Paper Type: Original Article

Tumor Detection in MRI Data using Deep Learning Techniques for Image Classification and Semantic Segmentation

Muhammad Abid [1],* (iD) **and Muhammad Shahid [2]** (iD)

[1] Department of Mathematics, North Carolina State University, 27695 NC Raleigh, USA; mabid@ncsu.edu.

[2] Department of Physics and Astronomy, Georgia State University, Atlanta, GA 30303, USA; mshahid1@gsu.edu.

Abstract

This study aims to apply image classification and semantic segmentation methods using deep learning for accurate and efficient tumor detection in brain MRI data. It looks at ways to get over issues including interpretability of the model and a lack of annotated data. A pre-trained VGG19 model and a basic convolutional neural network (CNN) were trained using a public dataset of 273 brain MRI images for the purpose of image classification. Techniques for regularization, transfer learning, and data augmentation were used. A U-Net architecture was trained using manually created masks from the same dataset for semantic segmentation, with several training setups investigated. With a mean test accuracy of 77.19%, the pre-trained VGG19 model surpassed the basic CNN in picture categorization. The U-Net model performed better for semantic segmentation after training on the entire dataset for two epochs as opposed to training on a subset for ten epochs. The work acknowledges issues including data scarcity and the requirement for model interpretability and scalability, which call for additional research, but also underlines the potential of deep learning for brain tumor identification.

Keywords: Brain Tumor Detection; Deep Learning, Image Classification; Semantic Segmentation; Magnetic Resonance Imaging (MRI).

1 | Introduction

The amount of data that may be analyzed has significantly increased in the field of radiology due to the quick developments in medical imaging technologies like magnetic resonance imaging (MRI). But manually interpreting such large datasets can be laborious, subjective, and prone to mistakes, especially when identifying and segmenting intricate structures such as brain tumors. Thankfully, computer vision, fueled by deep learning methods, has become a potent instrument to support automated medical image analysis, providing precise and effective answers for a range of tasks, such as segmentation and picture classification [1].

One of the fundamental tasks in computer vision is image classification, which is to group images according to their content into predetermined classes. This makes necessary to separating tumorous from non-tumorous pictures in the context of brain MRI analysis, that might act as an start screening step in the diagnostic process [2]. Precisely built for image data, Convolutional Neural Networks (CNNs) represent a family of deep learning

models that have shown remarkably effective in picture classification tasks, exceptionally conventional machine learning techniques for these models [3].

Semantic segmentation, which divides an image into several segments, each of which represents a different object or region of interest, is one of the most well-known uses of computer vision in medical imaging [4]. Semantic segmentation can accurately define the boundaries of tumors in brain MRI imaging, facilitating accurate quantification, localization, and additional analysis - all essential for treatment planning and follow-up monitoring [5].

Although deep learning methods have demonstrated encouraging outcomes in a range of medical imaging applications, their effectiveness is highly reliant on the availability of substantial, varied, and meticulously annotated datasets [6]. Nevertheless, gathering and annotating such datasets can be difficult, costly, and time-consuming in many real-world situations, especially in the medical field where proper annotations need specialist expertise [7].

Data augmentation is a well-researched strategy to address the problem of data scarcity. It entails creating synthetic data by transforming the available dataset in different ways [8]. The training data can be effectively made more diverse and larger by applying various transformations, such as rotation, flipping, scaling, and introducing noise [9]. utilizing transfer learning, that involves refining a pre-trained model on the target dataset after it has been trained on a sizable dataset from a related domain, is another successful strategy [10]. In view of the fact that the pre-trained model already captures pertinent properties that can be moved and improved for the target job, this course of action has been found to improve performance, particularly when dealing with limited type of the data [11].

As the weight decay and dropout are two most of the important regularization approaches which are also essential for avoiding from the overfitting, that is a typical kind of the problem when training deep learning models on short datasets [12]. The process of regularization improves the model's resilience and capacity for generalization by adding more restrictions or randomness during the training process [13]. So, these kinds of methods serve as a kind of control mechanism, striking a balance among the model's intricacy and its strong generalization to new data. Also, they encourage the model to learn more outstanding patterns by keeping it from overly remembering the training set. The procedure of the regularization that promotes a more consistent and dependable performance, especially in conditions with limited or noisy data.

One of the most widely adopted models is the U-Net architecture has been a mainstay in medical image analysis in recent years, especially for brain tumor identification and segmentation. The model's novel design, which contains the encoder and the decoder routes, allows it to recognize fine details in medical pictures while preserving spatial information that is necessary for accurate tumor delineation [14]. So, the U-Net productively captures hierarchical characteristics by implementing its convolutional layers, enabling reliable tumor localization. Additionally, it is a favored option in medical situations where data availability is frequently constrained due to its flexibility to limited training data [15]. The U-Net is a dependable tool that medical practitioners can rely on as research advancement, contributing prospective development in the early identification and management of brain tumors [16-17].

