EFFECTS OF ADAM OPTIMIZER VARIANTS ON BRAIN TUMOR SEGMENTATION TASK

Samson Offorjindu,
[0009-0001-2928-0654]

Marina Marjanović,
[0000-0002-9928-6269]

Timea Bezdan*
[0000-0001-6938-6974]

Singidunum University,
Belgrade, Srebia

Abstract:

In medical image analysis, accurately segmenting brain tumors is still very challenging, motivating researchers to explore advanced deep-learning methods. While U-Net models have produced promising results, improving their performance through optimized training techniques is still necessary. Given that Adam is commonly used as the default optimizer in such tasks, our study explores the impact of different Adam optimizer variants on U-Net performance using the well-known BraTS 2020 dataset. We evaluated Adam, AdamW, Adagrad, Adamax, Adafactor, and RMSprop optimizers, comparing their performance using key metrics such as training loss, validation loss, F-score, Intersection over Union (IoU), precision, and recall. The obtained results show that Adamax achieves the highest F-score (0.8120) and IoU score, demonstrating superior performance in segmenting tumor regions in medical images; AdamW also showed strong results with lower training and validation losses, as well as good precision and recall, highlighting its efficiency and accuracy. These findings emphasize the importance of selecting the right optimizer for Li-Net-based brain tumor segmentation and encourage further exploration into optimized training strategies in medical image analysis.

Keywords:

Artificial Intelligence, Computer Vision, Medical Image Segmentation, Convolutional Neural Network, Deep Learning.

INTRODUCTION

Brain Tumor Segmentation [1-5] has always focused on evaluating state-of-the-art methods for segmenting brain tumors in multimodal magnetic resonance imaging scans (MRIs). The BraTS 2020 dataset, a widely used benchmark in the field, utilizes multi-dimensional pre-operative MRI scans and primarily focuses on the segmentation task of intrinsically heterogeneous brain tumors, likely in appearance, shape, and histology, namely gliomas. Additionally, the dataset includes clinical information such as overall survival, the clinical assessment of disease progression, and uncertainty estimation for expected tumor subregions; BraTS multimodal imaging data include native T1-weighted (T1), post-contrast T1-weighted (T1Gd), T2-weighted (T2), and T2 Fluid Attenuated Inversion Recovery (T2-FLAIR) volumes, provided in Neuroimaging

Correspondence:
Timea Bezdan

e-mail:
tbezdan@singidunum.ac.rs

DOI: 10.15308/Sinteza-2025-41-47

Informatics Technology Initiative - NIfTI (.nii) format, these scans have been collected from various clinical procedures and imaging scanners at multiple institutions, all contributing to the dataset.

This research paper utilized a pre-trained deep-learning model for the brain tumor segmentation task using MRI scans. The model is trained on the BraTS 2020 dataset and is designed to identify and segment the different tumor subregions (whole tumor, tumor core, and enhancing tumor). The primary goal of the experiment is to evaluate the effectiveness of the other Adam optimizer variants in performing accurate segmentations on MRI scans while maintaining recall and precision without additional training. In contrast, we seek to determine whether and how the variations in the Adam optimizer's parameters affect the model's ability to accurately segment tumor subregions while maintaining a reasonable true-positive and true-negative rate. This study's central research question is of paramount importance: To what extent do different variants of the Adam optimizer influence the segmentation performance of a U-Net model (Figure 1) on the BraTS 2020 brain tumor dataset? This question is not only significant for our research but also for the broader field of medical image analysis and deep learning.

The Adam optimizer [6] is a widely recognized algorithm in deep learning and often the default choice; the method leverages the advantages of AdaGrad [7] and RMSProp [8], adapting the learning rates for each parameter based on their historical gradients. However, subtle variations in its parameters influence its convergence behavior and generalization performance. These differences often lead to variations in segmentation accuracy, mostly in complex medical image segmentation tasks where precise boundary outline is critical. We aim to provide a comparative analysis of six popular Adam optimizer variants: Adafactor [9], Adagrad, Adam, Adamax [10], AdamW, and RMSProp, specifically for our segmentation task. By examining their core mechanisms, strengths, and weaknesses, our analysis will provide insights into the suitability of each optimizer. More importantly, by leveraging a pre-trained model, we also aim to minimize computational costs while maintaining good segmentation accuracy, minimizing false rates, and retaining reasonable true positive rates. This research has practical implications for assisting radiologists in clinical decision-making, providing a practical and useable tool for the field of medical image analysis and deep learning.

