Taylor and Francis Book Chapter

LOC Page

Contents

Ensemble-Based Transfer Learning Approach for Brain Tumor Segmentation from MRI Images

A. F. M. Minhazur Rahman

Rajshahi University of Engineering & Technology, Dept. of CSE, Rajshahi, Bangladesh. Email: m.r.saurov@gmail.com **(Corresponding Author)**

Md. Ali Hossain

Rajshahi University of Engineering & Technology, Dept. of CSE, Rajshahi, Bangladesh. Email: ali.ruet@gmail.com

CONTENTS

M agnetic Resonance Imaging (MRI) is a medical imaging technique frequently used to visualize detailed anatomical structures in the brain, making it an indispensable tool for the detection and characterization of brain tumors. Due to the high variability and complexity of tumor structures, accurate segmentation of brain tumors from MRI images remains a challenging task. To tackle this challenge, this research introduces a transfer learning strategy that incorporates an ensemble tech-

nique. State-of-the-art pre-trained convolutional neural networks, such as Efficient-Net and DenseNet, are leveraged as encoders in a modified U-Net architecture and fine-tuned on a publicly available dataset of brain tumor MRI scans. The fine-tuning process adapts the models to the specific task of brain tumor segmentation. To further enhance performance, an ensemble approach is employed that combines the outputs from multiple models through majority voting and weighted voting strategies. Experiments demonstrate that the proposed combination of transfer learning and model ensemble approach achieves a Dice score of 91.96% and an IoU score of 85.15%, outperforming several existing methods, indicating that the proposed method holds great promise for enhancing the accuracy and reliability of brain tumor segmentation in clinical practice.

1.1 INTRODUCTION

Magnetic Resonance Imaging is a non-invasive method that utilizes a magnetic field and radio waves to obtain comprehensive images of the human brain, abdomen, breast etc. The MRI sensor provides excellent contrast for soft tissues and spatial resolution with no known health risks. These benefits of MRI lend to growing importance of semantic segmentation. Semantic segmentation of brain MRI images involves separating healthy tissue from abnormal tissues by classifying each individual pixel in the image. Fine-grained examination of tissue structure using segmented MRI can detect brain tumors, Alzheimer's disease, Sclerosis, etc.[1].

A brain tumor is defined as the atypical growth of brain cells, regardless of whether they are cancerous. Accurate segmentation of brain tumors in MRI images is vital for early diagnosis, precise prognosis, and effective treatment planning. The current limitations in segmentation techniques often lead to suboptimal clinical outcomes, emphasizing the need for continuous advancements in this field. With the rapid growth of medical imaging data, it is essential to develop reliable and efficient segmentation methods that can cope with the inherent complexities and heterogeneities of tumor structures. Improved segmentation techniques can provide clinicians with better decision-making tools and contribute to the overall advancement of patient care.

Classical methods for semantic segmentation of MRI images include pixel classification methods thresholding-based methods, and region-based methods[2]. However, classical methods require laborious feature engineering and still yield suboptimal results. The emergence of deep learning and its capacity for automatic representation learning has enabled efficient feature extraction from radiological images. Due to their ability at extracting valuable features from image data, deep learning approaches based on convolutional neural networks (CNNs) have gained traction for computer vision tasks such as classification, object detection, segmentation, etc.

Fully Convolutional Networks (FCNs)[3] were among the first architectures proposed for semantic segmentation tasks. FCNs replaced the fully connected layers in standard CNNs with convolutional layers, allowing them to handle input images of varying sizes and output spatially coherent segmentation maps. This pioneering work laid the foundation for subsequent developments in semantic segmentation.

Inspired by the success of FCNs, U-Net[4], a prevalent CNN-based architecture,

was proposed specifically for the semantic segmentation of biomedical images. U-Net retains the core idea of using convolutional layers for feature extraction and upsampling for generating the final segmented mask, similar to FCNs. However, U-Net introduces a significant innovations in the form of skip connections from the encoder to the decoder and larger number of feature maps in upsampling part. Skip connections facilitate the fusion of high-level feature maps with low-level ones, enabling accurate and detailed segmentation. Buda *et al.*[5] implemented the U-Net architecture for skull stripping and segmentation, integrating four levels of blocks within the encoder segment. Each block comprises convolutional layers with Rectified Linear Unit activation functions and a max-pooling layer for downsampling. The decoder portion utilizes up-convolutional layers for upsampling purposes.

