Analyzing the Impact of Preprocessing on Deep Learning-based Brain Tumor Classification Using Glioblastoma Magnetic Resonance Image

Nada Firdaus [1], Lailil Muflikhah [1,*], Rizal Setya Perdana [1], and Nashi Widodo [2]

[1] Department of Informatics Engineering, Faculty of Computer Science, Brawijaya University, Malang, Indonesia
[2] Department of Biology, Faculty of Mathematics and Natural Sciences, Brawijaya University, Malang, Indonesia
Email: nadafirdaus@student.ub.ac.id (N.F.); lailil@ub.ac.id (L.M.); rizalespe@ub.ac.id (R.S.P.); widodo@ub.ac.id (N.W.)
*Corresponding author

***Abstract*—Glioblastoma is the most aggressive primary brain tumor and demands reliable image-based analysis for accurate interpretation. This study analyzes how different magnetic resonance image preprocessing strategies affect visual representation learning and deep feature discrimination in deep learning-based glioblastoma detection. An EfficientNetB0 convolutional neural network with transfer learning was employed to learn hierarchical image features from brain magnetic resonance images. The dataset consisted of 4465 images spanning 15 tumor classes, including 204 glioblastoma cases. Five preprocessing scenarios were evaluated: baseline (no preprocessing), contrast enhancement (Contrast Limited Adaptive Histogram Equalization (CLAHE)), image denoising, skull stripping, and intensity normalization. The results indicate that intensity normalization yields the most stable and discriminative feature representations, achieving the highest accuracy of 89.70% with balanced glioblastoma sensitivity and precision. Conversely, aggressive preprocessing methods distort image intensity distributions and spatial cues, leading to degraded feature learning and reduced performance. Visual interpretability analysis using Gradient-weighted Class Activation Mapping (Grad-CAM) further demonstrates that intensity normalization enables consistent attention to tumor-relevant regions while preserving critical morphological structures. These findings highlight that simple intensity standardization is more effective than complex enhancement techniques for robust and visually meaningful glioblastoma image analysis.**

***Keywords*—medical image representation, feature learning, image preprocessing, magnetic resonance imaging, glioblastoma detection**

I. Introduction

Glioblastoma Multiforme (GBM) is characterized by rapid proliferation and marked intra-tumoral heterogeneity, with a median patient survival of only 12–15 months. While Magnetic Resonance Imaging (MRI) is the gold standard for diagnosis, manual interpretation is time-consuming and prone to inter-observer variability [1, 2]. These challenges motivate the development of reliable computer-aided diagnosis systems to support clinical decision-making.

In recent years, deep learning, particularly Convolutional Neural Networks (CNNs), has achieved state-of-the-art performance in brain tumor classification and segmentation tasks using MRI data. Numerous studies have demonstrated that CNN-based models can automatically learn hierarchical spatial and textural features, outperforming traditional machine learning approaches in accuracy and robustness [3, 4]. Transfer learning using pre-trained architectures such as ResNet, DenseNet, and EfficientNet has further improved performance by leveraging large-scale visual representations and reducing the need for extensive annotated medical datasets [5, 6]. More recently, hybrid and ensemble models, as well as transformer-based architectures, have been explored to further enhance classification accuracy and generalization across datasets [7, 8].

Despite these advances, a critical yet underexplored aspect of deep learning-based MRI analysis lies in the choice of image preprocessing techniques. MRI data are inherently affected by scanner-dependent intensity variations, noise, bias field inhomogeneity, and acquisition protocol differences. Preprocessing is therefore commonly applied to enhance image quality and reduce variability prior to model training. Existing studies have employed a wide range of preprocessing strategies, including intensity normalization, histogram equalization, denoising, skull stripping, and bias field correction [9, 10]. However, most prior works adopt preprocessing pipelines in an ad hoc manner, often combining multiple techniques without systematically analyzing their individual impact on model performance, particularly for clinically critical classes such as glioblastoma.

