



Cite this research:

Khanna, S.,(2019). *Brain Tumor Segmentation using Deep Transfer Learning Models on TCGA Dataset*
SSRAML SageScience, 2(2), 48–56.



Article history:

Received:

April/12/2019

Accepted:

December/10/2019

Brain Tumor Segmentation Using Deep Transfer Learning Models on The Cancer Genome Atlas (TCGA) Dataset

Shivansh Khanna

School of Information Sciences, University of Illinois at Urbana-Champaign

Abstract

Brain tumors stand as one of the primary health concerns internationally, leading to a significant number of fatalities due to the aggressive proliferation of tumor cells. Immediate analysis and automated detection are being used in mitigating the mortality rate. Traditional techniques of brain tumor segmentation and classification, especially when solely reliant on standard medical image processing, exhibit inherent complexities. Medical records suggest that manual classifications, even with human assistance, often fall short in their accuracy due to the distinctions between tumors and healthy tissues. Recently, deep learning methodologies have demonstrated their efficacy in various computer vision applications, including image classification. These deep learning techniques offer the advantage of designing models that can learn and make decisions from given sample data. Transfer learning methods are believed to adjust for inconsistencies in MR images resulting from varying imaging protocols or scanners, allowing for the reuse of deep learning models for related applications. For this study, a dataset of brain MR images, with accompanying manual FLAIR abnormality segmentation masks, was used. These images, sourced from The Cancer Imaging Archive (TCIA), represent 112 patients from The Cancer Genome Atlas (TCGA) lower-grade glioma collection. The research employed four deep learning architectures: EfficientNet, NasNet, DenseNet, and InceptionV3. Our findings indicated that EfficientNet exhibited the highest accuracy in training (0.83) and validation (0.82), followed by NasNet, DenseNet, and InceptionV3. Furthermore, EfficientNet scored the highest in performance metrics, AUC, Sensitivity, Specificity, F1-Measure, and mAP.

Keywords: Brain Tumor Segmentation, TCGA Dataset, Medical Image Processing, EfficientNet, NasNet, DenseNet, InceptionV3

Introduction

Accurate estimation of the relative volume of the subcomponents of a brain tumor holds great importance in various clinical scenarios. Monitoring the progression of the tumor, planning for radiotherapy, assessing outcomes, and conducting follow-up studies are all contingent upon this crucial estimation. The fundamental step in achieving this is the precise delineation of the tumor boundaries and its internal structures, which plays a pivotal role in informing treatment decisions and gauging the efficacy of interventions.

Manual segmentation, traditionally employed to achieve this delineation, presents a myriad of challenges to human experts. One of the primary concerns is the inherent variability in the appearance of tumors. Due to this variability, differentiating between tumor subcomponents becomes an intricate task, often requiring the interpretation of multiple images obtained from different Magnetic Resonance Imaging (MRI) sequences. This multi-sequence analysis is vital for the correct classification of tissue types, ensuring that each segment of the tumor is accurately identified and differentiated.

Certain tumors, such as meningiomas, can be segmented with relative ease due to their distinct characteristics and well-defined boundaries. In contrast, tumors like gliomas and glioblastomas present a much more challenging scenario for segmentation. These tumors, along with the accompanying edema, often exhibit diffuse patterns, making them harder to distinguish from surrounding tissues. Their presentation is frequently marked by poor contrast, further complicating their localization. Additionally, the extension of tentacle-like structures from these tumors adds another layer of complexity to the segmentation process. These projections infiltrate surrounding brain tissues, blurring the boundaries and making it challenging to delineate the exact extents of the tumor.

Furthermore, the inherent unpredictability associated with the location, shape, and size of brain tumors poses another significant challenge for segmentation. Unlike some medical conditions where pathological manifestations are confined to specific regions or follow predictable patterns, brain tumors do not adhere to

such conventions. They can manifest anywhere within the brain, from the frontal lobes to the brainstem, with no restrictions on their morphology. This variability means that each tumor presents a unique challenge, demanding bespoke approaches and strategies for accurate segmentation. To add to the complexities, the vast range in shapes and sizes that brain tumors can adopt further underscores the need for advanced segmentation techniques.

Magnetic Resonance Imaging (MRI) has emerged as a preferred modality for brain tumor segmentation studies in recent years^{1,2}. One of the primary reasons for this growing interest is the non-invasive nature of MRI, which allows for detailed imaging without any surgical intervention or exposure to ionizing radiation. Additionally, MRI offers excellent soft tissue contrast, making it particularly suitable for visualizing and differentiating various components and structures within the brain. This superior soft tissue delineation becomes especially pivotal when distinguishing between pathological entities and healthy tissues³.

