Dean&Francis
ISSN 2959-6157

Brain Tumor Classification in MRI Images Based on EfficientNetB0: Impact of Learning Rate, Optimizer and Batch Size

Weiye Dong

Department of Electrical Engineering and Automation, Hebei University of Water Resources and Electric Engineering, Cangzhou, China
zooincapital@gmail.com

Abstract:

Brain tumor classification based on Magnetic Resonance Imaging (MRI) images is essential for early diagnosis, yet traditional manual interpretation remains time-consuming and error-prone. This research delves into the ways in which deep learning models for brain tumor classification are impacted by changes in learning rate, optimizer, and batch size. The backbone model was created by training on a Kaggle brain MRI dataset that contained 3,264 grayscale images of four categories: glioma tumor, meningioma tumor, pituitary tumor, and no tumor. Preprocessing included normalization, grayscale-to-RGB conversion, and data augmentation. Using weights pre-trained by ImageNet, transfer learning was implemented. The model's final layer was adjusted for four-class output with Softmax activation. Experiments were conducted across four learning rates, two optimizers, and five batch sizes. There were 50 training iterations for each configuration, with validation loss serving as the basis for early termination. The following metrics were used to assess performance: recall, accuracy, precision, F1-score, and confusion matrices. Experimental results show that Adam generally delivers more stable and accurate outcomes than Stochastic Gradient Descent (SGD) under similar conditions. Higher learning rates combined with moderate batch sizes led to optimal performance. Pituitary tumors were accurately identified, while no_tumor and meningioma were often confused. These findings demonstrate how crucial hyperparameter adjustment is to enhance the accuracy of deep learning-based Identifying brain tumors.

Keywords: Brain tumor classification; EfficientNetB0; Magnetic Resonance Imaging (MRI).

ISSN 2959-6157

1. Introduction

Neurology has some of the most serious and life-threatening conditions, including brain tumors. Cells in the brain grow abnormally and uncontrollably, according to the characterization. These tumors can disrupt vital neurological functions, cause seizures, and in many cases, lead to death. According to data from the World Health Organization (WHO), brain and central nervous system tumors caused over 300,000 new cases worldwide in 2020. resulting in approximately 250,000 deaths [1]. glioma tumors, meningioma tumors, and pituitary tumors are the most common types of brain tumors, with either benign or malignant classification being the norm. Glioblastoma multiforme, one of the most aggressive forms, has a dismal five-year survival rate below 10% [2].

Detecting brain tumors at an early stage is critical to ensuring timely treatment. Traditionally, this relies heavily on radiologists manually examining Magnetic Resonance Imaging (MRI) scans require a lot of effort and take a lot of time that can vary from one specialist to another. In this situation, Artificial Intelligence (AI), especially deep learning models, can become useful assistants. These models can learn to extract meaningful features from huge amounts of imaging data and support clinicians in making faster, more reliable diagnostic decisions. By doing so, AI not only improves workflow efficiency but also minimizes the potential for human error [3]. In view of the global increase in brain tumor cases, integrating AI into diagnostic systems is both timely and essential.

Over the past decade, deep learning has revolutionized many scientific disciplines. Breakthroughs such as OpenAI's GPT-4 in natural language processing and DeepMind's AlphaFold in protein structure prediction have showcased in some complex cognitive tasks, machine learning models can match or surpass human performance [4, 5]. In healthcare, deep learning algorithms have demonstrated promising results across various diagnostic domains. For example, neural networks have been successfully applied to detect pneumonia in chest X-rays [6], evaluate cardiac risks [7], and diagnose diabetic retinopathy from retinal images [8].

The field of brain tumor classification has also witnessed significant advancements. In 2016, Pereira and colleagues presented a Convolutional Neural Network (CNN)-based approach for analyzing brain tumors using MRI data, which showed marked improvements over classical image processing methods [9]. Subsequently, Afshar et al. introduced capsule networks (CapsNet), which enhanced the model' s ability to preserve spatial hierarchies, leading to more accurate tumor classification [10]. More recent studies have leveraged attention mechanisms and multi-scale architectures to refine performance even further [11]. Despite these achievements, much of the current literature has centered around improving model design and boosting accuracy. However, the effect of tuning hyperparameters, such as learning rate, optimizer selection, and batch size, on classification results have not received the attention it deserves. These parameters play an important role in training dynamics and model generalization, and understanding their influence is crucial for practical deployments [12].

This paper aims to conduct a detailed analysis of how different hyperparameter settings influence the performance of neural network models in classifying brain tumor images. Using well-established brain MRI datasets, deep learning model called EfficientNetB0 across a range of rates that can be learned (0.0001, 0.001, 0.01, 0.1), optimizer algorithms (SGD and Adam), and batch sizes (8, 16, 32, 64, 128) are evaluated. Model effectiveness was assessed through multiple indicators, including overall accuracy, class-level precision and recall, the F1 metric, and confusion matrix analysis.