Recent state-of-the-art studies primarily focus on architectural innovation and optimization strategies, such as attention mechanisms, metaheuristic optimization, or

Manuscript received January 24, 2026; revised February 13, 2026; accepted March 12, 2026; published July 17, 2026.

multimodal data fusion, while treating preprocessing as a secondary or fixed component of the pipeline [8, 11, 12]. Moreover, several preprocessing techniques that improve visual contrast for human observers may inadvertently distort morphological or textural features that are essential for CNN-based feature learning, especially when transfer learning from natural image datasets is employed. Only a limited number of recent works have highlighted that aggressive preprocessing can degrade model generalization, yet comprehensive and controlled evaluations remain scarce [13, 14]. This lack of systematic analysis represents a significant research gap, particularly for GBM detection, where false negatives may delay life-saving interventions.

Therefore, there is a clear need for a systematic investigation into how different MRI preprocessing techniques influence deep learning-based glioblastoma detection performance under identical experimental conditions. Understanding this relationship is essential not only for improving classification accuracy but also for ensuring model reliability, interpretability, and clinical applicability. Furthermore, Explainable Artificial Intelligence (XAI) methods such as Gradient-weighted Class Activation Mapping (Grad-CAM) have recently gained attention as tools to verify whether models focus on clinically meaningful tumor regions rather than spurious artifacts introduced during preprocessing [15]. Integrating such interpretability analysis into preprocessing evaluation remains an open challenge.

In this context, the present study aims to bridge the identified research gap by providing a systematic and clinically oriented evaluation of commonly used MRI preprocessing techniques for glioblastoma detection using a modern transfer learning architecture. The main contributions of this work are summarized as follows:

(1) A comprehensive and controlled comparison of five MRI preprocessing strategies, including baseline resizing, contrast enhancement, denoising, skull stripping, and intensity normalization, was evaluated under identical training and testing conditions.
(2) A focused analysis on glioblastoma-specific performance using clinically relevant metrics such as sensitivity and precision, in addition to overall multi-class classification performance.
(3) An interpretability study using Grad-CAM to analyze how different preprocessing techniques influence model attention and feature localization within tumor regions.
(4) Practical insights and evidence-based recommendations for designing robust preprocessing pipelines in deep learning-based brain tumor classification, particularly when transfer learning is employed.

By addressing preprocessing as a first-class research component rather than a fixed precondition, this study contributes to the development of more reliable, interpretable, and clinically applicable deep learning systems for glioblastoma detection from MRI images.

II. Literature Review

Recent advances in deep learning have significantly transformed brain tumor detection and classification from MRI. A substantial body of research has investigated the role of preprocessing techniques in enhancing model performance, evolving from simple intensity adjustments to complex, multi-stage pipelines. However, the effectiveness of these preprocessing strategies varies considerably depending on tumor type, dataset characteristics, and model architecture.

Early studies in deep learning-based brain tumor classification primarily relied on basic preprocessing techniques such as intensity normalization and histogram equalization. These approaches aimed to improve global contrast and reduce inter-image intensity variability. For instance, Aulia and Rahmat [16] demonstrated that Contrast Limited Adaptive Histogram Equalization (CLAHE) combined with a Visual Geometry Group (VGG)-16 architecture achieved an accuracy of 90.37% for general brain tumor identification, suggesting that local contrast enhancement can benefit classification tasks. Nevertheless, their work focused on binary or coarse multi-class tumor categorization and did not specifically evaluate Glioblastoma Multiforme (GBM) as a distinct clinical entity. Moreover, the study did not provide a systematic comparison across multiple preprocessing methods, limiting conclusions about the relative effectiveness of CLAHE versus alternative techniques.