Although, a number of issues, together with the intrinsic complexity of brain architecture, the diversification of tumor appearance, and the existence of imaging artifacts, make it very complex to reliably detect and segment brain tumors using MRI data [18-19]. In addition to this, the development and assessment of robust computational algorithms is further impeded by the absence of publicly accessible, standardized datasets [20-24].
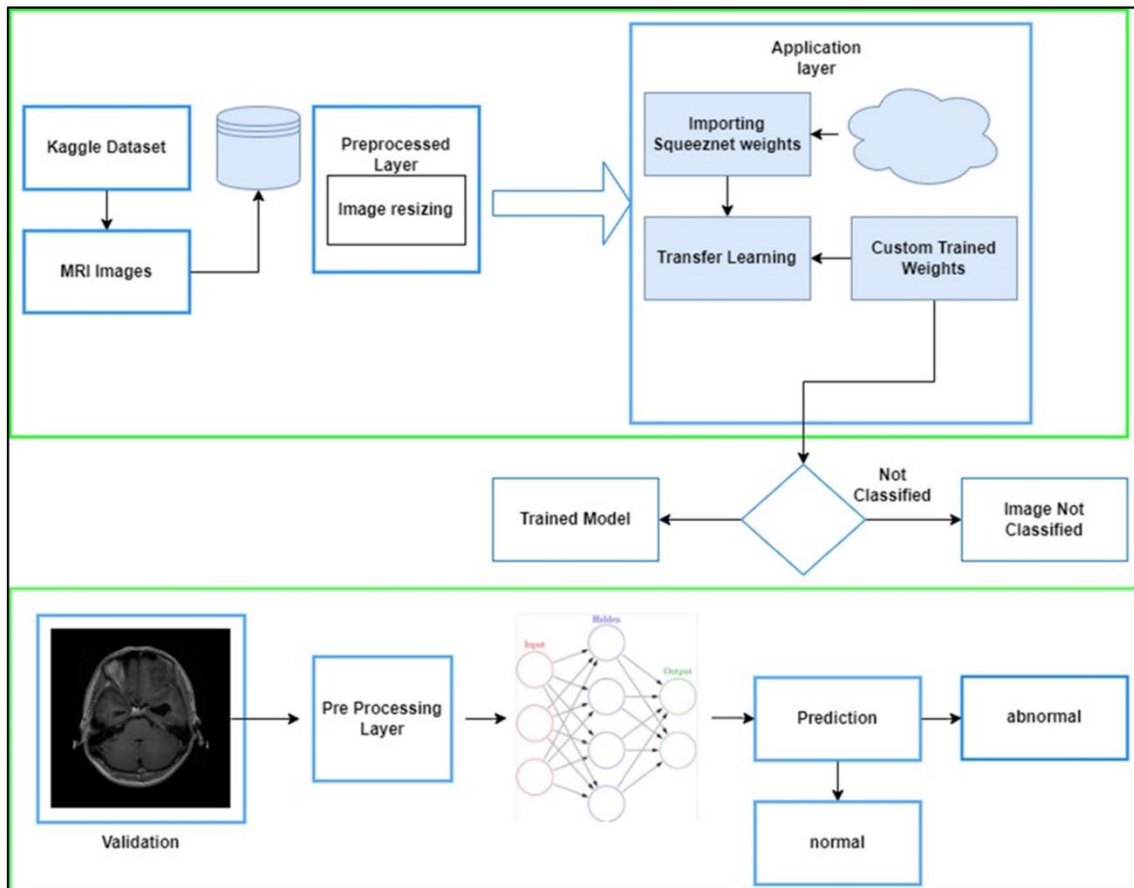
Figure 1. A comprehensive illustration of the suggested model's application-level architecture, demonstrating its incompatible components and their interactions inside the system.

To strengthen the accuracy of tumor detection and as well as the segmentation, recent research have investigated the integration of multi-modal data, for instance integrating MRI with other imaging modalities like Positron Emission Tomography (PET) or Computed Tomography (CT) [25-27]. In order that to address the data scarcity issue in medical imaging, researchers have also looked at the usage of generative adversarial networks (GANs) for data augmentation [28-29] and also the semi-supervised learning techniques to use unlabeled data [30]. Surprisingly with the massive advancements in the discipline, there are still a number of unresolved issues and potential avenues for furthermore study. Between these it include the creation of reliable and understandable models that can explain their choices, improving human-machine interaction, and also promoting confidence within the medical community [31-33]. Moreover, the performance and as well as the clinical applicability of these computational methods may be improved by adding the domain expertise and expert advice into the model training procedure in the future [34-36].

Investigating federated learning approaches is another major direction. These approaches enable collaborative model training during maintaining data security and also the privacy, allowing knowledge exchange across various institutions without jeopardizing patient confidentiality [37-40]. Now in addition, in order to enable real-time analysis and decision support in clinical settings, it is imperative to build efficient algorithms for processing large-scale medical imaging data, possibly employing cloud-based and distributed computing technologies. [41-44].

In summary, the use of deep learning and computer vision algorithms in brain MRI analysis for tumor detection and segmentation has demonstrated a great deal of promise, providing precise and effective solutions to enhance clinical decision-making and enhance patient care. To further improve this subject, interdisciplinary cooperation between computer scientists, medical experts, and domain specialists are necessary to overcome the difficulties of data scarcity, interpretability of models, and scalability. While brain

tumor identification and segmentation in MRI data has showed significant promise with deep learning and computer vision perspective, there are still a number of influential issues that need to be addressed. These incorporate scalability challenges in proceedings extensive medical imaging data for real-time analysis, a absence of big and well-annotated datasets for model training, and the inability of deep learning models to be interpreted, that undermines the confidence of the medical community.

In order to overcome the constraint of limited data, this study explores course of action such as data augmentation, transfer learning, and also the regularization while applying picture classification and semantic segmentation strategies employing architectures such as CNNs, VGG19, and U-Net. The significance is in offering precise and practical ways to enhance clinical judgment and brain tumor patient treatment. However further interdisciplinary cooperation among computer scientists, physicians, and subject matter experts is required to address issues with interpretability, deal with data scarcity, and also it create scalable algorithms. The formation of interpretable models, the integration of domain proficiency, federated learning for collaboration that protects privacy, and the establishment of effective algorithms for processing massive amounts of medical imaging data at scale are achievable future paths.

Introduction: Provides background on the identification of brain tumors in MRI data and uses deep learning methods, such as picture segmentation and classification, to inspire. Draws attention to issues like interpretability of the model and insufficient data.

Mathematical Formulations: The cross-entropy loss and also the Adam optimizer used in the paper are give details of in the following sections.

Image Classification: Explains the augmentation and preparation of data. Describes two models (pre-trained VGG19, simple CNN) and shows the classification outcomes of each.