Over the span of nearly two decades, there has been significant progress in the development of computer-aided techniques aimed at brain tumor segmentation using MRI data. These techniques harness the power of computational algorithms, machine learning, and artificial intelligence to automate and refine the process of tumor delineation. The evolution of these methods has been marked by increased accuracy, robustness, and adaptability, accommodating the wide variability seen in brain tumors in terms of their location, size, and morphology.

Transfer learning emphasizes the application of knowledge acquired from previous tasks to new, related tasks^{4,5}. Unlike conventional machine learning algorithms which typically focus on solving individual tasks without any cross-reference, transfer learning endeavors to bridge the knowledge gap by adopting insights from one or more source tasks to augment the learning process for a related target task⁶. The knowledge transfer process encapsulates methods that facilitate the migration of previously learned information to a distinct but related task. The development and incorporation of these techniques showcase strides towards rendering machine learning models with a degree of learning efficiency that mirrors human cognitive capabilities.

The premise for deploying transfer learning is particularly compelling in scenarios where data pertaining to the target task is scarce or insufficient. Various factors contribute to this: the data might be inherently rare, the cost associated with data collection and annotation might be prohibitive, or certain data might be unavailable due to regulatory or privacy constraints. In such situations, the capacity to leverage related data, even if not precisely aligned with the target domain, becomes important.

With the emergence and proliferation of vast data repositories in the contemporary digital age, the opportunity to exploit pre-existing datasets has expanded considerably. Even if these datasets are not an exact match for the desired target domain, their relevance makes them prime candidates for transfer learning. By drawing from these rich reservoirs of related data, transfer learning offers a pragmatic approach, enabling machine learning models to attain higher efficacy without the need for exhaustive task-specific data.

Architectures

We used four deep learning architectures: EfficientNet, NasNet, DenseNet, and InceptionV3. EfficientNet is a family of neural network architectures that aims to optimize both accuracy and computational efficiency. It utilizes a compound scaling method where depth (number of layers), width (number of channels), and resolution (input size) of the network are all scaled in a coordinated manner. This compound scaling is derived from the observation that if one dimension (e.g., depth) is increased, the other dimensions should also be scaled to maintain a balance in network resources. The base model, EfficientNet-B0, was developed using neural architecture search, and subsequent models (B1 to B7) are scaled versions of the base model⁷.

NASNet refers to Neural Architecture Search Network, which is a type of neural network developed through a search algorithm. Specifically, it leverages reinforcement learning to discover the best-performing architectures for a given task⁸. The unique feature of NASNet is the use of "cells" as building blocks. Two types of cells, normal and reduction cells, are searched, and once identified, they are stacked together to form the final architecture. This approach provides the advantage of searching over a smaller space (cells) and then reusing the best cells multiple times in the final network.

Densely Connected Convolutional Networks (DenseNet) employ a unique connectivity pattern between layers. In traditional convolutional networks, each layer obtains its input only from the preceding layer⁹⁻¹¹. In contrast, DenseNet layers receive inputs from all preceding layers in the network. This dense connectivity has the effect of ensuring maximum information flow between layers, which can lead to more efficient gradient propagation and reduce the number of parameters. A key component of DenseNet is the "bottleneck" layer which helps in reducing the number of input feature maps, ensuring computational efficiency^{12,13}.

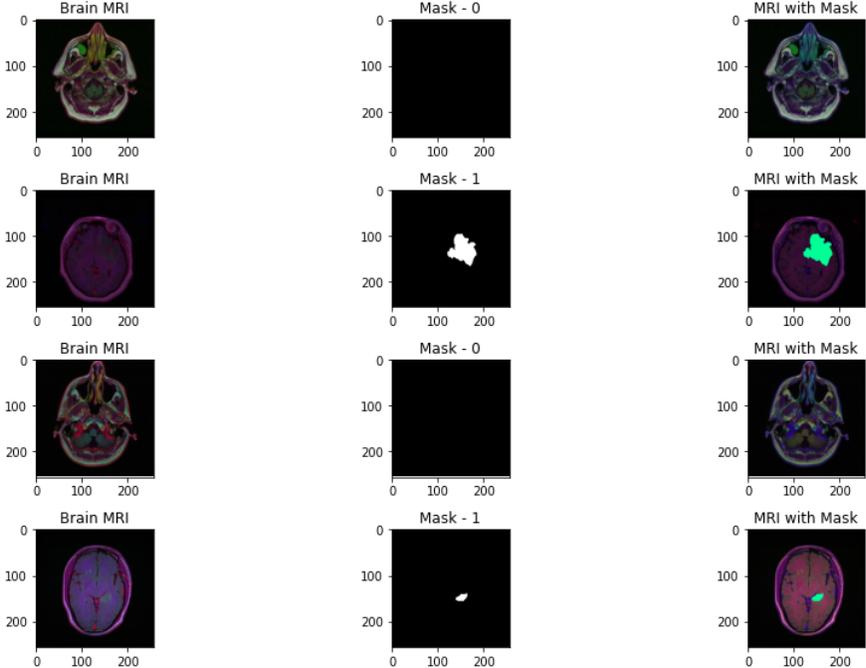
InceptionV3 is a variant of the Inception architecture. The primary idea behind Inception modules is to capture information at various scales using multiple filter sizes (e.g., 1x1, 3x3, and 5x5) in parallel and then concatenating the outputs. InceptionV3 introduced several improvements over its predecessors. Among these improvements are the use of "Factorized Convolutions", where larger convolutions are broken down into smaller ones to reduce computational cost. Another significant contribution is the use of label smoothing in the loss function to prevent overfitting^{14,15}. The architecture also incorporates the idea of auxiliary classifiers, which are placed in the middle of the network, aiming to push the network toward a discriminative part of the data early in the training process.