2. Method

2.1 Dataset Preparation

Kaggle was the source of the Brain Tumor Classification dataset [13]. Collection contains 3,264 grayscale MRI images sorted by category, with glioma tumor, meningioma tumor, pituitary tumor, and no tumor. Brain tumors, which make up 85% to 90% of all primary central nervous system (CNS) tumors, are depicted in the dataset as real-world clinical challenges. The five-year survival rate remains low, approximately 34% for males and 36% for females, emphasizing the urgency of early and accurate diagnosis. Magnetic Resonance Imaging continues to be the gold standard in diagnosing brain tumors. However, the time-consuming and error-prone manual interpretation process for radiologists is caused by the intricate structure of brain tissues and the visual similarities between different tumor types. The five-year survival rate remains low, approximately 34% for males and 36% for females, emphasizing the urgency of early and accurate diagnosis. MRI remains the most effective modality for brain tumor detection. However, manual interpretation by radiologists involves a lot of time and is prone to errors due to the intricate structure of brain tissues and the visual similarities among different tumor types.

Preprocessing steps included normalization of pixel values to the [0,1] range, enhancing training stability and convergence speed. The dataset was partitioned into training (88%) and testing (12%) subsets. In order to improve the

training data and prevent overfitting, various techniques, such as horizontal and vertical flipping, rotation, zooming, and brightness adjustments, were utilized to simulate real-world variability in medical imaging.

2.2 EfficientNetB0-Based Classification

Convolutional neural networks have proven highly effective for image classification, primarily due to their capability to extract multi-level spatial features from input images. The EfficientNetB0 model was chosen as the backbone network for this investigation.

The network's depth, width, and input resolution are all balancedly scaled by this architecture using a compound scaling technique. Such design allows it to achieve high accuracy while maintaining computational efficiency [14]. Swish activation functions and Mobile Inverted Bottleneck Convolutions (MBConv) are also included in EfficientNetB0 to enhance performance and lower parameter overhead. To take advantage of prior knowledge, the network was initialized with pre-trained ImageNet weights, implementing transfer learning. The final dense layer was modified to produce four outputs, corresponding to the target classes. To make the model appropriate for multi-class classification, a Softmax activation function was used to this layer, converting raw scores into class probabilities.

2.3 Implementation Details

The entire model development and experimentation process was carried out using TensorFlow 2.x with the Keras high-level API. This study methodically investigated how important hyperparameters, such as learning rate, optimizer, and batch size, impact the model's classification performance. Four learning rates, 0.0001, 0.001, 0.01, and 0.1, were evaluated to examine their effects on convergence speed and accuracy. In addition, the suitability of Stochastic Gradient Descent (SGD) and Adam for this task was determined by comparing two commonly used optimization algorithms. To assess the impact on training stability and model generalization, some batch sizes were tested, such as 16, 32, 64, and 128.

Early stopping was implemented to prevent overfitting and ensure optimal model performance during each training session, which was run for 50 epochs. For multi-class classification problems, categorical cross-entropy was used as the loss function. Although the original dataset consisted of grayscale images, each image was converted into a three-channel format by duplicating the grayscale channel. This preprocessing step ensured compatibility with EfficientNetB0's input requirements, resulting in an input shape of (224, 224, 3).

To comprehensively evaluate the model' s effectiveness, multiple performance metrics were adopted. Accuracy was used to measure the predictability of predictions as a whole, while precision assessed the proportion of positive predictions that are true and those that are not. Recall quantified the percentage of true positive predictions among all actual positive cases. The F1-score, as the harmonic mean of precision and recall, provided a balanced measure of the model's predictive ability. Moreover, the confusion matrix was analyzed to show the distribution of prediction errors among different tumor classes, providing valuable insights into the model's specific strengths and weaknesses. Finding the best hyperparameter settings and gaining a better understanding of the model's behavior under varying training conditions is possible with the use of these evaluation metrics as a robust framework.

In addition, this study also generated a series of visualizations to analyze their effects on key performance metrics. By varying learning rates, optimizers, and batch sizes, the model exhibited significant differences in classification outcomes across the metrics of accuracy, precision, recall, and F1-score. These variations were clearly reflected in the curve's charts, as well as in the confusion matrices.

3. Results and Discussion

3.1 Comparison of Adam and SGD Optimizers

3.1.1 General trend with Adam optimizer

Models trained with larger learning rates (LR = 0.01 and 0.1) generally outperformed models with smaller learning rates (LR = 0.0001 and 0.001) using the Adam optimizer. As shown in Fig.1, the precision-macro values for LR = 0.01 and LR = 0.1 are consistently higher than those for smaller LRs. This suggests that higher learning rates help the model converge more effectively with Adam.

Regarding batch size, performance improves as batch size increases up to a point (typically around 40–60), after which it stabilizes or slightly degrades. Fig.2 shows that the accuracy curve for LR = 0.01 increases with batch size and flattens beyond a certain threshold, indicating an optimal batch size range.