As deep learning architectures matured, subsequent studies increasingly emphasized sophisticated networks and elaborate preprocessing pipelines. Jadhav and Sudhagar [17] compared VGG-19, ResNet-50, and GoogleNet architectures for brain tumor classification, reporting a peak accuracy of 97.2% using VGG-19 in conjunction with normalization, segmentation, and extensive data augmentation. While their results underscore the importance of preprocessing, the combined use of multiple techniques makes it difficult to isolate the contribution of individual preprocessing steps to the final performance. Similarly, ZainEldin *et al*. [8] proposed a hybrid deep learning framework based on InceptionResNetV2 optimized using metaheuristic algorithms, achieving 99.98% accuracy. Although this study represents a strong state-of-the-art result, the aggressive augmentation and optimization strategies employed obscure the specific role of preprocessing, and the focus remains largely on architectural and optimization improvements rather than preprocessing analysis.

In contrast to general brain tumor classification, relatively fewer studies have focused explicitly on GBM detection. Contreras *et al*. [18] developed a custom CNN model to detect glioblastoma while analyzing potential demographic bias, achieving an F1-Score of 0.88 on the Erasmus Glioma Database. Their work highlighted the importance of model generalization and investigated the impact of skull stripping by comparing skull-stripped and non-skull-stripped images. However, the analysis was limited to a small subset of preprocessing variations and did not extend to a broader range of commonly used techniques such as contrast enhancement, denoising, or

intensity normalization. Consequently, the relative effectiveness of different preprocessing strategies for GBM detection remains insufficiently understood.

A limited number of recent studies have explicitly examined the impact of individual preprocessing techniques on model performance. Joseph [13] investigated preprocessing strategies for multimodal brain MRI classification in the context of stroke prognosis and concluded that denoising yielded the best performance improvements. While informative, these findings are domain-specific and may not generalize to glioblastoma detection, as GBM exhibits distinct morphological heterogeneity, necrotic regions, and infiltrative growth patterns that rely heavily on subtle textural cues. Preprocessing techniques that are beneficial for stroke or other neurological conditions may therefore degrade performance when applied to GBM classification.

While recent studies have explored various architectures, the systematic impact of individual preprocessing steps remains under-analyzed. Unlike existing works that often combine multiple techniques without isolating their effects, this study treats preprocessing as a first-class research component. Furthermore, recent advancements such as those by Kollem [19] have demonstrated the efficacy of optimized support vector machines for tissue segmentation, yet the specific compatibility of MRI preprocessing with transfer learning architectures like EfficientNet requires deeper investigation.

III. Materials and Methods

A. Dataset

The dataset used in this study was obtained from the publicly available Brain Tumor MRI Images collection hosted on Kaggle [20]. The dataset consists of 4465 2D axial MRI slices. While we acknowledge that 2D slices from public repositories like Kaggle may not fully capture the 3D volumetric complexity of institutional clinical data, they provide a standardized benchmark for evaluating feature discrimination. To ensure statistical robustness despite the limited number of glioblastoma cases (204 total; 31 in the test set), we employed stratified sampling to maintain consistent class distributions across training, validation, and testing subsets. This approach ensures that performance metrics for the minority GBM class are not biased by random selection.

The dataset comprises a total of 4465 axial brain MRI images categorized into 15 classes, including both pathological and normal cases. The class distribution consists of Astrocytoma (579 images), Carcinoma (248), Ependymoma (150), Ganglioglioma (61), Germinoma (90), Glioblastoma (204), Granuloma (78), Medulloblastoma (131), Meningioma (874), Normal (522), Neurocytoma (457), Oligodendroglioma (224), Papilloma (237), Schwannoma (465), and Tuberculoma (145). To ensure fair representation of all tumor types, the dataset was partitioned using stratified sampling into training (70%, 3,125 images), validation (15%, 670 images), and testing (15%, 670 images) subsets. This strategy preserved the original class distribution across all splits, resulting in 31 glioblastoma images in the test set. All MRI images were resized to 224 × 224 pixels to match the input requirements of the EfficientNetB0 architecture.

B. Preprocessing Techniques

The parameters for each technique were selected based on established literature to balance enhancement and artifact control: CLAHE used a clip limit of 3.0 to prevent over-amplification of noise, while Non-Local Means denoising used a filter strength of h = 15 to preserve structural edges. Min-Max Intensity Normalization was chosen over other methods (like Z-Score) because it provides a global linear transformation that standardizes the intensity range without distorting the relative intensity relationships or the morphological feature hierarchies required for ImageNet pre-trained weights.