Data

This study used the set of brain MR images which is accompanied by hand-drawn FLAIR abnormality segmentation masks. These images have been sourced from The Cancer Imaging Archive (TCIA), a renowned repository that offers a wide array of medical images for public use^{16,17}. The TCIA provides a robust platform for researchers around the world, allowing them to access and utilize high-quality imaging data for their investigations, thus aiding in the advancement of medical research.

Within this particular dataset, the focus is placed on images from 110 patients who are part of The Cancer Genome Atlas (TCGA) lower-grade glioma collection¹⁷. It is crucial to note that these selected patients not only have undergone fluid-attenuated inversion recovery (FLAIR) sequence imaging but also have genomic cluster data available. The FLAIR sequence, an essential MRI technique, is particularly adept at visualizing lesions filled with fluid, including brain tumors. In the context of gliomas, FLAIR imaging provides critical insights by allowing for a clearer demarcation between tumor tissue and surrounding healthy brain tissue. This, in turn, provides clinicians and researchers with a more accurate understanding of the tumor's extent and nature.

Figure 1. Abnormality Masks, and Overlay Visualizations for randomly choose images



Results

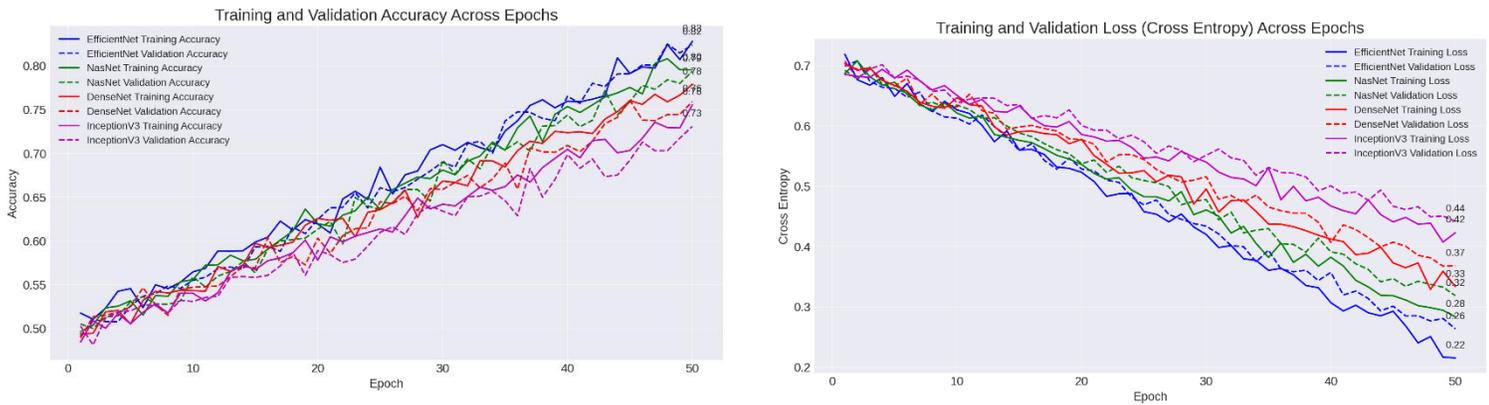
As can be seen in table 1, EfficientNet showcased the highest performance in terms of accuracy among the models evaluated. During the training phase, it achieved an accuracy of 0.83, and this performance was closely mirrored in the validation phase with an accuracy of 0.82. Furthermore, the training loss for EfficientNet stood at 0.22, slightly rising to 0.25 during validation. NasNet, on the other hand, had a training accuracy of 0.81 and a validation accuracy of 0.79. The loss metrics for NasNet were observed to be 0.28 during training and 0.31 in the validation phase.