Despite being a state-of-the-art segmentation method upon its introduction, several modifications have been proposed to enhance the standard U-Net architecture's efficacy. Yi *et al.*[6] presented an improved U-Net called SU-Net (Strong U-Net), incorporating the U-Net's encoder-decoder structure, adding a dense block to augment depth, and an inception module to widen the network. Ghosh *et al.*[7] proposed a U-Net modification by leveraging pre-trained VGG-16 model combined with dense blocks[8] as a feature extractor to extract high-level features from input images. These features were then concatenated with corresponding feature maps in the U-Net architecture's decoder path, enhancing performance. The authors demonstrated that this approach achieved superior pixel accuracies and outperformed state-of-the-art methods.

It is apparent from the literature that modification of the U-Net feature extractor part can lead to performance improvement. Using a pre-trained feature extractor has shown a gain in computational efficiency and better feature representation[9]. This has led to the increasing adoption of transfer learning in the MRI analysis domain[10]. However, most segmentation methods in the MRI analysis domain still rely on older pre-trained models, which may not fully capture the intricacies and nuances of the training dataset.

This study presents a technique that utilizes cutting-edge pre-trained CNN models, fine-tuned on a particular dataset, to serve as hierarchical feature extractors within the encoder component of the U-Net architecture for segmenting brain tumors. Goal of this approach is to capitalize on the knowledge acquired from the pre-trained models to enhance segmentation performance. The standard U-Net decoder is utilized for upsampling the encoded features to generate a segmentation mask that accurately delineates the target tumor regions.

To further improve the segmentation results, ensemble methods were employed to aggregate the outcomes from several state-of-the-art models. By incorporating the predictions of multiple models, the ensemble technique boosts the overall robustness and accuracy of the final tumor segmentation output. This combination of transfer learning in the encoder portion of the U-Net and ensemble methods presents a powerful approach to tackle challenging brain tumor segmentation tasks, potentially leading to better performance and generalization across diverse datasets.

1.2 METHODOLOGY

In this section, the discussion begins with the dataset utilized for binary semantic segmentation and the data preparation methods aimed at enhancing model performance and generalizability. Next, the modified U-Net architecture is introduced, detailing its structure and functionality. The discussion then transitions to incorporating transfer learning through various cutting-edge pre-trained models as encoders within the U-Net framework and examines the fine-tuning process, which refines the encoder's final layers to achieve better results. Lastly, an overview of the ensemble strategy is given, which combines predictions from multiple models to improve the overall performance and robustness of the final segmentation output.

1.2.1 Dataset and Data Preparation

This research work is focused on detecting low-grade glioma tumorous cells in MRI images. Glioma is a common brain tumor, that exhibits properties of benign tumors[11]. We used the TCGA-LGG Segmentation dataset[5] for our research. It consists of 3929 brain tumor images and corresponding FLAIR abnormality segmentation masks obtained from 110 patients.

The images are RGB and have dimensions of $256 \times 256 \times 3$. Corresponding masks are single-channel images with dimensions of $256 \times 256 \times 1$, with each pixel having a value of either 0 or 255. A pixel value of 0 indicates a non-tumorous pixel, while 255 indicates a tumorous pixel. Figure 1.1 shows a sample from this dataset.

Images from TCGA-LGG were partitioned into an 80% training and a 20% test set respectively. Additionally, 20% of the training set was allocated for validation purposes. The validation set was employed for hyperparameter tuning and model selection.

The TCGA-LGG dataset contains a relatively small number of images, with only a fraction of them having tumors present. To overcome this limitation, data augmentation techniques are employed to generate additional training samples. Random variations in the images, such as rotation, width and height shifts, and zooming, are introduced to simulate realistic changes that could occur in real-world scenarios. By augmenting the training set, a more diverse and robust dataset is created that enhances the model's ability to generalize to new and unseen data while minimizing the risk of overfitting.