Image Segmentation: Explains the LabelBox mask generating method. Analyzes segmentation performance under various training conditions and presents the U-Net architecture.

Conclusion: Summarizes the effective use of segmentation and classification in tumor detection. Identifies new directions, such as semi-supervised learning and multi-modal data, while acknowledging existing difficulties.

Future Research Directions: Discusses about creating interpretable models, including expert knowledge, using federated learning to protect privacy, and creating scalable algorithms for handling massive amounts of medical image data.

2 | Mathematical Formulations

The implementation of deep learning methods for semantic segmentation and also image classification on brain MRI data is make up in the paper. These type of methods incorporate Convolutional Neural Networks (CNNs) and U-Net architectures. The usage of standard loss functions and optimization strategies for training these models is introduced in the research, despite the insufficiency of detailed mathematical formulations for the model designs. The paper's represented mathematical formulations for the Adam optimizer and the cross-entropy loss function are described in the next to the following subsections.

2.1 | Cross-Entropy Loss

For this, the paper in reference to using the cross-entropy loss function for one and the other the image classification and segmentation tasks. In the direction of a binary classification problem with N samples, the cross-entropy loss is specified by:

$$L = -\frac{1}{N}\sum_{i=1}^{N}[y_i\log(p_i) + (1 - y_i)\log(1 - p_i)]$$

Consequently y_i is the true label (0 or 1) and also p_i is the predicted probability of the following positive class for the ith sample.

2.2 |Adam Optimizer

For the Adam optimization algorithm, along-established method for training machine learning models, incorporate the advantages of one and the other the AdaGrad and RMSProp algorithms. It adaptively adjusts the learning rates of the parameters by computing individual adaptive learning rates for contradictory parameters from estimates of the first and also the second moments of the gradients. The parameter improvements in the Adam optimization algorithm are computed as in the following way:

$$m_t = \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t \qquad \text{(Exponential moving average of gradients)} \tag{1}$$

$$v_t = \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot g_t^2 \qquad \text{(Exponential moving average of squared gradients)} \tag{2}$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t} \qquad \text{(Bias correction form)} \tag{3}$$

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t} \qquad \text{(Bias correction for v)} \tag{4}$$

$$\theta_{t+1} = \theta_t - \frac{\alpha.\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon} \tag{5}$$

Where θ represents the model parameters, g_t is the gradient at time t, m_t and v_t are the exponential moving averages of the gradient and squared gradient respectively. β_1, β_2 are hyperparameters controlling the decay rates, α is the learning rate, and ϵ is a small constant for numerical stability.

The paper specifies using Adam with $\alpha = 0.001$, momentum $= 0.9$, and weight decay $= 0.0004$ for regularization.

3 |Image Classification

3.1 |Classification Data and Processing

For both our image segmentation and classification tasks, we use a public data set of MRI images found on Kaggle [2]. The data consists of 273 axial images of brain MRIs. Each of these images is labeled as one of two classes. Either the MRI has been positively identified as containing a brain tumor, or not. We will first discuss how we prepared the data for binary classification. Following this discussion, we will show how we generated masks for the same dataset to prepare for image segmentation.

While binary classification is a fairly simple task, only 273 images to work with can pose challenges. With such few data points, we will need to employ several techniques to achieve high test accuracy and avoid overfitting our training data. Table 1 shows the class distribution of the dataset, as well as how we split our data from each classes into training and testing sets. Despite the number of images being slightly skewed towards the positive label, we will still consider the baseline model to have **50%** accuracy. Thus, we aim to build a classifier that achieves test accuracy greater than this baseline.

Table 1. Training and testing are split by class.

Class	Training	Testing
Yes	124	31
No	77	41

We performed exploratory data analysis to detect any abnormalities in the dataset. This was done by ensuring no corrupted images or duplicate image IDs were present for an MRI. After these basic checks, we began image processing. The images were very messy. Some came in RGB format, while others were already converted to grayscale. To further complicate things, many images had different resolutions and would need

Tumor Detection in MRI Data using Deep Learning Techniques for Image Classification ...

6

to be resized. Finally, the images were not cropped, which resulted in random artifacts beyond the border of the brain being present in the MRI. The left panel of Figure 2 illustrates a few examples of uncropped images from each class.
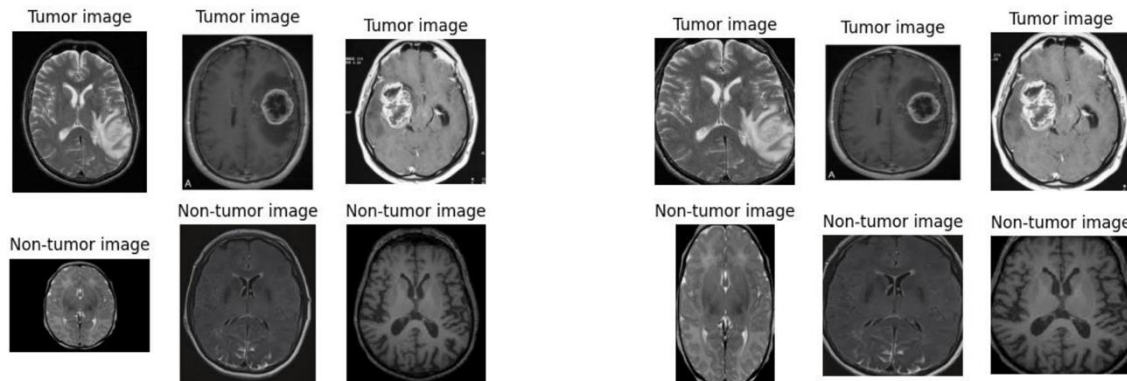

Figure 2. Left: Uncropped, unprocessed example images from both classes. Right: The same images, only now they have been cropped.