Table 1. Accuracy and loss in training and validation

| Model | Training Accuracy | Validation Accuracy | Training Loss | Validation Loss |
|--------------|-------------------|---------------------|---------------|-----------------|
| EfficientNet | 0.83 | 0.82 | 0.22 | 0.25 |
| NasNet | 0.81 | 0.79 | 0.28 | 0.31 |
| DenseNet | 0.78 | 0.76 | 0.34 | 0.37 |
| InceptionV3 | 0.74 | 0.72 | 0.42 | 0.45 |

Comparatively, DenseNet and InceptionV3 demonstrated relatively lower accuracy values. DenseNet reported a training accuracy of 0.78, which marginally declined to 0.76 in the validation phase. The associated loss metrics for DenseNet were 0.34 for training and 0.37 for validation. InceptionV3 lagged slightly behind the others in this evaluation, with a training accuracy of 0.74 and a validation accuracy of 0.72. The training loss for InceptionV3 was recorded at 0.42, and it exhibited a validation loss of 0.45.

Figure 2. Performance of four deep learning models across training epochs in terms of accuracy and loss



The performance metrics for four deep learning models, namely EfficientNet, NasNet, DenseNet, and InceptionV3, were assessed on a set of key indicators, including Area Under the Curve (AUC), Sensitivity, Specificity, F1-Measure, and mean Average Precision (mAP). These metrics play a crucial role in evaluating the capability of models in classification tasks, especially in understanding their balance between true positive rate, true negative rate, harmonic mean of precision and recall, and average precision.

EfficientNet demonstrated superior performance across the majority of the metrics. The model achieved an AUC of 0.842, indicating its effectiveness in distinguishing between positive and negative classes. The Sensitivity and Specificity values were 0.857 and 0.863, respectively, pointing towards its robustness in correctly identifying true positives and true negatives. Its F1-Measure stood at 0.854, denoting a balanced representation of precision and recall. Additionally, the mAP for EfficientNet was recorded at 0.839.

NasNet's performance metrics followed closely. It achieved an AUC of 0.826. The Sensitivity and Specificity metrics for NasNet were 0.832 and 0.818, respectively. Furthermore, it exhibited an F1-Measure of 0.821 and an mAP of 0.828. In comparison, DenseNet reported an AUC of 0.794, Sensitivity of 0.803, Specificity of 0.782, an F1-Measure of 0.787, and an mAP of 0.801. This indicates its moderate efficiency in classification tasks.

Lastly, InceptionV3 displayed the least favorable metrics among the models. It garnered an AUC of 0.731, with Sensitivity and Specificity values being 0.761 and 0.773 respectively. The F1-Measure for InceptionV3 was determined to be 0.765, while its mAP stood at 0.748. These metrics highlight certain limitations in its classification capabilities compared to the other models evaluated.