1.2.2 U-Net Architecture

U-Net, a fully convolutional neural network structure, has been developed for medical segmentation purposes. The symmetric, U-shaped architecture consists of an encoding pathway and a decoding pathway, linked by a bottleneck layer. The encoding pathway extracts spatial features and context through convolutional layers, ReLU activation functions, and max-pooling procedures. The bottleneck layer gathers high-level semantic data, bridging the encoder and decoder pathways. The decoding pathway restores spatial resolution by implementing upsampling processes, concatenating with corresponding feature maps from the encoding pathway, and using convolutional
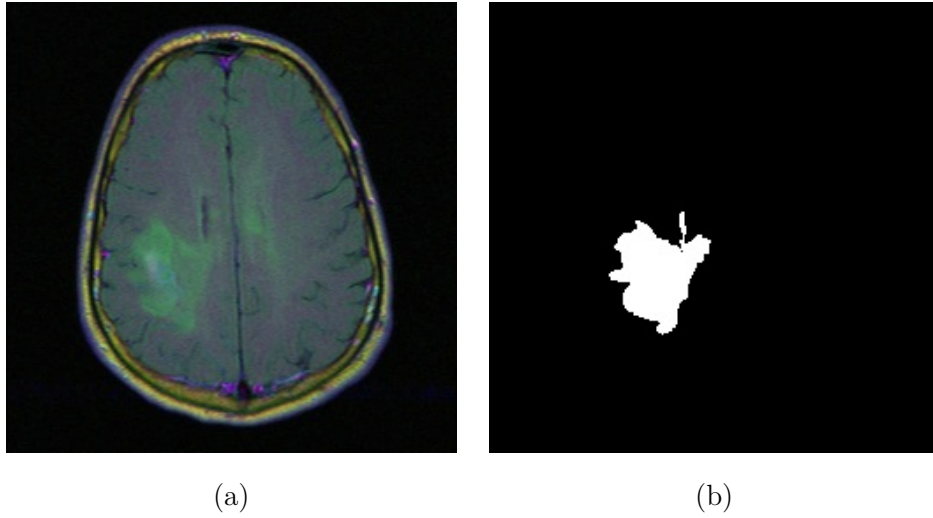
<div align="center">(a)   (b)</div>

Figure 1.1: A sample from TCGA-LGG dataset. (a) MRI image; (b) Ground truth mask.

layers with ReLU activations. Skip connections between the encoding and decoding paths help preserve fine-grained details and reduce spatial information loss[4]. A final 1×1 convolution layer with a softmax (in case of multi-class segmentation) or sigmoid (in case of binary segmentation) activation assigns pixel-wise class probabilities for each label, resulting in a segmented image produced by combining local and global contextual information learned by the network. Standard U-Net architecture is shown in Figure 1.2.

Since, this research is focused on identifying a single class— low-grade grade glioma tumorous cells from input images, the final convolution layer uses sigmoid activation function. Sigmoid activation outputs the individual probabilities of pixels being tumorous or non-tumorous. A threshold of 0.5 converts the probability score to either 0 (non-tumorous pixel) or 1 (tumorous pixel).

The conventional U-Net encoder possesses a restricted receptive field, implying that it encounters difficulties in capturing long-range contextual information. More advanced backbones[8, 12, 13] have deeper architectures and larger receptive fields, enabling them to capture more global context. Advanced backbones often incorporate mechanisms to improve gradient flow during training, such as skip connections, residual connections, or densely connected layers. These mechanisms can help alleviate the vanishing gradient problem, allowing for deeper architectures and more stable training. For these reasons, transfer learning and fine-tuning approach to semantic segmentation has potential for performance improvement.

1.2.3 Transfer Learning & Fine Tuning

Transfer learning is an effective technique that leverages models trained on expansive and diverse datasets such as ImageNet to improve performance on new, related tasks such as our low-grade glioma MRI segmentation task.
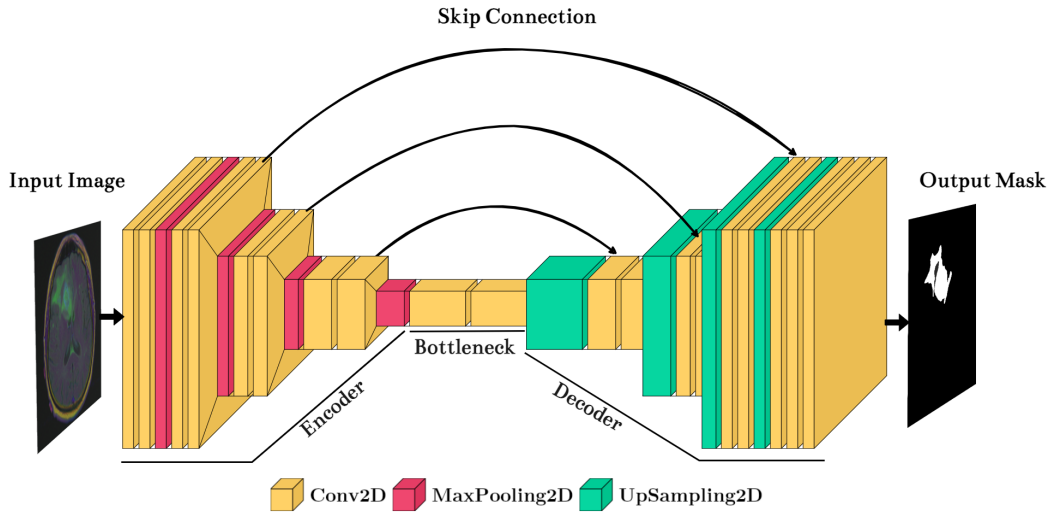
Figure 1.2: Standard U-Net architecture for semantic segmentation. Encoder portion is the contracting path (reduction in spatial dimension, expansion of channel dimension) in the network. Bottleneck layers connect the encoder to expanding (expansion in spatial dimension, reduction in channel dimension) decoder portion of the network.