To systematically investigate the influence of preprocessing on model performance, five preprocessing scenarios were evaluated independently. Each scenario was applied uniformly across training, validation, and test sets, except for data augmentation, which was restricted to the training set.

(1) Baseline: No explicit preprocessing was applied beyond resizing to 224 × 224 pixels. This scenario served as a reference to evaluate the inherent robustness of transfer learning with pre-trained CNNs.
(2) CLAHE: CLAHE was employed to enhance local contrast using a clip limit of 3.0 and a tile grid size of 16 × 16. The method divides the image into small regions and performs histogram equalization adaptively, aiming to improve local visibility while limiting excessive noise amplification.
(3) Denoising: Non-local means denoising was implemented with filter strength h = 15, template window size of 7, and search window size of 21. This approach reduces noise by averaging similar patches across the image while attempting to preserve structural edges.
(4) Skull stripping: Skull stripping was performed using Otsu thresholding followed by morphological operations (3 × 3 kernel, two erosion iterations and three dilation iterations). Connected component labeling was applied, and the largest connected region was selected as the brain mask to remove non-brain tissues.
(5) Intensity normalization: Global min-max normalization was applied to rescale pixel intensities to the range [0, 255] according to Eq. (1):

$$I_{normalized} = \frac{\left(I - I_{min}\right)}{\left(I_{max} - I_{min}\right)} \times 255 \quad (1)$$

where I denotes the original pixel intensity, and I_{min} and I_{max} represent the minimum and maximum intensities within the image.

Fig. 1 presents a visual comparison of the different preprocessing techniques applied to the same glioblastoma MRI sample.

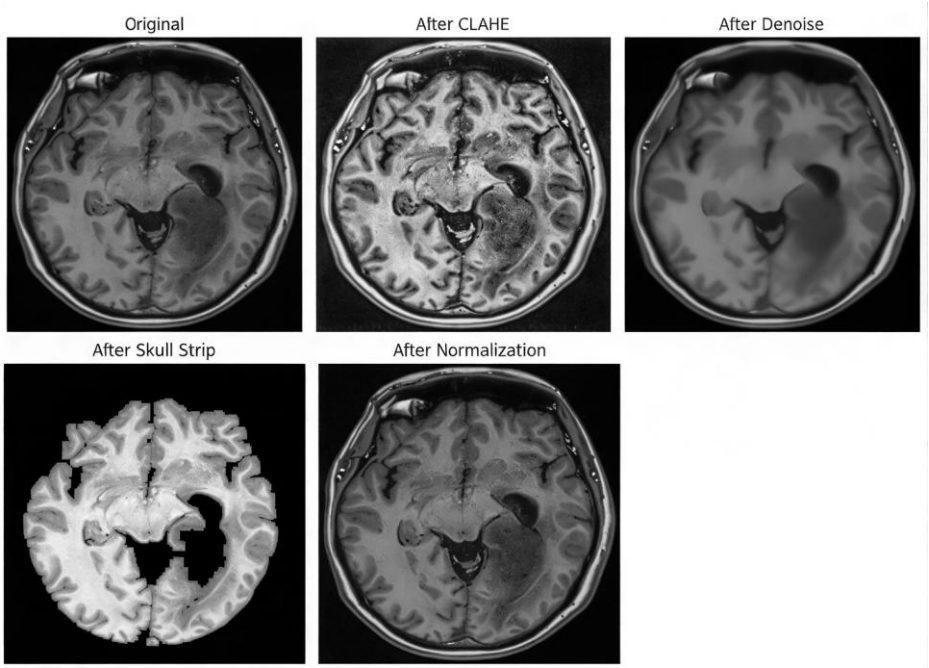

Fig. 1. Visual comparison of preprocessing techniques applied to the same glioblastoma MRI image.