ISSN 2959-6157

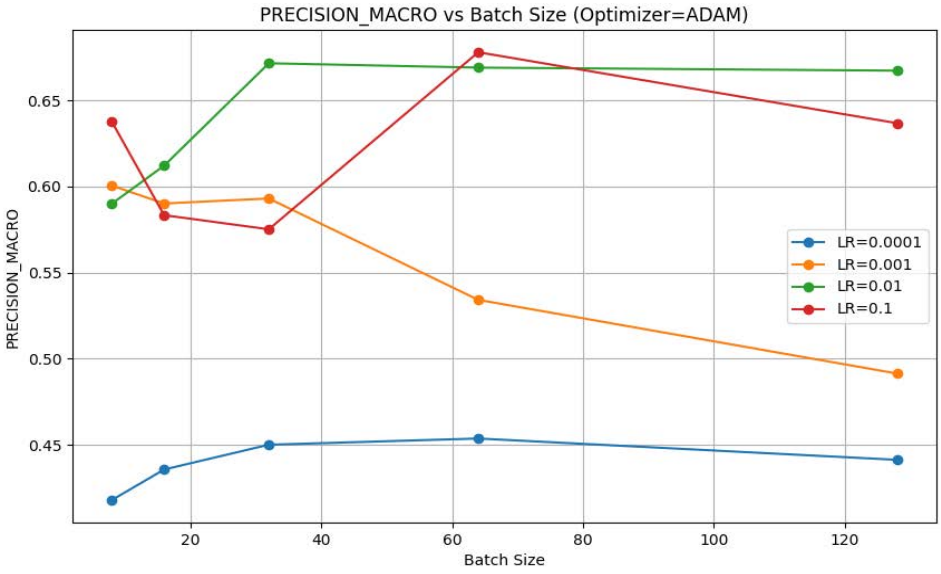

Fig. 1 The influence of learning rates on precision macro based on the Adam optimizer (Picture credit : Original)

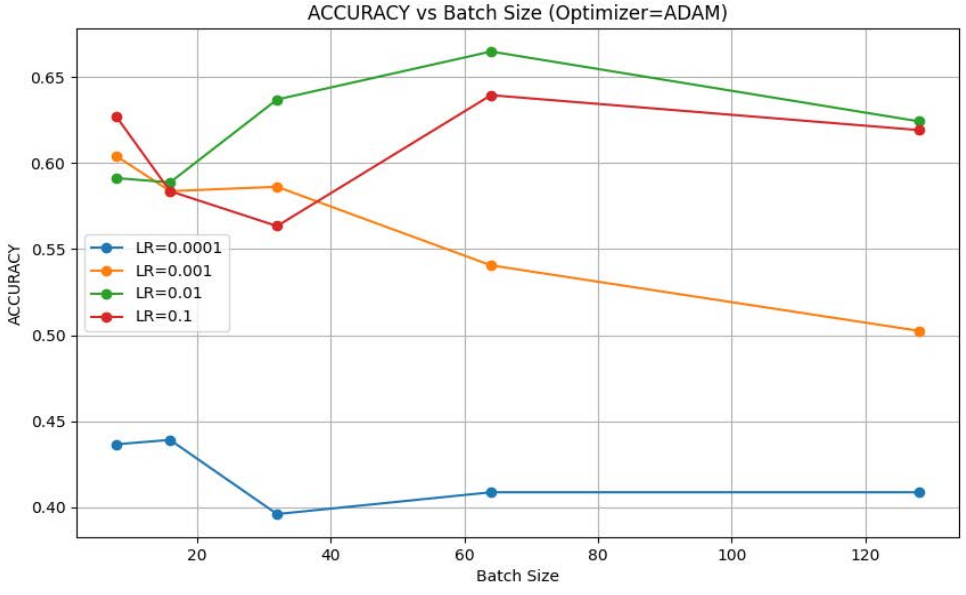

Fig. 2 The influence of batch size on accuracy using the EfficientNetB0 model under the Adam optimizer (Picture credit : Original)

3.1.2 General trend with SGD optimizer

When using the SGD optimizer, performance trends were more unstable. In Fig.3, precision-macro values across different learning rates fluctuated more significantly with changing batch sizes. For instance, LR = 0.001 shows a dip followed by a rise, suggesting sensitivity to parameter tuning. Despite variability, larger learning rates (0.01 and 0.1) tended to perform better on average. As shown in Fig.4, recall-macro for LR = 0.1 and 0.01 was generally higher than for smaller LRs, though not consistently.