The evolution of CNNs for computer vision began with VGG[14], which used deeper architectures and small filters to learn complex features. ResNet[12] introduced residual connections to enable even deeper networks without vanishing gradients. DenseNet[8] further connected layers densely for better feature reuse and reduced parameters. EfficientNet[13] optimized depth, width, and resolution using compound scaling, resulting in more accurate and efficient models. In the context of the semantic segmentation task, these architectures will be employed to modify the encoder portion of the U-Net architecture. The decoder portion will remain as described in Section 1.2.2.

Rather than fully training the modified U-Net architecture from randomly initialized weights, ImageNet weights of pre-trained backbones will be used as a starting point for model parameters. Utilizing a pre-trained encoder ensures faster convergence since the encoder has already learned generic features for visual recognition tasks. Another advantage of using these stable weights is the reduced amount of labeled data required for training, which is particularly beneficial for the TCGA-LGG segmentation dataset due to its small size.

Another consideration in transfer learning is whether to fully freeze the encoder or not. Fully freezing the encoder weights ensures the full preservation of learned features and significantly reduces computational challenges associated with training deep learning models from scratch, thus saving time and resources. However, unfreezing the last few encoder blocks or stages before the bottleneck stage can lead to improved performance[15]. This process is called *fine-tuning*. In the initial layers of a CNN, low-level image features like edges and corners are commonly learned, which are beneficial for many computer vision tasks, encompassing medical image segmen-
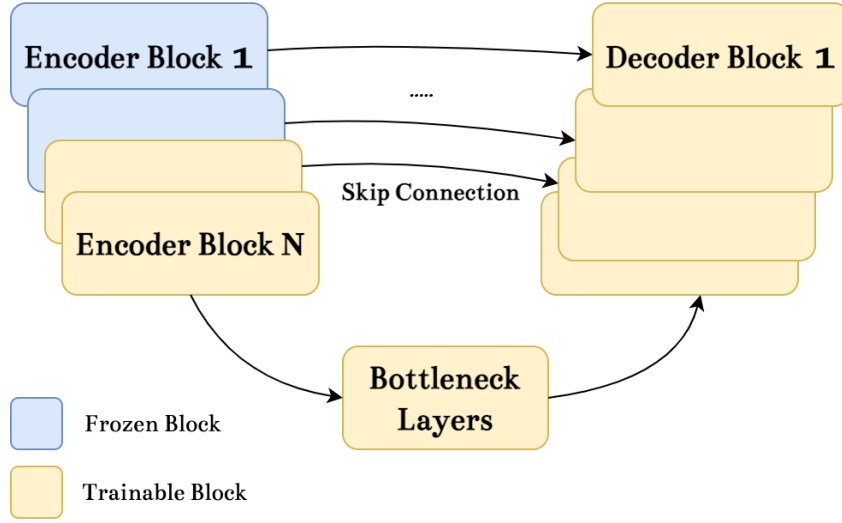
Figure 1.3: Fine-tuning in pre-trained encoder backbone based U-Net architecture. Final layers before the bottleneck layers are made trainable for learning high level features specific to LGG segmentation domain.

tation. In contrast, the later layers learn high-level features that ensure better domain adaptation[9]. Figure 1.3 shows the process of fine-tuning in U-Net architecture. Table 1.1 lists the models used as encoder for U-Net architecture.

Table 1.1: Models used for U-Net encoder and trainable blocks/stages for fine-tuning.