C. Data Augmentation

To improve model generalization and mitigate overfitting, data augmentation was applied exclusively to the training set. The augmentation pipeline included random rotations (±15°), width and height shifts (±10%), horizontal flipping, random zoom (±10%), and brightness adjustments within the range of 80–120%. Augmentation was performed on-the-fly during training, effectively increasing sample diversity without altering validation and test distributions.

D. Model Architecture

The proposed classification model employs EfficientNet-B0 as the backbone CNN for feature extraction. EfficientNet-B0 is designed using a compound scaling strategy that balances network depth, width, and input resolution, enabling efficient feature learning with relatively low computational cost. In this study, the EfficientNet-B0 backbone is initialized with ImageNet pre-trained weights and utilized as a fixed feature extractor.

The convolutional layers of EfficientNet-B0 remain frozen throughout the entire training process, allowing the model to preserve the general visual representations learned from large-scale natural image datasets. Freezing the backbone significantly reduces the number of parameters that must be optimized during training and helps mitigate overfitting, particularly when working with relatively limited medical imaging datasets. Consequently, only the newly added fully connected layers are trained to adapt the extracted features to the target classification task.

The overall architecture, as shown in Fig. 2, consists of the EfficientNet-B0 backbone followed by a custom classification head. After the final convolutional feature maps are generated, a Global Average Pooling (GAP) layer is applied to convert the spatial feature maps into a compact feature vector. This vector is then processed by two fully connected layers consisting of Dense (256 units, Rectified Linear Unit (ReLU)) and Dense (128 units, ReLU), each followed by Dropout layers (0.5 and 0.3) to improve model generalization. Finally, a SoftMax output layer with 15 neurons is used to produce probability predictions for the fifteen tumor classes.

The final architecture contains 4,412,331 parameters, consisting of 362,767 trainable parameters and 4,049,564 non-trainable parameters corresponding to the frozen EfficientNet-B0 backbone as shown in Table I. This configuration enables the model to leverage powerful pre-trained representations while maintaining computational efficiency and improving generalization performance.

TABLE I. MODEL PARAMETERS

Component	Parameters	Status
EfficientNet-B0 backbone	4,049,564	Frozen
Global Average Pooling	0	Trainable
Dense (256 units)	327,936	Trainable
Dense (128 units)	32,896	Trainable
Dense Output (15 units)	1935	Trainable

EfficientNet-B0 is particularly suitable for medical image analysis due to its Mobile Inverted Bottleneck Convolution (MBConv) blocks combined with Squeeze-and-Excitation (SE) attention mechanisms, which allow the network to capture both spatial and channel-wise feature relationships efficiently. In MRI analysis, these capabilities help the model learn discriminative representations of tumor morphology, texture heterogeneity, and intensity patterns, which are essential for accurate multi-class brain tumor classification.