The rest of this paper is organized as follows: Section 2 describes the methodology and model architecture. Section 3 presents the experimental setup, results, and discussion, and Section 4 concludes the paper with key findings and directions for future work.

2. METHODOLOGY

In digital image processing and computer vision, the process of partitioning an image into various multiple segments is known as the image segmentation task, where the goal is to simplify and transform the image representation to facilitate more efficient and accurate analysis. Image segmentation primarily identifies boundaries and objects, such as lines, curves, and other structures, within images. Image segmentation involves assigning a label to each pixel in an image so that pixels with the same label share a common set of characteristics. Our goal is to assign a unique label to each pixel, thereby outlining tumor boundaries in medical images.

The U-Net architecture is the most suitable architecture for our segmentation task due to its proven efficiency in medical image segmentation and its ability to handle limited training data, which results from several key characteristics:

i. Its encoder-decoder structure, based on the fully convolutional neural network principles proposed in [11], effectively captures both high-level contextual information and low-level spatial details. The encoder extracts hierarchical features while the decoder reconstructs the segmentation map. Hence, a compelling feature hierarchy capture.

ii. Implementing skip connections that bridge the encoder and decoder enables the propagation of the finely grained spatial information to higher-resolution layers. This is particularly vital in medical image segmentation, where precisely defining tumor boundaries is challenging. It enhances its robustness with limited data.

iii. The output of pixel-wise segmentation maps directly addresses the need for precise, detailed tumor subregion delineation.

The core concept involves augmenting a conventional contracting network by substituting pooling operations with upsampling operators. Consequently, these layers enhance the output resolution. Moreover, this data enables a subsequent convolutional layer to construct a precise output. A key innovation of U-Net is the increased density of feature channels in the upsampling pathway, facilitating the propagation of contextual information to higher-resolution layers. Therefore, the expansive pathway generates a U-shaped architecture and exhibits approximate symmetry with the contracting component. This tiling strategy is essential for applying the network to large images, as GPU memory limitations would otherwise constrain the resolution. The network utilizes only the valid portion of each convolution, excluding fully connected layers. The missing contextual information is extrapolated through input image mirroring to predict pixels in the image's border region.

Our research utilizes the proposed U-Net model for the tumor segmentation task. The model comprises three primary components: the ResNet encoder, a U-Net decoder, and a segmentation head. We utilized a ResNet encoder to extract a robust hierarchical feature. Its residual connections effectively address the vanishing gradient problem, enabling us to train deeper networks and capture more complex feature representations. The encoder begins with an initial convolutional layer, followed by batch normalization, ReLU activation, and max pooling. Subsequently, four sequential layers, composed of multiple Bottleneck blocks, perform feature extraction. Bottleneck blocks enhance computational efficiency while maintaining representational power. ResNet was used due to its strong ability to learn deep feature representations, which is particularly beneficial for analyzing complex medical images.

Moving on from the model, we will look at the Adam variant optimizers. Firstly, Adafactor offers significant memory efficiency, a crucial advantage when training large 2D U-Net models on high resolution images, potentially allowing for more complex architectures with limited resources. It provides adaptive learning rates that effectively handle the dataset's diverse intensity distributions and tumor characteristics. However, it possesses a slower convergence, which increases overall training times. Additionally, Adagrad adapts the learning rate for each parameter based on the cumulative history of gradients, which is beneficial for sparse features that might emerge. However, its weakness is the probability that the learning rate decays aggressively over time, leading to a slow convergence. Adam itself is a widely adopted optimizer that combines adaptive learning rates with momentum, making it computationally efficient; it possesses a deep, faster convergence due to its adjustments of individual parameter learning rates. Another variant of this optimizer uses the infinity norm and exhibits a stable behavior across different learning rates and gradient scales; Adamax exhibits robustness to extreme gradients, translating to a more stable training process. However, Adamax can still be sensitive to hyperparameter selection.