Family	Models	Trainable Blocks/Stages
EfficientNet	EfficientNetB0 to B7	Block 30 to 32
DenseNet	DenseNet169, DenseNet201	Block 7
ResNet	ResNet18, ResNet50, ResNet101, ResNet152	Stage 4
VGG	VGG16, VGG19	Block 5

1.2.4 Ensemble-based Approach

Leveraging the complementary strengths of different backbone architectures using ensemble models can mitigate model-specific weaknesses and provide more robust and accurate segmentation results. This is especially helpful for segmenting brain tumours, where the accurate delineation of tumor boundaries can greatly benefit clinical decision-making and treatment planning. Ensemble methods used in this research work are described in detail below.

- **Majority Voting Ensemble:** The first ensemble method we employed is majority voting, a simple yet effective technique for aggregating the predictions of

multiple models. For n models, final segmentation mask is computed by taking the mode value of the corresponding pixels in the predicted masks:

$$M_{final}(i,j) = \text{mode}\{M_1(i,j), M_2(i,j), ..., M_n(i,j)\} \tag{1.1}$$

Where $M_{final}(i,j)$ is the final segmentation mask and $M_k(i,j)$ is the segmentation mask produced by the k-th model, with $k \in 1, 2, ..., n$.

- **Weighted Ensemble:** Another approach used for improved segmentation is the weighted ensemble method. In this method, weights are assigned to the predictions of each model based on their individual performance on the validation set. The final segmentation mask is obtained by taking the weighted average of the predicted masks:

  For the weighted ensemble method with n models, Taking the weighted average of the predicted masks produces the final segmentation mask:

$$M_{final}(i,j) = \sum_{k=1}^{n} w_k \cdot M_k(i,j) \tag{1.2}$$

Where w_k is the weight assigned to the k-th model and $M_k(i,j)$ is the segmentation mask produced by the k-th model. Weights are assigned to the predictions of each model based on their individual IoU performance on the validation set.

$$w_k = \frac{IoU_k}{\sum_{l=1}^{n} IoU_l} \tag{1.3}$$

1.3 EXPERIMENTAL ANALYSIS

This section discusses the metrics used to evaluate the performance of the proposed approach in this research work, aiming to highlight its improvements. Additionally, important model hyperparameters for training are provided. Subsequently, the results from several models are compared, and the best models are selected for ensemble. The ensemble of models is then compared to other existing methods.

1.3.1 Evaluation Metrics

In the context of segmentation tasks, relying solely on accuracy, i.e. whether a pixel is labeled correctly, as a performance metric can be misleading due to its inherent limitations. Without taking into account the distribution of various classes or the spatial context of the predictions, accuracy counts the percentage of properly categorized pixels in the image. In this segmentation task, the classes are imbalanced, with background class (healthy tissues) significantly outnumbering the tumor class.

Furthermore, accuracy does not capture the overlap between the predicted segmentation mask and the actual mask, which is crucial for assessing the quality of prediction. Metrics such as IoU and Dice score, on the other hand, are designed to

evaluate the similarity between the predicted mask and actual mask by considering the spatial overlap between them.

Moreover, the accuracy metric fails to represent the degree of overlap between the model estimated segmentation mask and the true mask, an essential aspect for evaluating prediction quality. In contrast, measures like IoU and Dice score aim to assess the resemblance between the predicted and actual masks by taking into account the spatial overlap between them.

- **Intersection over Union (IoU):** IoU score, also referred to as the Jaccard index, assesses the overlap between predicted and ground truth segmentations. The IoU score is determined by ratio between intersection area and union area of the two segmentations. The formula for IoU is given by:

$$\text{IoU Score} = \frac{\text{Predicted Mask} \cap \text{Ground Truth Mask}}{\text{Predicted Mask} \cup \text{Ground Truth Mask}} \tag{1.4}$$

- **Dice Coefficient:** Dice coefficient, commonly referred to as the Dice score or F1 score, is another popular tool for assessing segmentation performance. It is computed as the harmonic mean of precision and recall, or equivalently, as twice the area of intersection divided by the sum of the areas of the two segmentations. The formula for the Dice coefficient is:

$$\text{Dice Coefficient} = \frac{2(\text{Predicted Mask} \cap \text{Ground Truth Mask})}{\text{Predicted Mask} + \text{Ground Truth Mask}} \tag{1.5}$$

IoU and Dice coefficient both have a range of 0 to 1, where higher values signify superior segmentation performance.