In order to fix these issues, we performed several steps of image processing. The first, and most involved step, was to crop each image to only include the extreme points of the height and width of the borders of the brain. This was done using various functions in the CV2 Python package [6]. The results of this cropping are shown in the right panel of Figure 2. We then checked each image to ensure that the cropping was successful.

Because of the small amount of training data, we used Pytorch transformations to perform data augmentation. The idea here is to slightly alter the images for each batch of training data loaded into the forward pass of our models. This has the effect of increasing the size of the dataset, without actually generating any new images. To do this, we resized each image to $(224,224)$, performed a horizontal flip and rotation with a 50 applied a Gaussian blur with a kernel size of 3, and then applied normalization. This was done on each crop of training data. For our test data, we only perform the resizing, tensor transformation, and normalization.

3.2 | Methods and Results for Classification

Now that we had our data processed and was ready to be augmented, we trained two classifiers. For our first classifier, we implemented a simple untrained CNN with two convolutional layers and two fully connected layers. Each convolutional layer uses a 3×3 kernel with a padding and stride of 1. After each of these layers, we applied a ReLU activation function on the output maps and then performed max pooling. We use a cross-entropy loss function and an Adam optimizer with a learning rate of 0.01, a momentum of 0.9, and a weight decay of 0.0004. The model architecture and parameters are further described in Tables 2 and 3.

Table 2. Model architecture.

Layer	Layer Type	Output Size
1	Conv2d(3,16,3,1,1)	224×224
2	MaxPool2d(2,2)	112×112
3	Conv2d(16,32,3,1,1)	112×112
4	MaxPool2d(2,2)	56×56
5	Ler(32*56*56,128)	128
6	Linear (128, 2)	2

The added weight decay introduces regularization to our model, which is crucial for a dataset as small as ours. Regularization helps prevent overfitting by discouraging overly complex models that may fit noise in the training data. With weight decay, we impose a penalty on the magnitude of the weights, effectively shrinking them towards zero during training. This encourages the model to focus on the most important features and reduces the chances of overfitting.

We trained the model on 50 epochs and tracked the training and validation accuracy after each epoch. This iterative method that allows us to monitor how proficiently the model is learning from the training data and generalizing to concealed data. Alongside comparing the training and also the validation accuracies, we can notice the signs of overfitting or underfitting. Overfitting takes place when the model executes well on the training data but poorly on the validation data, indicating that it has memorized the training examples as an alternative to learning the underlying patterns. Underfitting, on the other hand, recommends that the model is too simple to capture the patterns in the data, resulting in substandard performance on the training and also on validation sets.

Table 3. Adam parameters.

Parameter	Value
Learning Rate (α)	0.001
Momentum	0.9
Weight Decay	0.0004

After that, we plotted a confusion matrix for the final predictions on a test set. Now a confusion matrix is a beneficial tool for evaluating the performance of a classification model. It also provides a summary of the predictions made by the model compared to the actual labels in the test data. Alongside examining the confusion matrix, also we can identify which classes are being confused with each other and estimate the overall accuracy of the model. Visualizing the confusion matrix gives the advantage of gaining insights into where the model is making errors and can guide further improvements to the model architecture or training process. The results demonstrated in Figure 3 showcase the effectiveness of our approach and it also provides valuable comprehension for future iterations of the model.
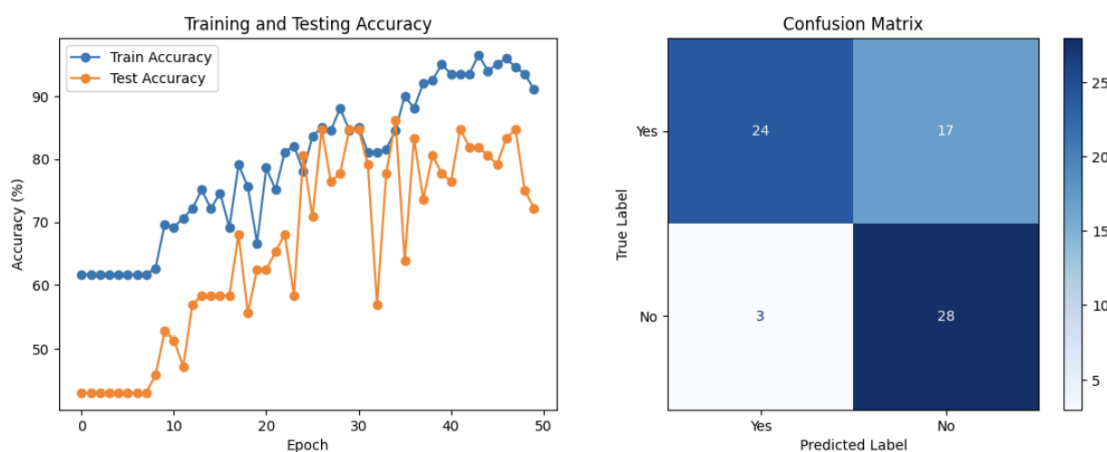

Figure 3. Plot for the training and test accuracy of the CNN we trained from scratch.

The results of this model were surprisingly good. It outperformed the baseline model, with both training and testing accuracy tending to increase with more epochs. However, as we can see in the plot of training accuracy against test accuracy, the model test accuracy is highly variable. Further, there is still a decent amount of misclassified images reported in the confusion matrix. Instead of trying our luck at training this model for a few more hours, we decided to move to a pre-trained model to see if transfer learning could help improve accuracy.