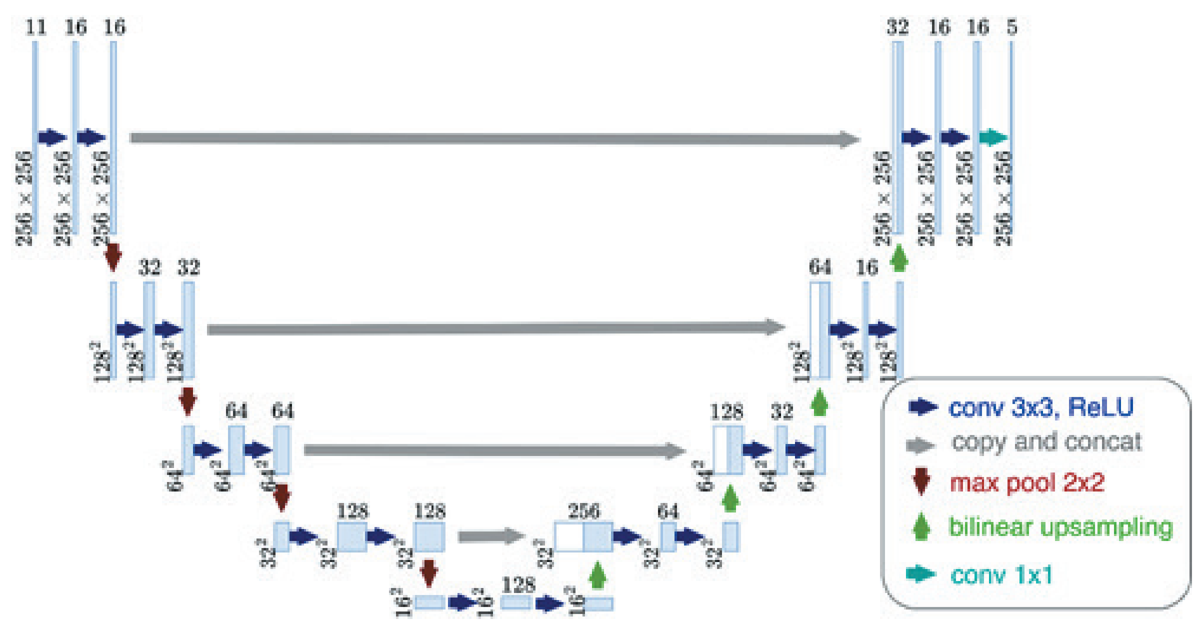

Figure 1. Model architecture

Regarding the Adam optimizer, AdamW improves upon the standard Adam optimizers by decoupling the weight decay process, leading to a more effective L2 regularization [12], giving it a higher performance potential for this task. Lastly, RMSprop is another optimizer that uses a moving average of squared gradients to normalize each parameter's learning rate, allowing it to handle parameters with diverse features. However, its performance is sensitive to the choice of hyperparameters and its slow convergence in most cases.

3. EXPERIMENTAL RESULTS AND DISCUSSION

The experimental setup in this study is designed to evaluate the performance of various Adam optimizer variants when training a U-Net model for brain tumor segmentation. The goal is to conduct a systematic comparison of these optimizers to determine how each influences segmentation accuracy, training dynamics, and generalization ability. By analyzing key performance metrics, we aim to reveal the strengths and limitations of each optimizer variant within the context of medical image segmentation.

3.1. EXPERIMENTAL SETUP AND DATASET

Dataset preparation and setup for training involved key stages to ensure proper data organization, preprocessing, and transformations. The dataset was initially sourced from the BraTS2020 training data, which is structured by extracting subject IDs from the CSV holding name-mapping data. Afterward, file paths are organized into separate categories corresponding to different MRI modalities, including the FLAIR, T1, T1CE, T2, and Segmentation masks (Figure 2), after which the currently structured dataset is subsequently stored in a CSV file for efficient access. The next stage involves loading the images using the NiBabel library that allows handling medical image files such as the one in question, NIfTI (.nii). The images are passed onto a center cropping function, which ensures that the images maintain uniform dimensions across the dataset. Normalization is then applied to standardize pixel intensities, which contributes to stabilizing the training process, speeds up the training process, and improves the model's convergence.

Furthermore, data augmentation techniques were carried out to enhance model generalization. We utilized augmentation libraries from Albumentations and TorchVision for transformation operations such as image flipping, rotation, and contrast adjustments. These steps were taken to collectively ensure that the dataset is well-prepared for training our deep learning model - as represented in Table 1.

Additionally, the dataset was divided into training (60%), validation (20%), and test (20%) sets. Individual MRI slices were extracted and stored in .npy format for good loading and processing. The U-Net model was trained using the previously mentioned optimizers, each with a learning rate 0.0001. At the same time, dice loss was applied to the model's softmax2d output to optimize multi-class segmentation, and the batch size was set to 16 for training. We evaluated model performance using several metrics, including Intersection over Union (IoU) with a threshold of 0.5, F-score, precision, recall, training loss, and validation loss.