- **Dice Loss:** The Dice coefficient ranges from 0 to 1. To train a deep network responsible for segmenting input images with the maximum possible Dice score, Dice loss is formulated such that a low Dice score penalizes the parameters of the neural network during the training process. This is simply achieved by the following formula:

$$\text{Dice Loss} = 1 - \text{Dice Score} \tag{1.6}$$

1.3.2 Implementation Details

All models were trained using an NVIDIA A6000 GPU. Model architectures were implemented with the help of Keras deep learning library. The models underwent training for 120 epochs, utilizing the Adam optimizer. Selected batch size was 32. To minimize overfitting and decrease training time, early stopping was implemented with a patience threshold of 20. Dice loss was selected as the loss function. Best performing model parameters were selected based on maximum validation Dice score via model checkpointing.
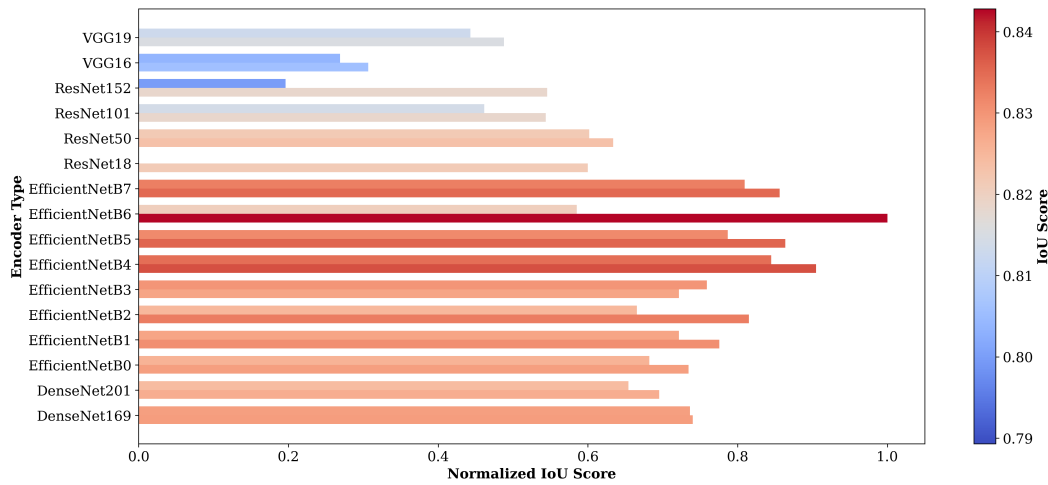
Figure 1.4: Min-Max normalized validation IoU scores of trained models. Each group of bars represents the performance of fully frozen (top bar) and fine-tuned (bottom bar) encoders.

1.3.3 Result Analysis & Discussion

A total of 32 models were trained on the TCGA-LGG segmentation dataset, with 16 models trained without fine-tuned encoder and 16 models trained with fine-tuned encoder. Figure 1.4 illustrates the performance difference observed between these models. Analyzing the results in Figure 1.4 and Table 1.2, following trends can be observed:

(i) The EfficientNet family of encoders consistently outperforms the other encoder types across various architectures. Among them, EfficientNetB6 with trainable layers achieved the highest IoU score of 84.28%.

(ii) The DenseNet family also demonstrates good performance, with DenseNet169 and DenseNet201 showing similar results. In both cases, the trainable versions outperform the fully frozen versions, albeit by a very small margin.

(iii) The ResNet family displays varying performance, with ResNet50 achieving the best IoU scores among the group. The fully frozen versions consistently underperform compared to their fine-tuned counterparts.

(iv) The VGG family exhibits the lowest overall performance among the tested encoder types. Both VGG16 and VGG19 show better results when their final layers are trainable compared to their fully frozen counterparts.

EfficientNet models, with their balanced scaling of network width, depth, and resolution, excel in performance. This excellence is further amplified by fine-tuning which refines the pre-trained encoder parameters with TCGA-LGG dataset, thereby improving its predictive performance. Following EfficientNet, DenseNet stands out

Table 1.2: Best model from each family and its performance (in percentage).

Encoder Family	Best Model	IoU	Dice
DenseNet	DenseNet169 (Fine-tuned)	82.89	90.65
EfficientNet	EfficientNetB6 (Fine-tuned)	84.28	91.47
ResNet	ResNet50 (Fine-tuned)	82.32	90.30
VGG	VGG19 (Fine-tuned)	81.54	89.83

with its densely connected architecture that enables efficient gradient flow and feature reuse. Just like with EfficientNet, fine-tuning plays a crucial role in amplifying these inherent advantages of DenseNet, aligning the model more closely with the specific characteristics of the segmentation task and thus boosting its performance.