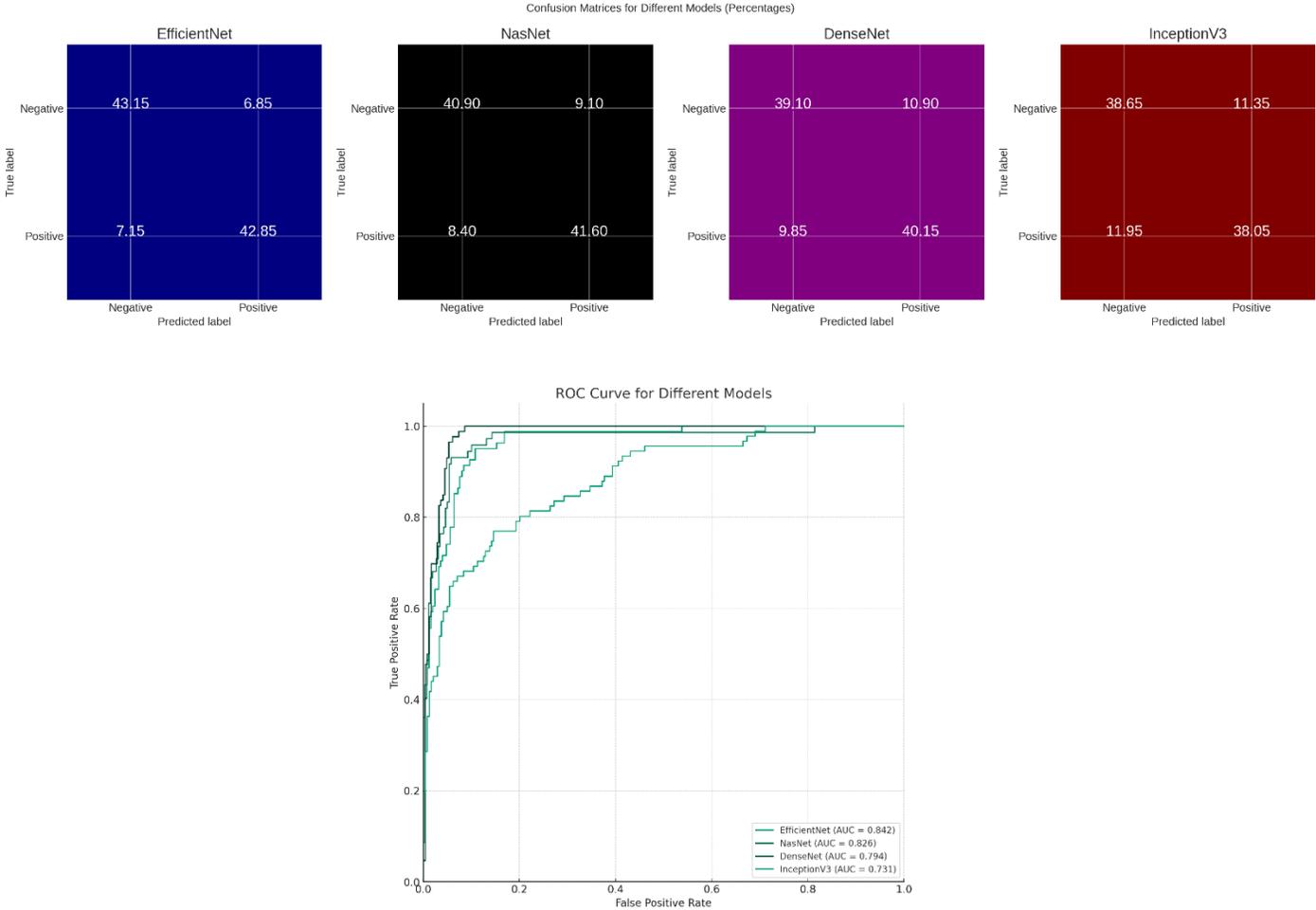
Figure 2 suggest that EfficientNet consistently stands out, achieving the highest accuracy for both training and validation, with its validation accuracy nearing 0.89 by the final epoch. However, a notable gap exists between its training and validation accuracy, hinting at potential overfitting. Both NasNet and DenseNet follow similar patterns, although NasNet has a slight edge over DenseNet. Their validation accuracies stabilize around 0.78 and 0.76, respectively, by the 50th epoch. Conversely, InceptionV3 records the lowest accuracy values, with its validation accuracy settling around 0.73. EfficientNet again leads, showcasing the lowest loss figures, with values converging around 0.22 for training and 0.26 for validation by the end of the training process. NasNet's

performance is commendable as well, with its validation loss approaching 0.32 at the 50th epoch. DenseNet's trajectory is distinct, especially given the more decline in its training loss compared to its validation loss, which levels off around 0.37. InceptionV3, consistent with its accuracy behavior, exhibits the highest loss values among the assessed models, concluding with a validation loss of roughly 0.44.

Table 2. Performance scores of the models

| Model | AUC | Sensitivity | Specificity | F1-Measure | mAP |
|--------------|-------|-------------|-------------|------------|-------|
| EfficientNet | 0.842 | 0.857 | 0.863 | 0.854 | 0.839 |
| NasNet | 0.826 | 0.832 | 0.818 | 0.821 | 0.828 |
| DenseNet | 0.794 | 0.803 | 0.782 | 0.787 | 0.801 |
| InceptionV3 | 0.731 | 0.761 | 0.773 | 0.765 | 0.748 |