3.2. EXPERIMENTAL RESULTS AND DISCUSSION

This subsection describes the results of the experiments conducted using the Adam optimizer variants mentioned to train the U-Net model on the BraTS2020 dataset. The performance of these optimizer variants in terms of IoU, precision, recall, and F-score, as well as train and validation losses, provided insights into why Adamax had the best results. The comparative analysis is presented in Table 2.

Table 1. Parameter configuration

Parameter	Value
Batch size	16
Learning rate	1e-4
Loss function	DiceLoss
Activation function	ReLU
Epochs	200

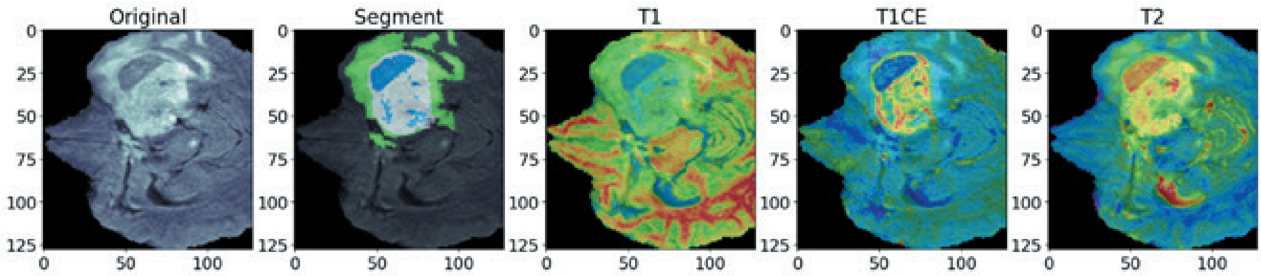

Figure 2. Sample modality visualizations

Table 2. Comparative analysis

Metric	Adam	AdamW	Adagrad	Adamax	Adafactor	RMSprop
Training loss	0.0025	0.0027	0.0171	0.0030	0.0097	0.0026
Validation loss	0.0108	0.0108	0.0183	0.0104	0.0124	0.0108
F-score	0.7964	0.7979	0.7105	0.8120	0.7661	0.8027
IoU-score	0.7087	0.7105	0.6125	0.7238	0.6743	0.7136
Precision	0.8711	0.8684	0.8670	0.8678	0.8713	0.8521
Recall	0.7810	0.7860	0.6869	0.8005	0.7486	0.7978