On the other hand, ResNet, which employs residual connections to address the vanishing gradient problem common in deep networks, may not be as efficient as DenseNet in feature reuse. Despite this, fine-tuning helps make the ResNet model more task-specific, leading to an improvement in its performance, albeit not entirely bridging the gap with DenseNet.

Lastly, VGG, despite its simpler architecture, struggles to match the performance of the other models due to its limited ability to adapt to the segmentation task and capture complex features. While fine-tuning does manage to improve VGG's performance, it isn't enough to compensate for the architectural limitations, resulting in a lower performance compared to the other models.

1.3.3.1 Model Ensemble

Based on validation IoU and Dice score, majority voting and weighted ensemble method were applied by selecting top 5 models.

Table 1.3: IoU and Dice scores (in percentage) of Top 5 Models and their ensemble.

Encoder	Val IoU	Val Dice	Test IoU	Test Dice
EfficientNetB6 (Fine-tuned)	84.28	91.49	84.05	91.32
EfficientNetB4 (Fine-tuned)	83.77	91.13	83.91	91.26
EfficientNetB5	83.55	91.01	83.77	91.17
EfficientNetB7 (Fine-tuned)	83.51	90.99	83.72	91.13
EfficientNetB4	83.45	90.94	84.12	91.38
Majority Voting Ensemble	85.05	91.88	85.13	91.95
Weighted Ensemble	**85.07**	**91.89**	**85.15**	**91.96**

In Table 1.3, it can be observed that EfficientNetB6 encoder based U-Net has the highest Validation and Test IoU scores among the individual models, indicating that it performs best on its own. Furthermore, both the weighted ensemble and majority voting ensemble of top 5 models show improvements in Validation and Test IoU scores compared to the individual models, demonstrating the effectiveness of

<div align="center">(a) MRI Image   (b) Ground Truth   (c) Predicted Mask</div>
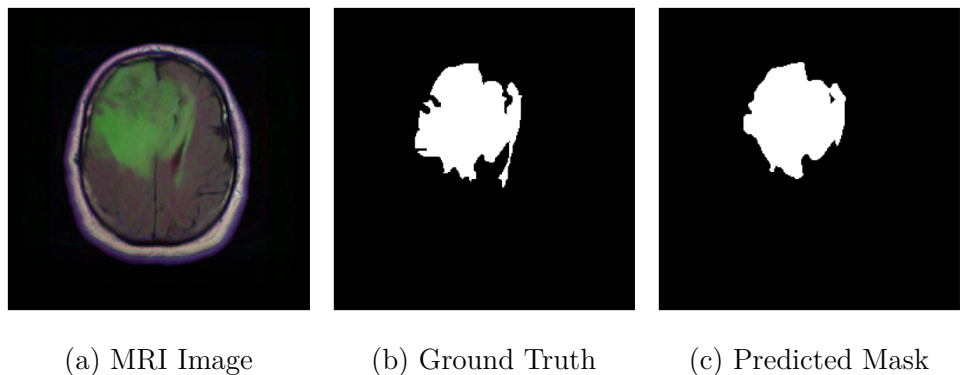
Figure 1.5: Segmentation of MRI image using weighted ensemble of top 5 models. IoU score and Dice score for this sample is 87.04%, 93.07% respectively.

ensemble techniques in this case. The weighted ensemble method slightly outperforms the majority voting ensemble method in both validation and test scores, suggesting it may be the preferred ensemble technique for these models. Figure 1.5 shows the mask prediction of the proposed method on a sample MRI image from the test set.

1.3.4 Comparison with Other Methods

The proposed approach was compared with three other approaches that also experimented with the TCGA-LGG dataset.

Table 1.4: Comparison between various approached on TGCA-LGG dataset.

Authors	Approach	Dice (%)
Buda et. al[5]	Standard U-Net Architecture	82.00
Yi et. al[6]	Strong U-Net with inception module and dense block	78.50
Ghosh et. al[7]	Improved U-Net with VGG encoder and dense-convolutional blocks	91.81
Proposed	**Ensemble of fine-tuned encoder U-Nets**	**91.96**

Table 1.4 compares the Dice scores of various approaches applied to the TCGA-LGG dataset. Buda *et al.* achieved a Dice score of 82.00% using a standard U-Net architecture, while Yi *et al.* utilized a Strong U-Net with inception module and dense block, resulting in a 78.50% Dice score. Ghosh *et al.* significantly improved upon these scores, achieving a 91.81% Dice score by incorporating a pre-trained VGG encoder and dense-convolutional blocks within an improved U-Net. The proposed approach of this research work, which employs an ensemble of fine-tuned encoder U-Nets, further enhances the Dice score to 91.96%, demonstrating the effectiveness of the combination of transfer learning and ensemble method in achieving superior performance.