For our pre-trained model, we decided to go with VGG19 from the PyTorch pre-trained models library. We again used a cross-entropy loss function with Adam optimization. The same learning rate, momentum, and weight decay were used as in the previous model. To use the model, we froze all pre-trained layers and changed the head of the model to perform binary classification. We processed our images to be (224, and 224), which is the image size VGG19 accepts. We again ran the model with 50 epochs and plotted the results in Figure 4.
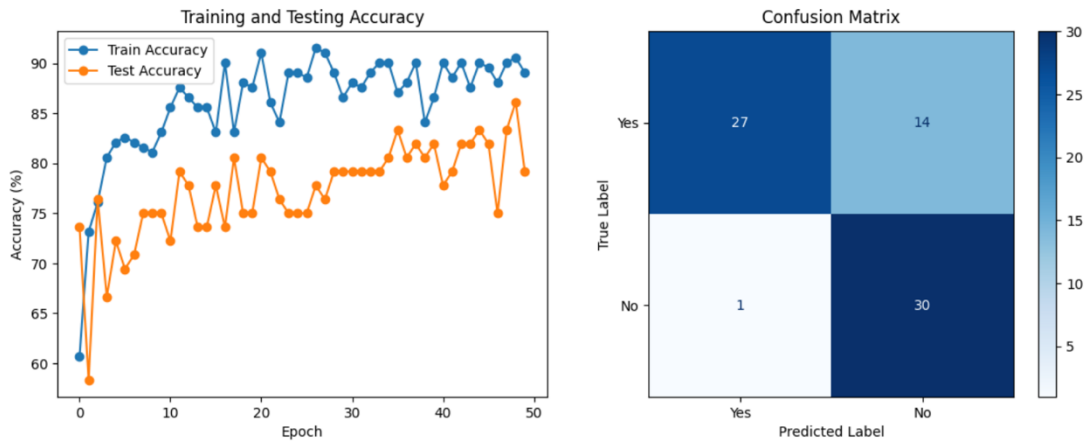
Figure 4. Plot for the train and test accuracy using the VGG19 pre-trained model.

Here we can see a huge improvement over the previous model. Specifically, we see less variance in the testing accuracy than that of the previous model. We also see an increase in train and test accuracy as the number of epochs increases. The confusion matrix improves from the previous model, but the model still seems to struggle with producing false negatives. The mean test accuracy across each epoch is given for both models is given in Table 4. We see that both models outperform the baseline, with VGG19 having the highest mean test accuracy.

Table 4. Mean test accuracy for each model against the baseline accuracy.

Model	Mean Test Accuracy
Baseline (12)	50.03%
ResNet (16)	59.61%
EfficientNet-B7 (17)	64.52%
Basic CNN (18)	66.22%
VGG19	77.19%

4 | Image Segmentation

4.1 | Data Preparation for Image Segmentation

As previously stated, we worked with the same Kaggle data set of brain MRI's for our segmentation task. To perform image segmentation, we first needed to generate masks for our original dataset, since it did not come with any. We manually created our segmentation masks for these images using LabelBox [1], a tool that allows for precise image masking. To do this we utilized several of LabelBox's tools for mask generation, identifying the boundaries of each image and then accessing the stored data from their cloud server to download and run on our local machines. The masks generated through LabelBox were then utilized as the ground truth data against which the model's predictions are compared.

It is important to note however that none of us are neuroscientists, but we highlighted the tumors in each image as best we could. We illustrate some of our difficulties in Figure 5.

Initially, each MRI scan and mask was resized to a resolution of 572×572. This is the image size that U-net takes as input. We then normalized both the brain MRI and the mask and converted each set of images to a grayscale format. This is also required by the U-net model. With this resizing and, normalization, and format transformation complete, the images and masks were now ready to be trained on.
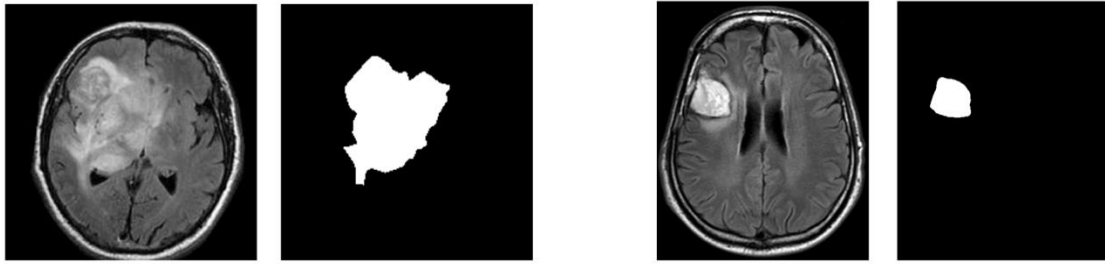
Figure 5. Left: An MRI with difficulty in creating the mask. Right: An MRI image that was easy to generate the mask for.

4.2 | Image Segmentation Methods and Results

One of the most commonly used models used for medical data image segmentation is U-net. The inputs for this model are images and corresponding masks (in our case brain MRI's and our masks) and the output is the predicted segment (in our case, the brain tumor). For this model, the output is a mapping of the binary image of pixels, the same as used for the masks shown above in the data processing section.

The U-Net model comprises three primary components: an encoder, a decoder, and the model core itself. The model is described as follows and in Table 5 below:

Encoder: The encoder consists of several blocks, each containing two convolutional layers with ReLU activation followed by a max pooling layer. Each convolutional layer uses 3×3 filters and operates with padding to preserve the dimensions of the input image. The max-pooling layers, with a $2x2$ filter and stride of 2, progressively reduce the spatial dimensions of the feature maps, effectively compressing the input data and enabling the model to capture essential features at multiple scales.

Bottleneck: The bottleneck works as the transition between the encoder and the decoder. It consists of two convolutional layers with 1024 filters each, again using 3×3 kernels and ReLU activation. This section aims to process the most abstracted features extracted by the encoder.

Decoder: The decoder incrementally reconstructs the spatial dimensions of the output. Each decoder block starts with an up-sampling layer followed by a convolutional layer that reduces the number of filters by half, compared to the preceding block. Each up-sampling step is followed by a concatenation operation with corresponding feature maps from the encoder, which reintroduces spatial and contextual information lost during downsampling. This is followed by two more convolutional layers with ReLU activation to refine the features.