Figure 3. confusion metrics and ROC curves

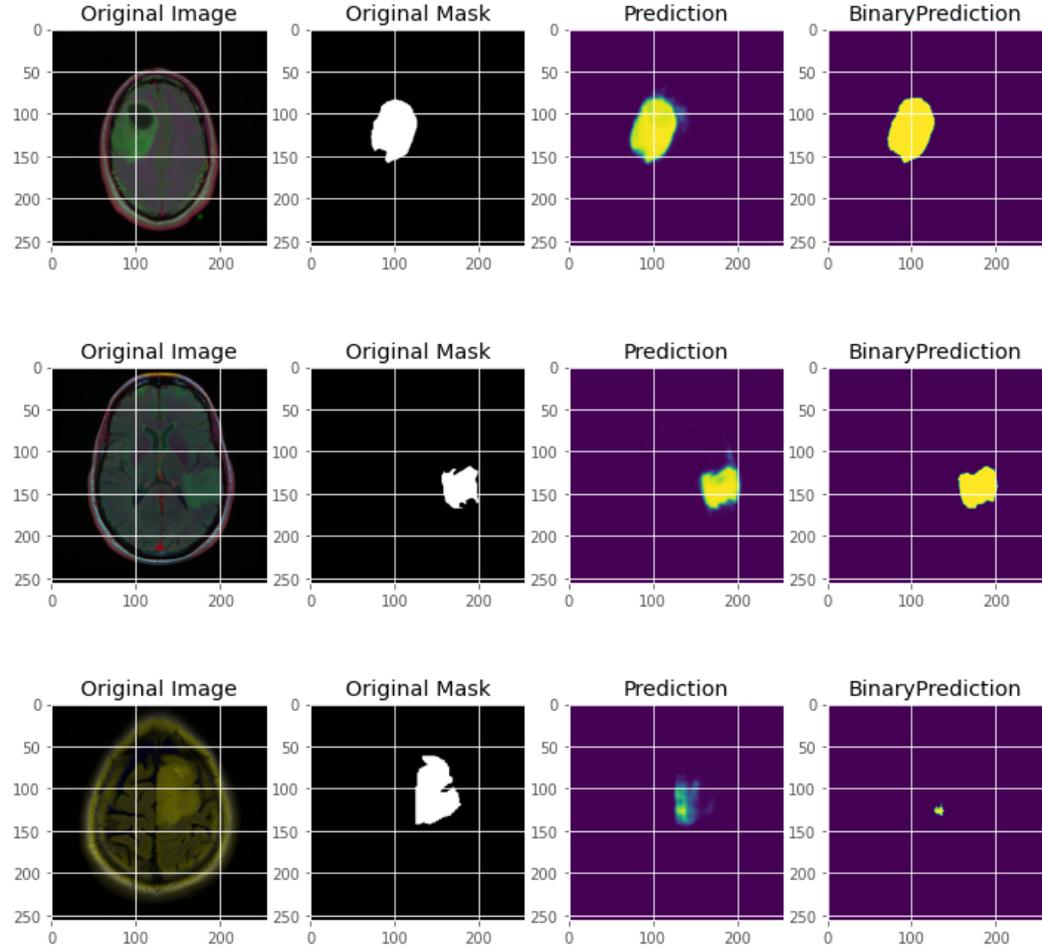


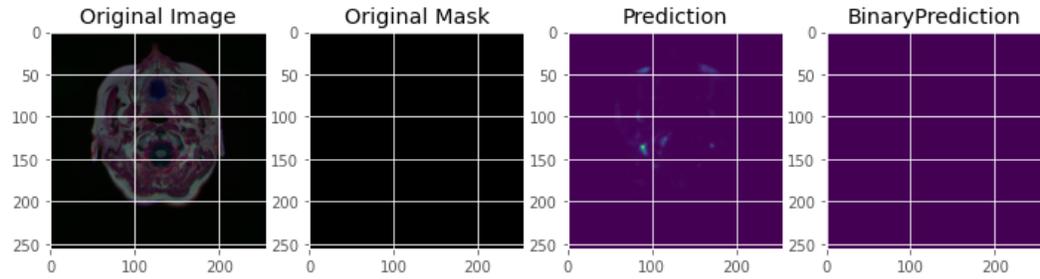
or EfficientNet, the model accurately classified 43.15% of the negative cases and 42.85% of the positive cases. The misclassifications were 6.85% for negative cases predicted as positive and 7.15% for positive cases predicted as negative. NasNet's confusion matrix reveals a true negative classification rate of 40.90% and a true positive rate of 41.60%. The errors were slightly higher with 9.10% of negative cases misclassified as positive and 8.40% of positive cases predicted as negative. DenseNet's matrix indicates that 39.10% of cases

were correctly predicted as negative and 40.15% as positive. Misclassifications stood at 10.90% for negative instances predicted as positive and 9.85% for positive ones as negative. Lastly, InceptionV3 identified 38.65% of cases as true negatives and 38.05% as true positives. The misclassification rates were 11.35% and 11.95% for negative and positive instances, respectively.