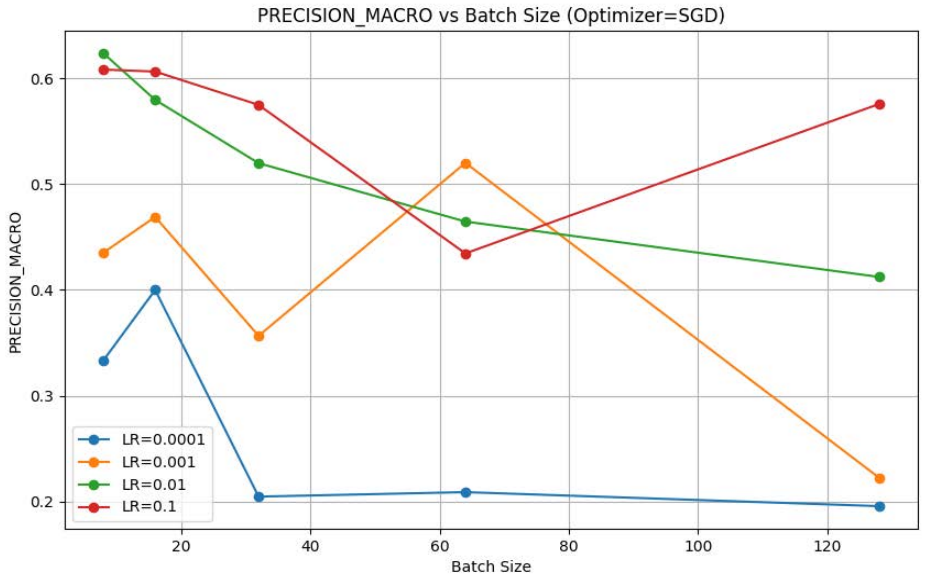

Fig. 3 The influence of batch size on precision macro based on the SGD optimizer (Picture credit : Original)

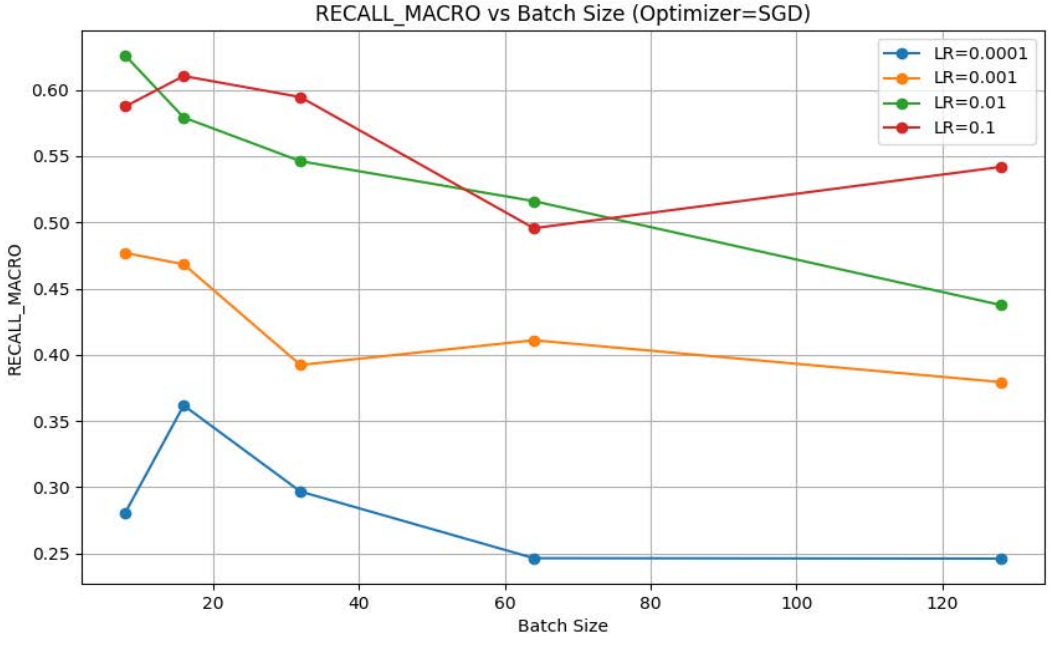

Fig. 4 The influence of batch size on recall macro based on the SGD optimizer (Picture credit : Original)

3.2 Impact of Learning Rate

3.2.1 Small learning rates (0.0001 and 0.001)

Models with small learning rates generally showed lower performance. With Adam, precision and recall values were consistently lower at LR = 0.0001 and 0.001 (Fig.1, Fig.5), and similar trends were observed with SGD (Fig. 3). This indicates limited gradient updates and the possibility of getting stuck in sub-optimal solutions.

ISSN 2959-6157

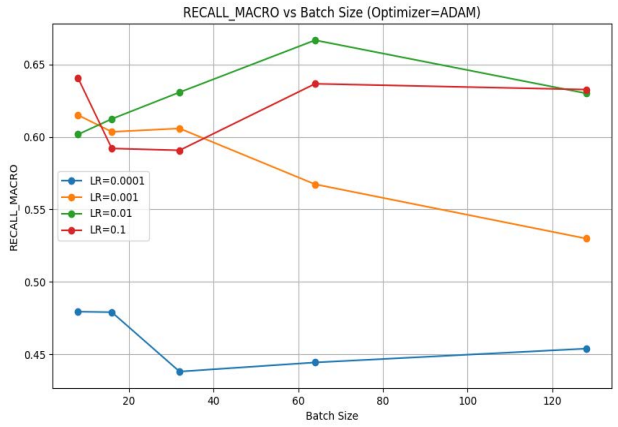

Fig. 5 The influence of learning rate on recall_macro based on the Adam optimizer (Picture credit : Original)

3.2.2 Large learning rates (0.01 and 0.1)

With larger learning rates, Adam achieved higher precision, recall, and F1 scores. Fig.5 illustrates better classification results under LR = 0.01 and 0.1. For SGD, results were more erratic, but performance was still better on average at higher LRs. However, instability can occur with very large learning rates in SGD.

3.3 Optimal Batch Size Analysis

With Adam, performance improved as batch size increased from 10 to about 40–60 and then plateaued, as seen in Fig. 2 and Fig. 5. This suggests that a batch size in this range balances computational cost and model performance.