Output: The final layer of the model is a convolutional layer with a $1x1$ kernel. It maps the feature maps to the desired number of output classes - here, just one, using a sigmoid activation function to produce a binary mask indicating the presence or absence of a tumor in each pixel.

After training the model on brain tumor images with generated masks, we achieved remarkable results in image segmentation predictions. Employing two distinct models based on U-net architecture facilitated insightful comparisons. The initial model utilized the entire dataset but constrained by computational limitations, we could only run it for 2 epochs. To address this, we devised a second model, leveraging only a fraction of the dataset, yet allowing for a more extensive training regime of 10 epochs. Figure 6 elegantly illustrates a representative outcome from each of these trained models, showcasing the nuances in performance and highlighting the trade-offs between dataset comprehensiveness and computational efficiency.

Table 5. U-Net Model table. We use an Adam optimizer with a learning rate of 0.001 and cross-entropy loss function.

Layer	Type of Layer	Output S
Input	Input Layer	(256,256,1)
Encoder 1	Conv2D + ReLU + Conv2D + ReLU + MaxPool	(128,128,64)
Encoder 2	Conv2D + ReLU + Conv2D + ReLU + MaxPool	(64,64,128)
Encoder 3	Conv2D + ReLU + Conv2D + ReLU + MaxPool	(32,32,256)
Encoder 4	Conv2D + ReLU + Conv2D + ReLU	(32,32,512)
Bottleneck	Conv2D + ReLU + Conv2D + ReLU	(32,32,1024)
Decoder 4	UpSampling2D + Conv2D + Concatenate + Conv2D + ReLU + Conv2D + ReLU	(64,64,512)
Decoder 3	UpSampling2D + Conv2D + Concatenate + Conv2D + ReLU + Conv2D + ReLU	(128,128,256)
Decoder 2	UpSampling2D + Conv2D + Concatenate + Conv2D + ReLU + Conv2D + ReLU	(256,256,128)
Decoder 1	UpSampling2D + Conv2D + Concatenate + Conv2D + ReLU + Conv2D + ReLU	(256,256,64)
Output	Conv2D (Sigmoid)	(256,256,1)
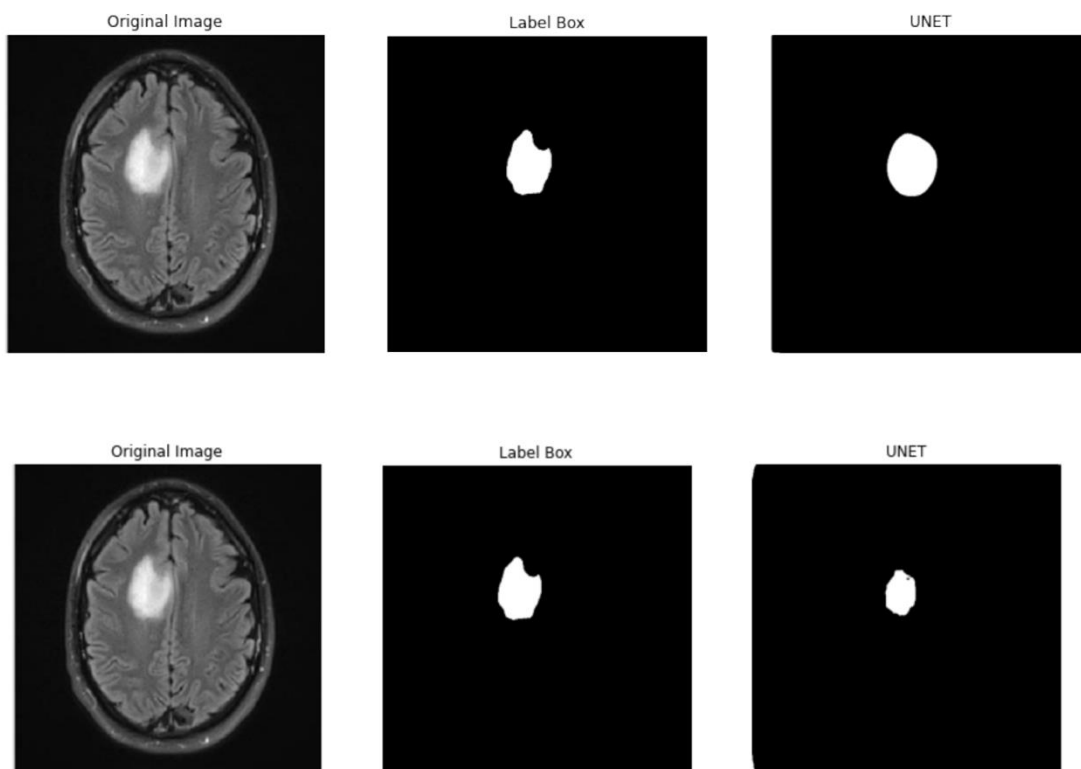
Figure 6. Row 1: Shows the input image, ground truth mask, and U-net prediction trained on an entire dataset for 2 epochs. Row 2: Now the similar configuration, but this time the model is run for 10 epochs and we train with less data.

The entanglement of these findings recommends a significant interplay between the quantity of training data and the number of epochs in achieving optimal model performance. This underscores the significance of striking a balance between the dataset size and also training duration. Regardless of its inherent simplicity, the U-net architecture demonstrates extraordinary adaptability in scenarios with limited training instances, showcasing its robustness in semantic segmentation assignments. We can see that these comprehensions not only enhance our understanding of deep learning methodologies but also provide practical guidance for optimizing model training strategies in resource-constrained environments. Exploration into the dynamics of model convergence and generalization could primarily contribute deeper insights into the mechanisms underlying the observed performance variations. Such kinds of investigations hold promise for refining existing methodologies and advancing the field of machine learning.