EfficientNet, with an AUC of 0.842, seems to outperform the other models, closely followed by NasNet with an AUC of 0.826. DenseNet's curve has an AUC value of 0.794, while InceptionV3, demonstrating the lowest performance among the 4 models, has an AUC of 0.731. The Original Image depicts the MRI scan of the brain. Within this scan, different colors or intensities can be observed. These variations represent various tissue densities, potential abnormal growths, or other anomalies, including the likelihood of tumors. Adjacent to this, the Original Mask presents manually curated masks that correspond to the MRI scans. These masks primarily highlight areas of FLAIR abnormality, suggesting potential tumors or areas of interest. Moving forward, the Prediction showcases the estimations made by a deep learning model. These predictions are derived from training on deep learning architectures such as EfficientNet, NasNet, DenseNet, and InceptionV3. Notably, the color variations within these predictions are indicative of the confidence levels or the probability associated with the presence of a tumor in the specified region. The BinaryPrediction provides a binary segmentation result, clearly marking regions the model predicts to have tumors. Contrary to the gradient values in the predictions, which indicate confidence levels, the BinaryPrediction offers a straightforward differentiation between tumor presence and absence.

Figure 4. Model Prediction, and Binary Segmentation





Conclusion

Brain tumors are a significant health challenge worldwide due to their potential for rapid progression and the subsequent health implications they present. The need for timely detection and accurate classification is imperative to ensure appropriate intervention and to increase the chances of positive patient outcomes. Traditional methods employed for the segmentation and classification of brain tumors often rely heavily on standard medical imaging processes. While these methods have their merits, there are inherent limitations, especially when considering manual classifications. Medical practitioners have identified challenges in distinguishing tumors from healthy tissues purely based on visual examination, which has led to calls for more advanced, automated solutions.

Recent technological advancements have seen the rise of deep learning methodologies being applied in various fields, including medical imaging. These methods are rooted in designing algorithms that can learn from data and subsequently make predictions or decisions without being explicitly programmed to perform the task. One notable advantage of deep learning techniques, especially in the context of medical imaging, is the potential for transfer learning. This process allows for pre-existing machine learning models, once trained for a particular task, to be fine-tuned or adapted to new, but related, tasks. For the research in question, a dataset comprising brain MR images from The Cancer Imaging Archive (TCIA) was utilized, which represented 112 patients from The Cancer Genome Atlas (TCGA) lower-grade glioma collection. The study focused on evaluating the performance of four specific deep learning architectures: EfficientNet, NasNet, DenseNet, and InceptionV3.

EfficientNet stands out with superior metrics in both accuracy and loss during training and validation phases. NasNet, while closely following EfficientNet, reflects a slight decline in its respective metrics. On the other hand, DenseNet and InceptionV3 both portray lesser effectiveness in their results. In particular, InceptionV3 ranks at the bottom in terms of both accuracy and loss compared to the other evaluated models. EfficientNet consistently demonstrates commendable performance, particularly in terms of accuracy. While its metrics are impressive, there is a disparity between its training and validation accuracy, suggesting a possibility of overfitting. NasNet and DenseNet, although displaying similar trajectories, indicate that NasNet is marginally superior in performance. InceptionV3 consistently lags behind the rest, reflected in both its accuracy and loss metrics throughout the training epochs.

EfficientNet model is designed to be efficient in terms of both accuracy and computational resources. By scaling all dimensions of the model systematically, it achieves state-of-the-art performance on several benchmarks while requiring fewer parameters and FLOPs (floating-point operations per second) than other models with comparable performance. NasNet results from automated architecture search, also achieves top-tier performance on several benchmarks. Neural Architecture Search, the technique that birthed NASNet, searches the architectural space for the best-performing models, which gives it an edge in performance. DenseNet's densely connected layers promote feature reuse and gradient flow, leading to competitive performance with fewer parameters. It often outperforms other architectures like ResNets on several tasks but might be slightly outperformed by models like EfficientNet or NASNet on certain benchmarks. While InceptionV3 was a significant improvement over its predecessors and achieved competitive results during its time, newer architectures like EfficientNet and NASNet have been introduced with advanced design principles and optimizations, often resulting in better performance on several benchmarks. This includes addressing challenges related to data privacy, real-time processing, and ease of use.

Setting an optimal learning rate is essential to ensure that the model converges to a solution efficiently without overshooting or getting stuck in local minima. Similarly, the batch size, which dictates the number of training samples used in each iteration for updating the model parameters, can also significantly impact both the training

speed and the model's ability to generalize. A smaller batch size often provides a regularizing effect and lower generalization error, while a larger batch size usually allows the model to converge faster but may lead to overfitting. The dropout rate, another hyperparameter, is often used as a regularization technique to prevent overfitting by randomly setting a fraction of input units to zero during training. To systematically explore the optimal combination of these hyperparameters, various search methods can be utilized. Grid search is one of the most straightforward approaches, wherein a predefined set of values for each hyperparameter is exhaustively tested to find the combination that yields the best performance. However, grid search can be computationally expensive and time-consuming, especially when dealing with a large number of hyperparameters and possible values. As an alternative, Bayesian optimization can be employed. Unlike grid search, Bayesian optimization builds a probabilistic model of the objective function and uses it to select the most promising hyperparameters to evaluate in the actual objective function, thereby making the search process more efficient. Exploring how these models can be seamlessly integrated into existing medical imaging systems would be a practical direction.