For SGD, performance was more volatile. While moderate batch sizes (20–60) provided better results on average, identifying an optimal size was difficult due to frequent fluctuations (Fig.3, Fig.4) .

3.4 Confusion Matrix Analysis: Tumor Classification Details

3.4.1 Misclassification patterns by tumor category

Fig. 6 illustrates that the no_tumor category was consistently prone to misclassification, frequently being confused with meningioma_tumor. This implies a notable degree of feature similarity between non-tumor samples and meningioma cases, posing a persistent challenge for classification models. In contrast, pituitary tumor samples were reliably and correctly classified in most configurations, as indicated by prominent diagonal values in the confusion matrices. This suggests that features associated with pituitary tumors are more distinct and easier for the model to learn.

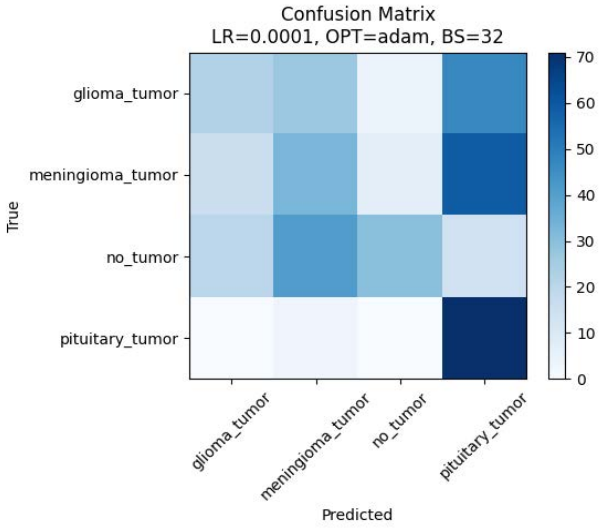

Fig. 6 The influence of class type on confusion matrix based on the misclassification between no tumor and meningioma tumor (Picture credit : Original)

3.4.2 Optimizer comparison: Adam vs. SGD

When using the Adam optimizer, misclassifications between no_tumor and meningioma_tumor decreased gradually as batch size increased at a learning rate of 0.01. In Fig.7, there are four different batch sizes from 16 to 128 (each multiplied by 2) to compare the differences in model performance on image classification tasks. This trend highlights Adam's ability to enhance feature discrimination when training on larger batches.

Furthermore, at a higher learning rate of 0.1, Adam also improved the separation between glioma_tumor and meningioma_tumor, as shown in Fig. 8. In contrast, the SGD optimizer exhibited highly unstable classification behavior. Even under the same learning rate (e.g., LR = 0.1), confusion between glioma and meningioma fluctuated substantially as batch size varied (Fig.9) underscoring SGD's sensitivity to batch-related noise and its difficulty in consistently learning discriminative features.