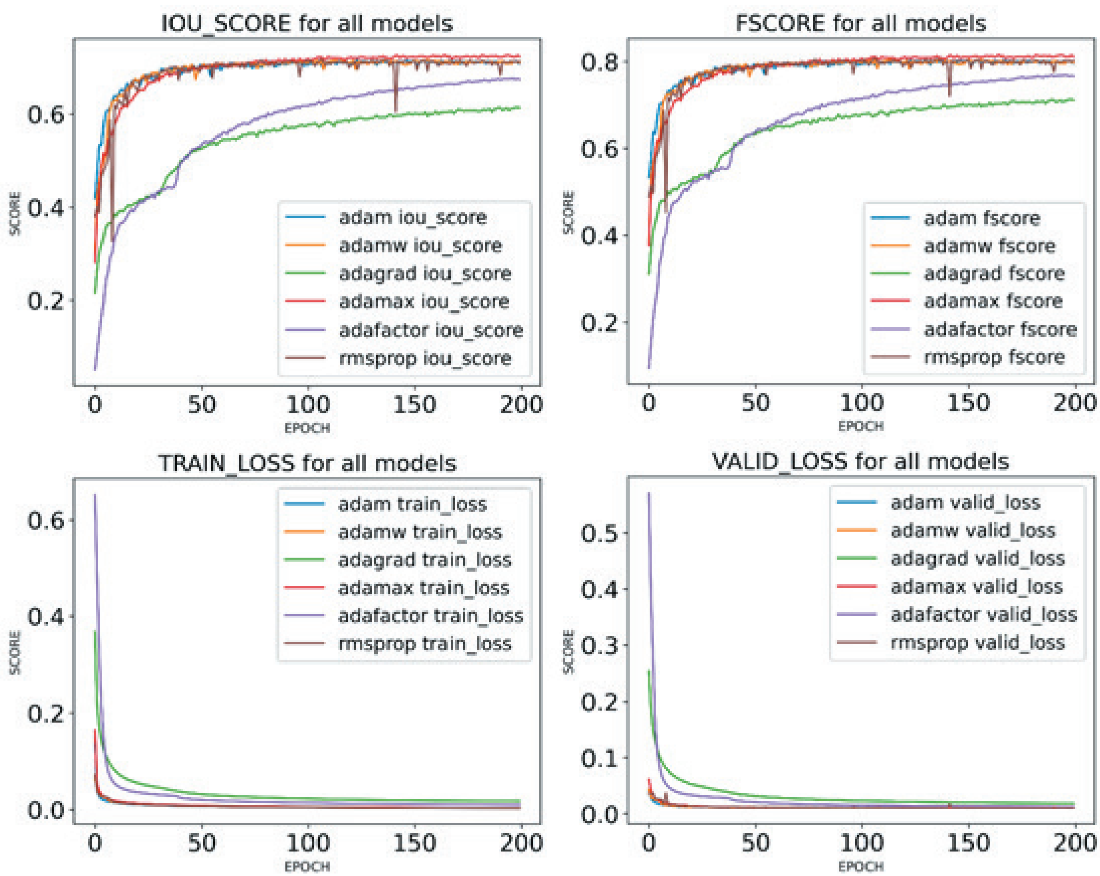

Figure 3-6. The first plot (top left) shows the Intersection of Union score across the epochs. The second plot (top right) displays the F-score with the third (on the right) and the validation loss across epochs

The experimental results (Figure 3-6) first showed a rapid initial loss reduction in both training and validation. The training dynamics across these optimizers showed a consistent pattern. Most of the methods show a rapid loss reduction, with convergence to stable loss values between 25 and 30 epochs; the training and validation curves mirrored each other with exceptional precision, including minimal overfitting and robust learning mechanisms, which suggests that the Adam optimizer family that is characterized by adaptive learning rates and momentum-based updates are very-well suited for the task. Within the first 75-100 epochs, we also observed stabilization for the IoU scores and precision and recall, after which the performance gains became increasingly marginal. This observation underscored the importance of early training stages and suggested that extended training may yield diminishing returns. However, Adamax's infinity norm variant appeared adept at handling parameter magnitudes variations, which should explain its performance advantage.

The precision scores, which represent the ability of the model to avoid false positives, were relatively similar across all three optimizers, with Adam achieving the highest precision (0.8711), followed closely by AdamW (0.8684) and Adamax (0.8678). This suggests that all three optimizers are reasonably effective at minimizing the prediction of tumor regions where none exist. However, Adamax demonstrated the highest recall (0.8005), followed by AdamW (0.7860) and Adam (0.7810). Recall measures the ability of the model to identify all actual positive cases, meaning Adamax was more successful in detecting all the tumor regions present in the BRATS2020 dataset compared to the other two optimizers. Overall, the initial analysis of the performance metrics suggests that Adamax outperformed both Adam and AdamW on key segmentation metrics and generalization ability despite a slightly higher training loss. AdamW showed a marginal improvement over Adam in terms of the F-score and IoU score while maintaining the same validation loss.

While these results, as well as the predicted segmentations (Figure 7) seem rather compelling, they were not without limitations. Our analysis and experiments were conducted on a single dataset and experimental setup, which implies that generalizability will require further investigation and study. Nevertheless, the findings align closely with the existing context of adaptive optimization techniques, reinforcing the effectiveness of the Adam optimizer family. From a practical perspective, the results provide a clear recommendation. Adamax emerges as the preferred choice for tasks requiring a balanced performance, while the standard Adam serves as a robust alternative. The minimal variability between the tested optimizers suggests that other researchers and practitioners can confidently select from this optimizer family with a relatively low risk of significant performance degradation.

4. CONCLUSION

The findings of this study provide practical insights for researchers in medical image segmentation, particularly those using the BraTS dataset. Based on the obtained results, Adamax is the most effective optimizer among the other evaluated optimizers, offering superior segmentation accuracy and generalization capabilities. The outcome also highlights the potential benefits of Adamax's robustness to extreme gradients and noisy data, which are a widespread challenge in medical image analysis and many other deep-learning tasks.

