm for the validation of image segmentation," *IEEE Trans. Med. Imaging*, vol. 23, no. 7, pp. 903–921, 2004.
- [16] F. Segonne *et al.*, "A hybrid approach to the skull stripping problem in MRI," *NeuroImage*, vol. 22, no. 3, pp. 1060–1075, 2004.
- [17] M. Yapici, R. Karakis, and K. Gurkahrman, "Improving brain tumor classification with deep learning using synthetic data," *Comput. Mater. Continua*, vol. 74, no. 3, pp. 5049–5067, 2023.
- [18] N. Siddique, S. Paheding, C. P. Elkin, and V. Devabhaktuni, "U-Net and its variants for medical image segmentation: A review of theory and applications," *IEEE Access*, vol. 9, pp. 82031–82057, 2021.

- [19] J. Kleesiek *et al.*, “Deep MRI brain extraction: A 3D convolutional neural network for skull stripping,” *NeuroImage*, vol. 129, pp. 460–469, 2016.
- [20] H. Hwang, H. Z. U. Rehman, and S. Lee, “3D U-Net for skull stripping in brain MRI,” *Appl. Sci.*, vol. 9, no. 3, Art. no. 569, 2019.
- [21] Q. Zhang *et al.*, “FRNET: Flattened residual network for infant MRI skull stripping,” *arXiv preprint arXiv:1904.05578*, 2019.
- [22] K. Gurkahraman and C. Dasgin, “Brain extraction from magnetic resonance images using UNet modified with residual and dense layers,” *Türk Doga ve Fen Dergisi*, unpublished.
- [23] A. Hoopes, J. S. Mora, A. V. Dalca, B. Fischl, and M. Hoffmann, “SynthStrip: Skull-stripping for any brain image,” *NeuroImage*, vol. 260, Art. no. 119474, 2022.
- [24] X. Feng, C. Wang, S. Cheng, and L. Guo, “Automatic liver and tumor segmentation of CT based on cascaded U-Net,” in *Proc. Chin. Intell. Syst. Conf.*, 2019, pp. 155–164.
- [25] Y. Ding *et al.*, “A stacked multi-connection simple reducing net for brain tumor segmentation,” *IEEE Access*, vol. 7, pp. 104011–104024, 2019.
- [26] S. K. R. Chinnam, V. Sistla, and V. K. K. Kolli, “Multimodal attention-gated cascaded U-Net model for automatic brain tumor detection and segmentation,” *Biomed. Signal Process. Control*, vol. 78, Art. no. 103907, 2022.
- [27] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, “UNet++: A nested U-Net architecture for medical image segmentation,” in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, ser. Lecture Notes in Computer Science, vol. 11045, Springer, 2018, pp. 3–11.
- [28] G. Li, K. Wang, P. Lv, *et al.*, “Multistage feature fusion knowledge distillation,” *Scientific Reports*, vol. 14, article no. 13373, 2024.