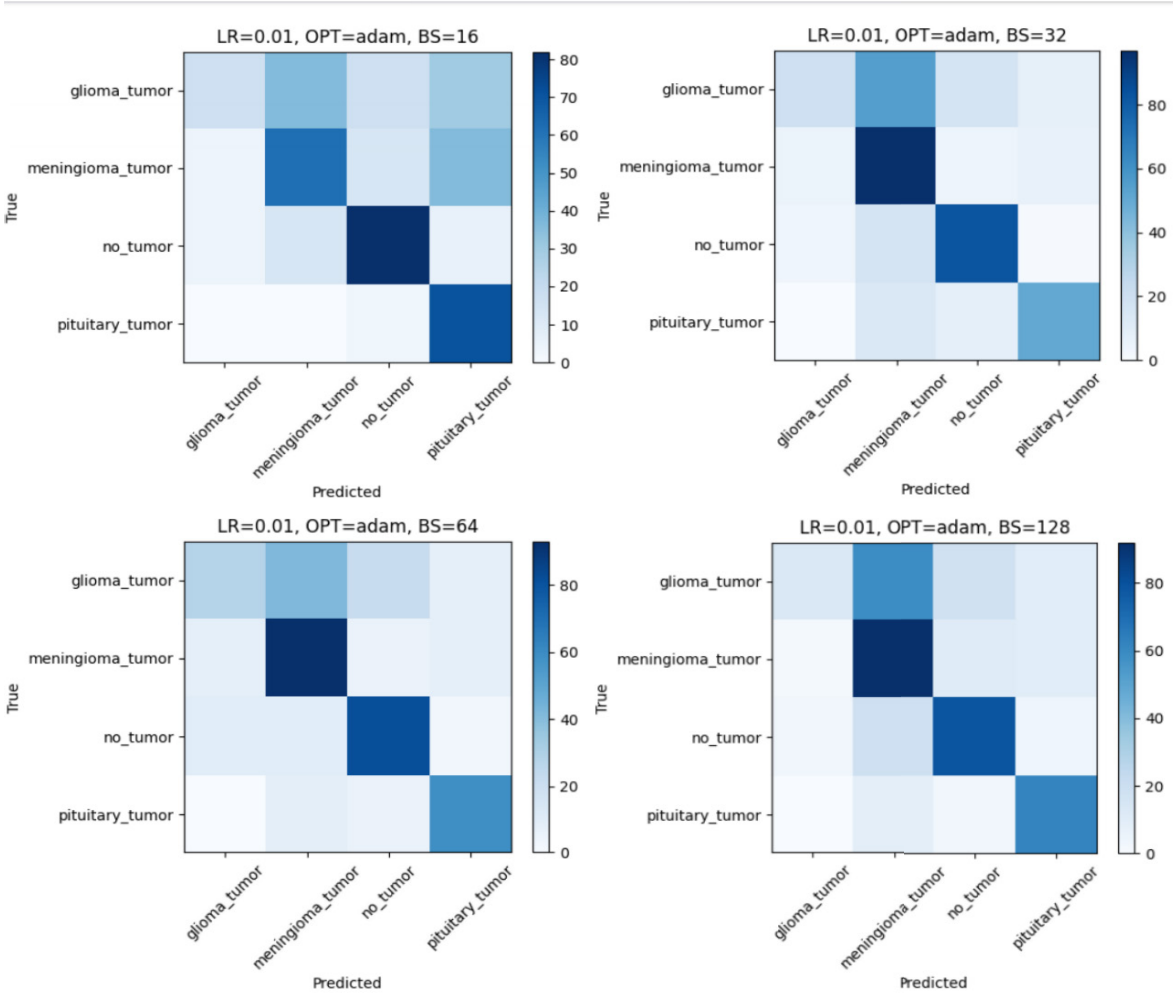

Fig. 7 The influence of batch size (ranging from 16 to 128) on the confusion matrix based on the Adam optimizer with a learning rate of 0.01 (Picture credit : Original)

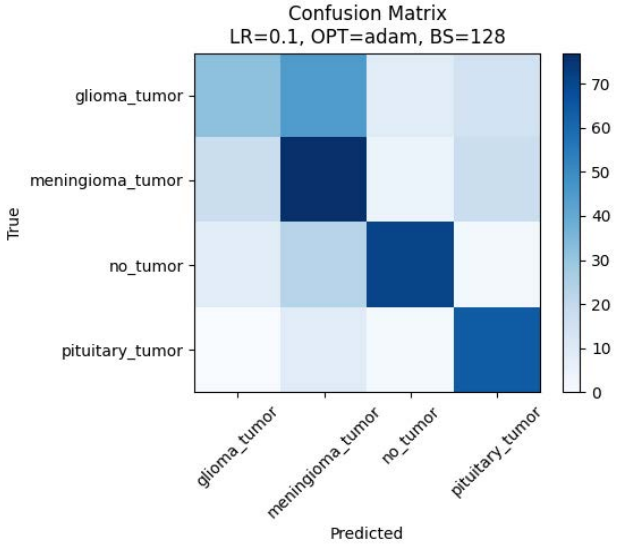

Fig. 8 Confusion matrix with a high learning rate (0.1) using the Adam optimizer and a batch size of 128 (Picture credit : Original)