11

Abid and Shahid | Sustain. Mach. Intell. J. 9 (2024) 1-13

5 | Conclusion

Comprehensively, we are fairly advantageous with the results for both our binary classification and also two-dimensional semantic image segmentation tasks. For the binary classification task, we were successfully able to implement crucial data processing steps to prepare our images for each model. We overcame the problem of having a small dataset by utilizing data augmentation, regularization, and using a pre-trained model. Our first CNN had decent results, but variable test accuracy led to a poor mean testing accuracy score. Improvements were immediately seen after implementing VGG19, with less variability and higher mean test accuracy. In future work, we would like to use a held-out validation set to implement hyperparameter tuning of our learning rate and weight decay.

For the image segmentation task, we were able to generate high-quality image masks using the Labelbox software to use to train our model. From this, we then were able to run multiple iterations with several different parameters over the data set to generate masks and found that the model with 2 epochs over the entire data set outperformed the model with 10 epochs run over a third of the data set. In future work, we would ideally be able to have expert verification in the creation of the masks in addition to recording model accuracy through a measurement such as pixel accuracy or the Jaccard Index. Also working with a different data set with multiple slices of the same tumor we could use the image segmentation techniques to generate a three-dimensional segmentation.

6 | Future Research Directions

One important direction for future research is the development of interpretable and trustworthy models that can provide explanations for their decisions. Also, this would enable recommended human-machine collaboration and foster trust in the medical community regarding the use of this kind of computational methods. Over and above that, integrating domain knowledge and expert guidance into the model training process could additionally enhance the performance and clinical relevance of this kind of technique.

Additionally, one more promising path lies in exploring federated learning proceed toward, which authorize collaborative model training over a period of time preserving data privacy and security. This would authorize the sharing of knowledge across several institutions without compromising patient confidentiality, a critical concern in the medical domain. Also developing more efficient and scalable algorithms for processing large-scale medical imaging data is also an essential part of this research, potentially leveraging distributed computing and cloud-based solutions to qualify real-time analysis and decision support in clinical settings.

Assimilating multi-modal data, something like combining MRI with other imaging modalities like Positron Emission Tomography (PET) or Computed Tomography (CT), could further enhance the accuracy of tumor detection and also segmentation. In addition to the above, investigating generative adversarial networks (GANs) for data augmentation and semi-supervised learning techniques and methods to leverage unlabeled data could help get the best out of the challenge of data scarcity in medical imaging.

Overall, this paper highlights that driving further advancements and innovations in this field will require interdisciplinary collaborations among computer scientists, medical experts, and domain specialists. These kinds of collaborations could bridge the gap between cutting-edge computational techniques and the practical needs of healthcare professionals, eventually leading to more productive and clinically relevant solutions.

Acknowledgments

Author Contributions

All authors contributed equally to this work.

Funding

This research was conducted without external funding support.

Data Availability

The datasets generated during and/or analyzed during the current study are not publicly available due to the privacy-preserving nature of the data but are available from the corresponding author upon reasonable request.

Conflicts of Interest

The author declares that there is no conflict of interest in the research.

Ethical Approval

This article does not contain any studies with human participants or animals performed by any of the authors.

References

[1] Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542(7639), 115-118.

[2] Akkus, Z., Galimzianova, A., Hoogi, A., Rubin, D. L., & Erickson, B. J. (2017). Deep learning for brain MRI segmentation: state of the art and future directions. Journal of digital imaging, 30(4), 449-459.

[3] Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. Medical image analysis, 42, 60-88.

[4] Shen, D., Wu, G., & Suk, H. I. (2017). Deep learning in medical image analysis. Annual review of biomedical engineering, 19, 221-248.

[5] Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., & Lanczi, L. (2015). The multimodal brain tumor image segmentation benchmark (BRATS). IEEE transactions on medical imaging, 34(10), 1993-2024.

[6] M. Saqlain et al., A Dynamic Investigation to Analyzing Divorce effects within Wolbachia Models, Nonlinear Convex Anal. & Optim., Vol. 2 No. 2, 55-73.

[7] Jia, H., Cheng, F., Guan, Y., Zhang, D., & Yiu, S. M. (2021). The challenges of making machine learning available in clinical practice. International Journal of Medical Informatics, 146, 104341 .

[8] Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. Journal of Big Data, 6(1), 60.

[9] Saqlain, M. (2023). Evaluating the readability of English instructional materials in Pakistani Universities: A deep learning and statistical approach. Education Science and Management, 1(2): 101-110.

[10] Raghu, M., Zhang, C., Kleinberg, J., & Bengio, S. (2019). Transfusion: Understanding transfer learning for medical imaging. Advances in neural information processing systems, 32.

[11] He, K., Girshick, R., & Dollár, P. (2019). Rethinking imagenet pre-training. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 4918-4927).

[12] Lohia, R., Nagi, J., Kapoor, A., Kannan, R., & Viswanath, P. (2019). Regularizing by dropout and weight decay helps in overfitting but what helps in generalization?. arXiv preprint arXiv:1905.11278.

[13] Meharunnisa, Saqlain, M., Abid, M., Awais, M., and Stević, Ž. Analysis of software effort estimation by machine learning techniques. Ingénierie des Systèmes d'Information, 28(2023),1445-1457.

[14] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham.

[15] Saqlain, M. (2023). Sustainable hydrogen production: A decision-making approach using VIKOR and intuitionistic hypersoft sets. Journal of Intelligent Management Decision, 2(3): 130-138.

[16] Hu, Y., Wang, L., Cheng, D., Xu, Q., Lu, Y., Liu, Y., & Yang, T. (2022). Brain tumor segmentation using multi-cascaded convolutional neural networks and conditional random field. IEEE Access, 10, 3459-3472.

[17] Crombe, A., Duron, V. B., Cherifi, S., Harbec, L. L., Mercier, L., Langs, G., & Biailly, A. (2021). A review of datasets for machine learning-based medical image segmentation. IEEE Transactions on Biomedical Engineering, 69(3), 847-865.