1.4   CONCLUSION AND FUTURE WORK

This research demonstrates the effectiveness of employing transfer learning, fine-tuning, and ensemble techniques on various pre-trained models in the context of semantic segmentation of brain tumors. By leveraging the U-Net architecture and adapting the last layers of the encoder, improved segmentation results were achieved across a range of model families. The application of weighted ensemble further enhanced the overall performance and robustness of the segmentation output, providing a strong case for the use of these techniques in similar tasks.

The successful implementation of this methodology suggests that it has the potential to be applied to other datasets and tasks beyond binary semantic segmentation, potentially leading to improved results in a variety of applications. In future research, the aim is to explore the generalizability of this approach by testing it on different datasets and domains. Additionally, investigating alternative ensemble techniques and model architectures to refine the methodology further is planned. Overall, this study provides valuable insights and a solid foundation for future work in the field of semantic segmentation and related areas.

Bibliography

[1] Yamanakkanavar, N., Choi, J. & Lee, B. MRI segmentation and classification of human brain using deep learning for diagnosis of alzheimer's disease: a survey. *Sensors.* **20**, 3243 (2020)

[2] Lefkovits, S., Lefkovits, L. & Szilágyi, L. HGG and LGG brain tumor segmentation in multi-modal MRI using pretrained convolutional neural networks of Amazon sagemaker. *Applied Sciences.* **12**, 3620 (2022)

[3] Long, J., Shelhamer, E. & Darrell, T. Fully convolutional networks for semantic segmentation. *Proceedings Of The IEEE Conference On Computer Vision And Pattern Recognition.* pp. 3431-3440 (2015)

[4] Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. *Medical Image Computing And Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18.* pp. 234-241 (2015)

[5] Buda, M., Saha, A. & Mazurowski, M. Association of genomic subtypes of lower-grade gliomas with shape features automatically extracted by a deep learning algorithm. *Computers In Biology And Medicine.* **109** pp. 218-225 (2019)

[6] Yi, L., Zhang, J., Zhang, R., Shi, J., Wang, G. & Liu, X. SU-Net: an efficient encoder-decoder model of federated learning for brain tumor segmentation. *Artificial Neural Networks And Machine Learning–ICANN 2020: 29th International Conference On Artificial Neural Networks, Bratislava, Slovakia, September 15–18, 2020, Proceedings, Part I.* pp. 761-773 (2020)

[7] Ghosh, S., Chaki, A. & Santosh, K. Improved U-Net architecture with VGG-16 for brain tumor segmentation. *Physical And Engineering Sciences In Medicine.* **44**, 703-712 (2021)

[8] Huang, G., Liu, Z., Pleiss, G., Maaten, L. & Weinberger, K. Convolutional Networks with Dense Connectivity. *IEEE Transactions On Pattern Analysis And Machine Intelligence.* **44**, 8704-8716 (2022)

[9] Tajbakhsh, N., Shin, J., Gurudu, S., Hurst, R., Kendall, C., Gotway, M. & Liang, J. Convolutional neural networks for medical image analysis: Full training or fine tuning?. *IEEE Transactions On Medical Imaging.* **35**, 1299-1312 (2016)

[10] Valverde, J., Imani, V., Abdollahzadeh, A., De Feo, R., Prakash, M., Ciszek, R. & Tohka, J. Transfer learning in magnetic resonance brain imaging: A systematic review. *Journal Of Imaging.* **7**, 66 (2021)

[11] Wadhwa, A., Bhardwaj, A. & Verma, V. A review on brain tumor segmentation of MRI images. *Magnetic Resonance Imaging.* **61** pp. 247-259 (2019)

[12] He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. *Proceedings Of The IEEE Conference On Computer Vision And Pattern Recognition.* pp. 770-778 (2016)

[13] Tan, M. & Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. *International Conference On Machine Learning.* pp. 6105-6114 (2019)

[14] Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. *ArXiv Preprint ArXiv:1409.1556.* (2014)

[15] Mormont, R., Geurts, P. & Marée, R. Comparison of deep transfer learning strategies for digital pathology. *Proceedings Of The IEEE Conference On Computer Vision And Pattern Recognition Workshops.* pp. 2262-2271 (2018)