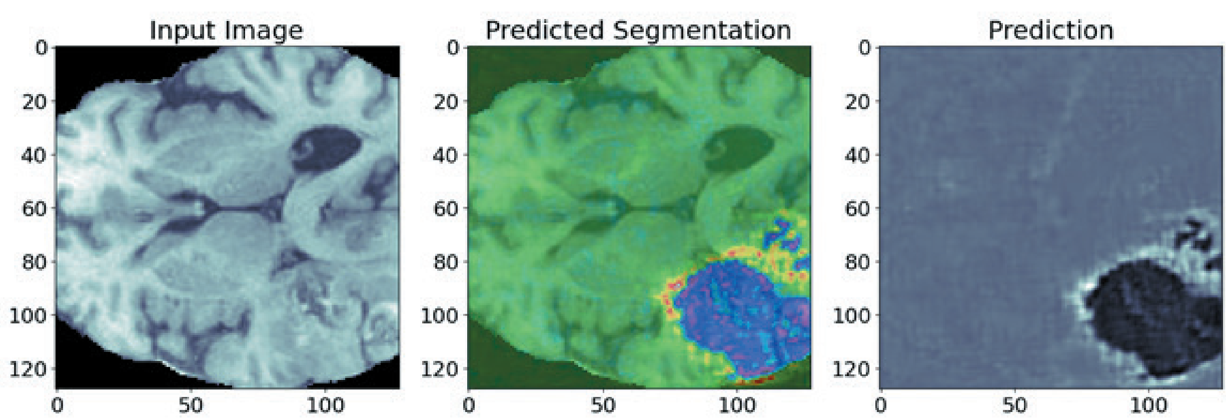

Figure 3. Visualization for the model's predicted segmentation sample

Hence, for tasks that resemble or are of the BraTS challenge, optimizers such as Adamax stand as strong candidates due to their ability to handle the complexities of medical image gradients. However, it is crucial to be mindful of its potential sensitivity to hyperparameter selection and conduct thorough tuning. AdamW remains a generally recommended optimizer for deep learning tasks, especially when dealing with complex datasets such as BraTS 2020, as well as models where regularization plays a vital role. The slight improvement observed over ordinary Adam further supports its use in medical image segmentation to achieve a good balance between performance and stability. Finally, future research should investigate the performance of these optimizers further across different BraTS datasets, as well as other models, potentially looking to hybrid models that utilize U-Net as a base model - potentially with more extensive hyperparameter tuning.

REFERENCES

[1] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, et al. "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)", *IEEE Transactions on Medical Imaging* 34(10), 1993-2024 (2015) DOI: 10.1109/TMI.2014.2377694

[2] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J.S. Kirby, et al., "Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features", *Nature Scientific Data*, 4:170117 (2017) DOI: 10.1038/sdata.2017.117

[3] S. Bakas, M. Reyes, A. Jakab, S. Bauer, M. Rempfler, A. Crimi, et al., "Identifying the Best Machine Learning Algorithms for Brain Tumor Segmentation, Progression Assessment, and Overall Survival Prediction in the BRATS Challenge", *arXiv preprint arXiv*:1811.02629 (2018)

[4] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. Kirby, et al., "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection", *The Cancer Imaging Archive*, 2017. DOI: 10.7937/K9/TCIA.2017.KLXWJJ1Q

[5] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. Kirby, et al., "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection", *The Cancer Imaging Archive*, 2017. DOI: 10.7937/K9/TCIA.2017.GJQ7R0EF

[6] Z. Zhang, "Improved Adam Optimizer for Deep Neural Networks," *2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS)*, Banff, AB, Canada, 2018, pp. 1-2, doi: 10.1109/IWQoS.2018.8624183.

[7] J. -K. Fang, C. -M. Fong, P. Yang, C. -K. Hung, W. -L. Lu and C. -W. Chang, "AdaGrad Gradient Descent Method for AI Image Management," *2020 IEEE International Conference on Consumer Electronics - Taiwan (ICCE-Taiwan)*, Taoyuan, Taiwan, 2020, pp. 1-2, doi: 10.1109/ICCE-Taiwan49838.2020.9258085.

[8] Huk, M. (2020). Stochastic Optimization of Contextual Neural Networks with RMSprop. In: Nguyen, N., Jearanaitanakij, K., Selamat, A., Trawiński, B., Chittayasothorn, S. (eds) Intelligent Information and Database Systems. ACIIDS 2020. *Lecture Notes in Computer Science*(), vol 12034. Springer, Cham.

[9] Shazeer, Noam & Stern, Mitchell. (2018). Adafactor: Adaptive Learning Rates with Sublinear Memory Cost. 10.48550/arXiv.1804.04235.

[10] C. Arora, G. Raj, A. Ajit and A. Saxena, "ADAMAX-Based Optimization of Efficient Net V2 for NSFW Content Detection," *2023 IEEE International Conference on Contemporary Computing and Communications (InC4)*, Bangalore, India, 2023, pp. 1-6, doi: 10.1109/InC457730.2023.10263203.

[11] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 3431–3440.

[12] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA: MIT Press, 2016.