13

Abid and Shahid | Sustain. Mach. Intell. J. 9 (2024) 1-13

[18] Ortiz-Ramón, R., Lazcano, R., Bian, W., Madabhushi, A., Rosen, B., & Ge, Y. (2022). Multimodal data fusion for brain tumor segmentation. Journal of Digital Imaging, 1-13.

[19] Abid, M. & Saqlain, M. (2023). Utilizing Edge Cloud Computing and Deep Learning for Enhanced Risk Assessment in China's International Trade and Investment. Int J. Knowl. Innov Stud., 1(1), 1-9.

[20] Abid, M., & Saqlain, M. (2023). Decision-Making for the Bakery Product Transportation using Linear Programming. Spectrum of Engineering and Management Sciences, 1(1),1-12.

[21] Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence, 1(5), 206-215.

[22] Haq, H. B. U., Akram, W., Irshad, M. N., Kosar, A., & Abid, M. (2024). Enhanced Real-Time Facial Expression Recognition Using Deep Learning. Acadlore Trans. Mach. Learn., 3(1), 24-35.

[23] Sheller, M. J., Reina, G. A., Edwards, B., Martin, J., & Bakas, S. (2020). Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. Scientific reports, 10(1), 1-12.

[24] Hamid, M. T., & Abid, M. (2024). Decision Support System for Mobile Phone Selection Utilizing Fuzzy Hypersoft Sets and Machine Learning. J. Intell. Manag. Decis., 3(2),104-115.

[25] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. The journal of machine learning research, 15(1), 1929-1958.

[26] Ren, J., Liu, P. J., Fertig, E., Snavely, J., Poplin, R., Depristo, M., & Norman, T. (2020). A comprehensive survey of neural architecture search: Challenges and solutions. ACM Computing Surveys (CSUR), 54(4), 1-42.

[27] Isensee, F., Jaeger, P. F., Kohl, S. A., Petersen, J., & Maier-Hein, K. H. (2021). No New-Net. In Machine Learning for Medical Image Reconstruction (pp. 234-244). Springer, Cham.

[28] Yi, X., Walia, E., & Babyn, P. (2019). Generative adversarial network in medical imaging: A review. Medical image analysis, 58, 101552.

[29] Kelly, C. J., Karthikesalingam, A., Suleyman, M., Corrado, G., & King, D. (2019). Key challenges for delivering clinical impact with artificial intelligence. BMC medicine, 17(1),1-13.

[30] Chang, J., Chou, Y. C., Wang, R., Li, W., Kim, M., Deng, J., & Liu, Y. (2022). Semisupervised medical image segmentation via cross-teaching between heterogeneous 3D and 2D encoders. arXiv preprint arXiv:2204.03830.

[31] Tajbakhsh, N., Jeyaseelan, L., Li, Q., Chiang, J. N., Wu, Z., & Ding, X. (2020). Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation. Medical Image Analysis, 63, 101693.

[32] Tahoor, Abid, M., Mushtaq, A. & Bibi M. (2024). Exploring the Strong Metric Dimension of Hollow Coronoid Struc- tures: Applications and Implications. Complexity Analysis and Applications,1(1), 1-13.

[33] Wang, S., Zhou, Y., & Mehio-Ranahane, M. (2018). Scalable cloud computing for large-scale medical imaging applications. Journal of Cloud Computing, 7(1), 1-12.

[34] Ullah, N., Khan, J.A., Khan, M.S., Khan, W., Hassan, I., Obayya, M.I., Negm, N., & Salama, A.S. (2022). An Effective Approach to Detect and Identify Brain Tumors Using Transfer Learning. Applied Sciences, 12(11), 5645.

[35] Hencya, F.R., Mandala, S., Tang, T.B., Soperi, M., & Zahid, M. (2023). A Transfer Learning-Based Model for Brain Tumor Detection in MRI Images. Jurnal Nasional Teknik Elektro, 12(2).

[36] Abdalla, P.A., Mohammed, B.A., & Saeed, A.M. (2023). The impact of image augmentation techniques of MRI patients in deep transfer learning networks for brain tumor detection. Journal of Electrical Systems and Information Technology, 10, 51.

[37] Asif, S., Zhao, M., Tang, F., & Zhu, Y. (2023). An enhanced deep learning method for multi-class brain tumor classification using deep transfer learning. Multimedia Tools and Applications, 82, 31709-31736.

[38] Savaş, S., & Damar, Ç. (2023). Transfer-learning-based classification of pathological brain magnetic resonance images. ETRI Journal.

[39] Alhatemi, R.A.J., & Savaş, S. (2022). Transfer learning-based classification comparison of stroke. Computer Science, 2022, 192-201.

[40] Ghosal, P., Nandanwar, L., Kanchan, S., Bhadra, A.K., Chakraborty, J., & Nandi, D. (2019). Brain Tumor Classification Using ResNet-101 Based Squeeze and Excitation Deep Neural Network. In 2019 Second International Conference on Advanced Computational and Communication Paradigms (ICACCP), 25-28 February, Gangtok, India.

[41] Terzi, D.S., & Azginoglu, N. (2023). In-Domain Transfer Learning Strategy for Tumor Detection on Brain MRI. Diagnostics, 13(12), 2110.

[42] Islam, R., Akhi, A.B., & Akter, F. (2023). A fine tune robust transfer learning based approach for brain tumor detection using VGG-16. Bulletin of Electrical Engineering and Informatics, 12(6), 3861-3868.

[43] Hurtík, P., Molek, V., Hula, J., Vajgl, M., Vlasánek, P., & Nejezchleba, T. (2022). Poly-YOLO: higher speed, more precise detection, and instance segmentation for YOLOv3. Neural Computing and Applications, 34, 8275-8290.

[44] Esteva, A., et al. (2021). Deep learning-enabled medical computer vision. NPJ digital medicine, 4(1),5.