References

1. Sheller MJ, Reina GA, Edwards B, Martin J, Bakas S. Multi-Institutional Deep Learning Modeling Without Sharing Patient Data: A Feasibility Study on Brain Tumor Segmentation. *Brainlesion*. 2019;11383:92-104.
2. Iqbal S, Ghani Khan MU, Saba T, et al. Deep learning model integrating features and novel classifiers fusion for brain tumor segmentation. *Microsc Res Tech*. 2019;82(8):1302-1315.
3. Kamnitsas K, Ferrante E, Parisot S, et al. DeepMedic for Brain Tumor Segmentation. In: *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. Springer International Publishing; 2016:138-149.
4. Pan SJ, Yang Q. A survey on transfer learning. *IEEE Trans Knowl Data Eng*. 2010;22(10):1345-1359.
5. Wei Y, Zhang Y, Huang J, Yang Q. Transfer Learning via Learning to Transfer. In: Dy J, Krause A, eds. *Proceedings of the 35th International Conference on Machine Learning*. Vol 80. Proceedings of Machine Learning Research. PMLR; 10--15 Jul 2018:5085-5094.
6. Zamir AR, Sax A, Shen W, Guibas L, Malik J, Savarese S. Taskonomy: Disentangling Task Transfer Learning. In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE; 2018:3712-3722.
7. Tan M, Le QV. EfficientNet: Rethinking model scaling for convolutional Neural Networks. Chaudhuri K, Salakhutdinov R, eds. *arXiv [csLG]*. Published online 09--15 Jun 2019:6105-6114. <https://proceedings.mlr.press/v97/tan19a.html>
8. Qin X, Wang Z. NASNet: A Neuron Attention Stage-by-Stage Net for Single Image Deraining. *arXiv [eessIV]*. Published online December 6, 2019. https://www.researchgate.net/profile/Xu-Qin-12/publication/337830328_NASNet_A_Neuron_Attention_Stage-by-Stage_Net_for_Single_Image_Deraining/links/5e193ee3299bf10bc3a34e94/NASNet-A-Neuron-Attention-Stage-by-Stage-Net-for-Single-Image-Deraining.pdf
9. Liu W, Zeng K. SparseNet: A Sparse DenseNet for Image Classification. *arXiv [csCV]*. Published online April 15, 2018. <http://arxiv.org/abs/1804.05340>
10. Huang G, Liu S, van der Maaten L, Weinberger KQ. CondenseNet: An Efficient DenseNet using Learned Group Convolutions. *arXiv [csCV]*. Published online November 25, 2017:2752-2761. http://openaccess.thecvf.com/content_cvpr_2018/html/Huang_CondenseNet_An_Efficient_CVPR_2018_paper.html
11. Iandola F, Moskewicz M, Karayev S, Girshick R, Darrell T, Keutzer K. DenseNet: Implementing Efficient ConvNet Descriptor Pyramids. *arXiv [csCV]*. Published online April 7, 2014. <http://arxiv.org/abs/1404.1869>
12. Zeng M, Xiao N. Effective combination of DenseNet and BiLSTM for keyword spotting. *IEEE Access*. 2019;7:1-1.

13. Zhang K, Guo Y, Wang X, Yuan J, Ding Q. Multiple Feature Reweight DenseNet for Image Classification. *IEEE Access*. 2019;7:9872-9880.
14. Lin C, Li L, Luo W, Wang KCP, Guo J. Transfer learning based traffic sign recognition using Inception-v3 model. *Period Polytech Trans Eng*. 2018;47(3):242-250.
15. Tio AE. Face shape classification using Inception v3. *arXiv [csCV]*. Published online November 15, 2019. <http://arxiv.org/abs/1911.07916>
16. Mazurowski MA, Clark K, Czarnek NM, Shamsesfandabadi P, Peters KB, Saha A. Radiogenomics of lower-grade glioma: algorithmically-assessed tumor shape is associated with tumor genomic subtypes and patient outcomes in a multi-institutional study with The Cancer Genome Atlas data. *J Neurooncol*. 2017;133(1):27-35.
17. Buda M, Saha A, Mazurowski MA. Association of genomic subtypes of lower-grade gliomas with shape features automatically extracted by a deep learning algorithm. *Comput Biol Med*. 2019;109:218-225.