ISSN 2959-6157

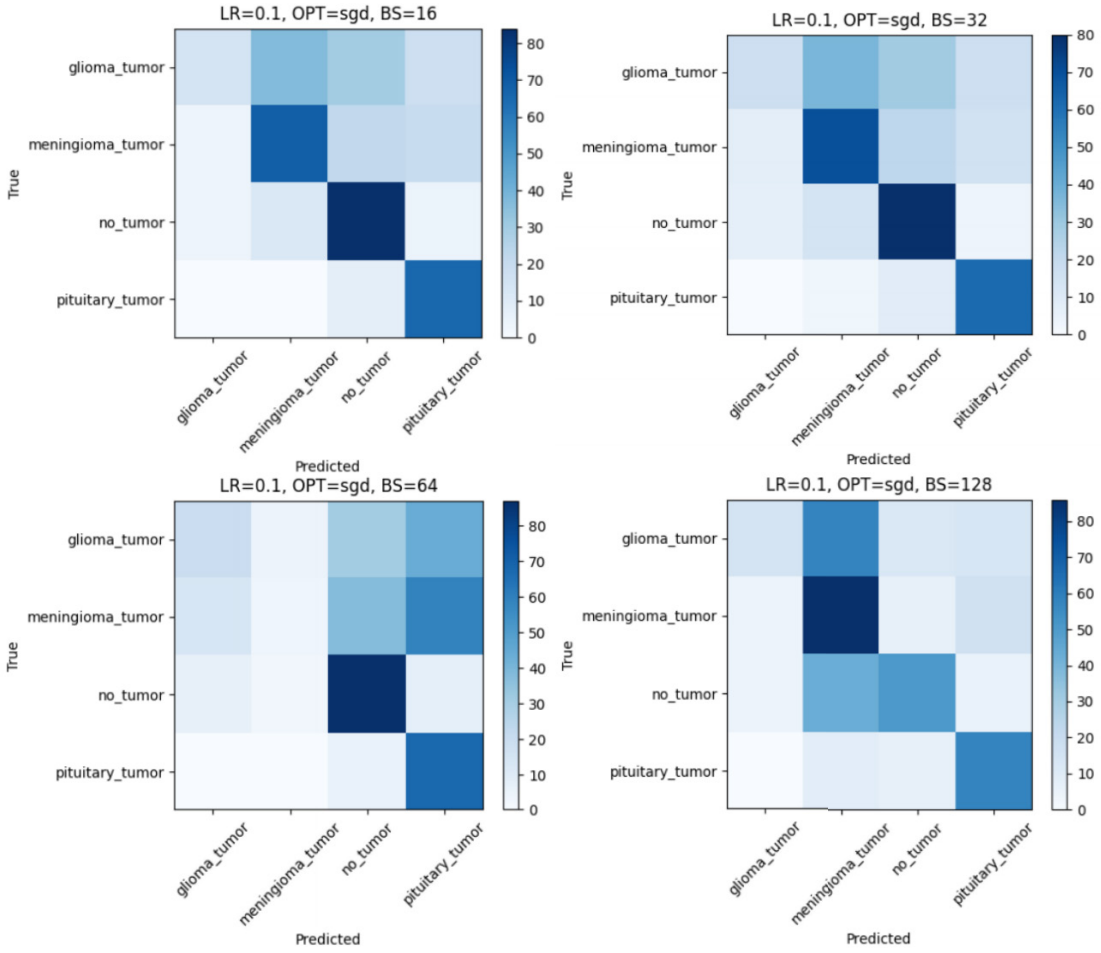

Fig. 9 The influence of batch size (ranging from 16 to 128) on the confusion matrix based on the SGD optimizer with a learning rate of 0.1 (Picture credit : Original)

3.4.3 Learning rate and batch size interaction

At lower learning rates (0.0001 and 0.001), both optimizers struggled to effectively distinguish between no_tumor and meningioma_tumor, as evidenced by persistent off-diagonal errors ,as seen in Fig. 10. This may be due to insufficient gradient updates, limiting the model's ability to refine subtle feature differences.

At higher learning rates (0.01 and 0.1), Adam exhibited steady improvements with increasing batch sizes, confirming its robustness and adaptability. On the other hand, SGD's performance remained inconsistent, suggesting its greater susceptibility to training instability at elevated learning rates or batch sizes, as seen in Fig. 11.

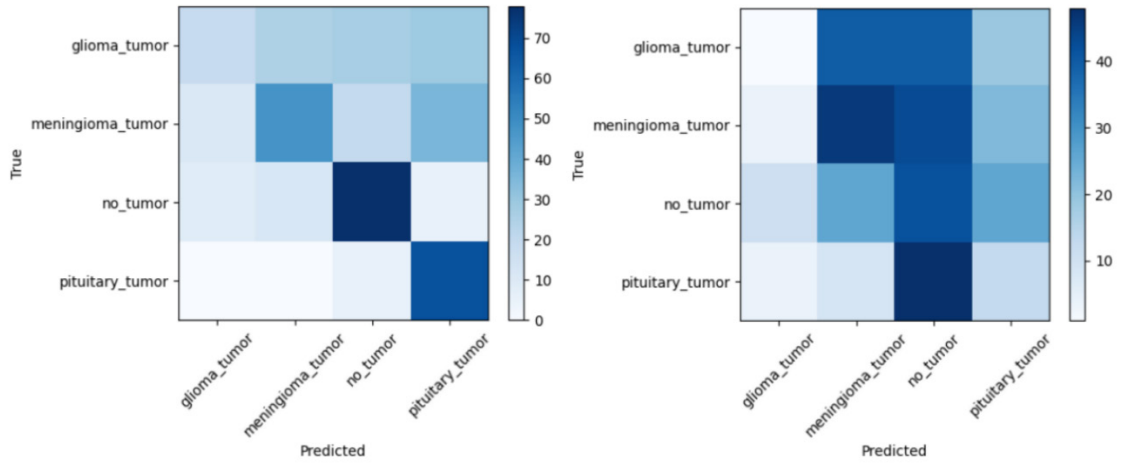

Fig.10 The influence of optimizer type on confusion matrix based on a batch size of 64 with different learning rates (Adam: 0.001, SGD: 0.0001) (Picture credit : Original)