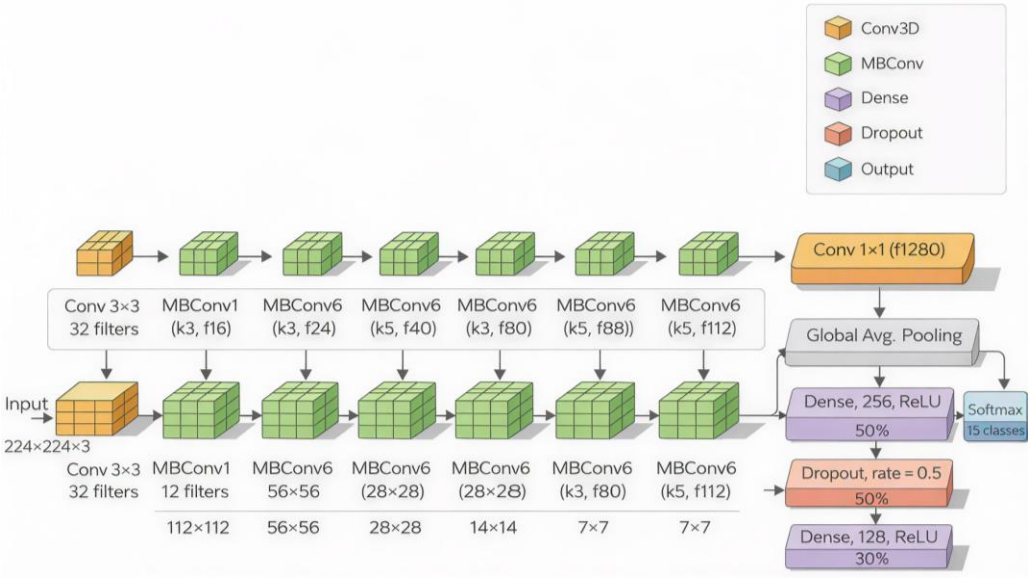

Fig. 2. Block diagram of the proposed EfficientNetB0-based architecture for brain tumor classification.

E. Training Configuration

All models were trained under identical conditions to ensure a fair comparison between preprocessing strategies. The Adam optimizer was used with a learning rate of 1×10^{-4} and categorical cross-entropy loss. Training was conducted with a batch size of 16 for a maximum of 500 epochs. Early stopping was applied by monitoring validation loss with a patience of 10 epochs, restoring the best-performing weights. Additionally, a learning rate reduction on a plateau decreased the learning rate by a factor of 0.5 after five epochs without improvement.

F. Evaluation Metrics

Model performance was assessed using multiple metrics derived from the confusion matrix, including accuracy, precision, recall, and F1-Score as defined in Table II. Accuracy measured overall classification correctness across all 15 classes, while precision, recall, and F1-Score were computed per class and averaged.

TABLE II. CONFUSION MATRIX COMPONENTS

Metric	Definition	Description
True Positive (TP)	Correctly predicted as positive	Model correctly identifies a tumor as belonging to the target class
True-Negative (TN)	Correctly predicted as negative	Model correctly identifies a tumor as not belonging to the target class
False Positive (FP)	Incorrectly predicted as positive	Model incorrectly identifies a tumor as belonging to the target class (Type I error)
False-Negative (FN)	Incorrectly predicted as negative	Model incorrectly identifies a tumor as not belonging to the target class (Type II error)

For clinical relevance, glioblastoma-specific metrics were computed independently. Glioblastoma sensitivity (recall) quantified the proportion of true glioblastoma cases correctly identified, whereas glioblastoma precision measured the reliability of positive glioblastoma predictions. Confusion matrices were generated for each preprocessing scenario to visualize detailed misclassification patterns across tumor types.

Based on these fundamental metrics, overall accuracy is measured by the proportion of correctly classified images across all 15 classes, as shown in Eq. (2), providing a single-number summary of general classification performance.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

Precision computed for each class indicated the reliability of positive predictions, representing the proportion of predicted positive cases that were truly positive, as defined in Eq. (3). Average precision was calculated as the mean precision across all 15 classes.

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

Recall computed for each class measured the model's ability to identify positive cases, representing the proportion of actual positive cases that were correctly detected, as shown in Eq. (4). Average recall was calculated as the mean recall across all 15 classes.

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

F1-Score calculated as the harmonic mean of precision and recall provided a balanced metric that accounts for both false positives and false negatives, as expressed in Eq. (5).

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{5}$$

In addition to accuracy and F1-Score, we included specificity to evaluate the model's ability to correctly identify non-glioblastoma cases, which is critical for minimizing false alarms in clinical settings. Specificity is defined as in Eq. (6).

$$Specificity = \frac{TN}{TN + FP} \tag{6}$$

For clinical relevance, glioblastoma-specific metrics were computed independently of the overall multi-class metrics. Glioblastoma sensitivity, equivalent to recall specifically for the glioblastoma class, measured the proportion of actual glioblastoma cases correctly identified by the model, representing the true positive rate for this critical tumor type. This metric is particularly important clinically because false negatives missing glioblastoma cases, could delay urgent treatment. Glioblastoma precision measured the proportion of images predicted as glioblastoma that were truly glioblastoma, indicating the reliability of positive glioblastoma predictions. High precision is clinically valuable as it reduces unnecessary anxiety and additional testing for patients incorrectly flagged as having glioblastoma. Confusion matrices were generated for each model to visualize the detailed pattern of correct classifications and misclassifications across all class pairs, enabling identification of which tumor types are most confused with each other.