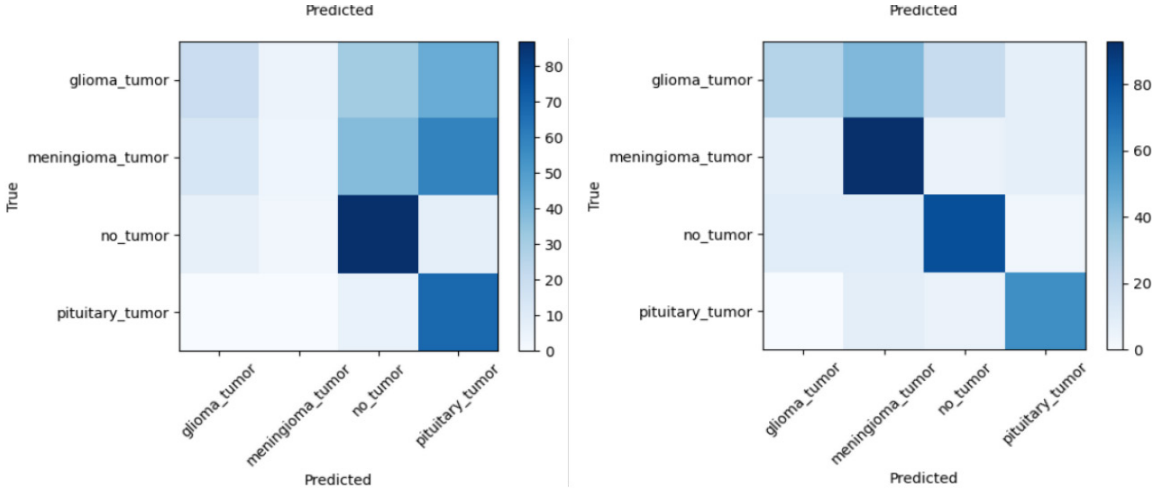

Fig.11 The influence of optimizer type on confusion matrix based on a batch size of 64 with different learning rates (SGD: 0.1, Adam: 0.01) (Picture credit : Original)

4. Conclusion

The performance of the EfficientNetB0 model in classifying brain tumor images from MRI is investigated in this study by examining the influence of learning rate, optimizer, and batch size. The Adam optimizer, when paired with higher learning rates (0.01 and 0.1) and moderate batch sizes (20–60), results in better stability and a significant reduction in misclassification between no tumor and meningioma tumor. In contrast, the SGD optimizer exhibits greater performance fluctuations. Among the tumor types, pituitary tumors are easier to identify due to their distinct features, whereas no tumor and meningioma tumor are more difficult to distinguish due to overlapping characteristics. These findings offer valuable guidance for hyperparameter tuning in AI-assisted brain tumor diagnosis and contribute to improving diagnostic efficiency and accuracy.

Although this study offers helpful information about the impact of hyperparameters on brain tumor classification, it still has some limitations. The first step was to train and test the model on a single public dataset, which may not accurately represent the variability in real clinical settings. Second, only one network architecture (EfficientNetB0) was explored, limiting comparisons across models. In future work, expanding the dataset with more diverse samples and testing across multiple models or medical centers could improve generalizability. Integrating explainable AI would improve clinicians' comprehension of the method used by the model for making

ISSN 2959-6157

decisions, enhancing trust and clinical applicability.

References

[1] Li C, He WQ. Comparison of primary liver cancer mortality estimates from World Health Organization, global burden disease and global cancer observatory. Liver International. 2022 Oct;42(10):2299-316.

[2] Ostrom QT, Cioffi G, Waite K, Kruchko C, Barnholtz-Sloan JS. CBTRUS statistical report: primary brain and other central nervous system tumors diagnosed in the United States in 2014–2018. Neuro-oncology. 2021 Oct 1;23(Supplement_3):iii1-05.

[3] Akkus Z, Galimzianova A, Hoogi A, Rubin DL, Erickson BJ. Deep learning for brain MRI segmentation: state of the art and future directions. Journal of digital imaging. 2017 Aug;30:449-59.

[4] Achiam J, Adler S, Agarwal S, Ahmad L, Akkaya I, Aleman FL, Almeida D, Altenschmidt J, Altman S, Anadkat S, Avila R. Gpt-4 technical report. arXiv preprint arXiv:2303.08774. 2023 Mar 15.

[5] Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A. Highly accurate protein structure prediction with AlphaFold. nature. 2021 Aug;596(7873):583-9.

[6] Rajpurkar P, Irvin J, Zhu K, Yang B, Mehta H, Duan T, Ding D, Bagul A, Langlotz C, Shpanskaya K, Lungren MP. Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. arXiv preprint arXiv:1711.05225. 2017 Nov 14.

[7] Kagiyama N, Tokodi M, Sengupta PP. Machine learning in cardiovascular imaging. Heart Failure Clinics. 2022 Apr 1;18(2):245-58.

[8] Alyoubi WL, Shalash WM, Abulkhair MF. Diabetic retinopathy detection through deep learning techniques: A review. Informatics in Medicine Unlocked. 2020 Jan 1;20:100377.

[9] Zhao L, Jia K. Multiscale CNNs for brain tumor segmentation and diagnosis. Computational and mathematical methods in medicine. 2016;2016(1):8356294.

[10] Afshar P, Mohammadi A, Plataniotis KN. Bayescap: A bayesian approach to brain tumor classification using capsule networks. IEEE Signal Processing Letters. 2020 Oct 29;27:2024-8.

[11] Chen X, Yao L, Zhang Y. Residual attention u-net for automated multi-class segmentation of covid-19 chest ct images. arXiv preprint arXiv:2004.05645. 2020 Apr 12.

[12] Shankar K, Zhang Y, Liu Y, Wu L, Chen CH. Hyperparameter tuning deep learning for diabetic retinopathy fundus image classification. IEEE access. 2020 Jun 26;8:118164-73.

[13] Bhuvaji S. Brain Tumor Classification (MRI). Kaggle. https://www.kaggle.com/datasets/sartajbhuvaji/brain-tumor-classification-mri. 2020.

[14] Tan M, Le Q. Efficientnet: Rethinking model scaling for convolutional neural networks. InInternational conference on machine learning 2019 May 24 (pp. 6105-6114). PMLR.