G. Interpretability Analysis

To enhance model transparency, Gradient-weighted Class Activation Mapping (Grad-CAM) was employed [15]. Grad-CAM generates heatmaps by computing gradients of the predicted class with respect to the final convolutional layer, highlighting regions that contribute most strongly to the model's decision. This analysis enabled qualitative comparison of attention patterns across preprocessing methods and ensured that predictions were grounded in medically relevant tumor regions.

IV. Result and Discussion

A. Overall Performance Comparison

The comparative evaluation revealed that preprocessing choices have a profound impact on classification performance. Intensity normalization and the baseline model both achieved a high glioblastoma specificity of 96.77%. While the accuracy improvement of intensity normalization (89.70%) over baseline (88.51%) is 1.19 percentage points, it represents a more stable and clinically reliable feature representation. In contrast, aggressive preprocessing techniques such as CLAHE, denoising, and skull stripping led to substantial performance degradation. Tables III and IV present comprehensive performance metrics across all preprocessing scenarios evaluated on the 670-image test set.

B. Baseline Performance Analysis

The baseline model, which applied only image resizing without additional preprocessing, achieved a strong overall accuracy of 88.51%, demonstrating the effectiveness of transfer learning from large-scale natural image data to the medical imaging domain. In glioblastoma detection, the model correctly identified 26 out of 31 cases, yielding a sensitivity of 83.87% and a precision of 81.00%. The confusion matrix presented in Fig. 3 shows pronounced diagonal dominance, indicating robust classification performance across most tumor categories, with 593 correct predictions out of 670 test images. Misclassified glioblastoma samples were predominantly confused with meningioma, suggesting partial overlap in visual appearance between these tumor types. Specifically, glioblastoma classification resulted in 26 true positives and 5 false negatives, with most errors arising from meningioma-related misclassification. Visual interpretability analysis using Grad-CAM further supports these quantitative findings. As illustrated in Fig. 4, the model consistently attends to diagnostically relevant tumor regions in correctly classified glioblastoma samples. The activation heatmaps, shown as cyan-to-blue gradients, exhibit focused responses over tumor masses with high confidence scores of 99.7% and 99.4%, indicating effective feature localization and the absence of severe overfitting.

TABLE III. Overall Performance Metrics

Scenario	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Baseline	88.51	91.56	83.13	87.14
CLAHE	68.21	71.14	64.46	67.64
Denoise	58.96	63.31	46.03	53.30
Skull Stripping	40.15	37.93	30.37	33.73
Intensity Norm.	89.70	90.84	85.07	87.86

TABLE IV. Glioblastoma Performance Metrics

Scenario	GBM Recall (%)	GBM Precision (%)
Baseline	83.87	81.00
CLAHE	22.58	77.78
Denoise	64.52	44.44
Skull Stripping	48.39	34.88
Intensity Norm.	80.65	83.33

C. Effect of Aggressive Preprocessing Techniques

CLAHE preprocessing led to a substantial performance degradation, with overall accuracy decreasing to 68.21% and glioblastoma sensitivity collapsing to 22.58%. The confusion matrix in Fig. 5 highlights extensive misclassification across tumor categories, with only 457 out of 670 test images correctly classified. Glioblastoma detection was particularly affected, yielding just 7 true positives alongside 24 false negatives. Visual interpretability analysis further explains this decline.

As shown in Fig. 6, Grad-CAM reveals diffuse and unstable attention patterns extending beyond tumor regions. This indicates that aggressive local contrast enhancement distorted global intensity relationships critical for consistent feature extraction. The resulting model uncertainty is reflected in reduced confidence scores of 44.1% and 31.1%.

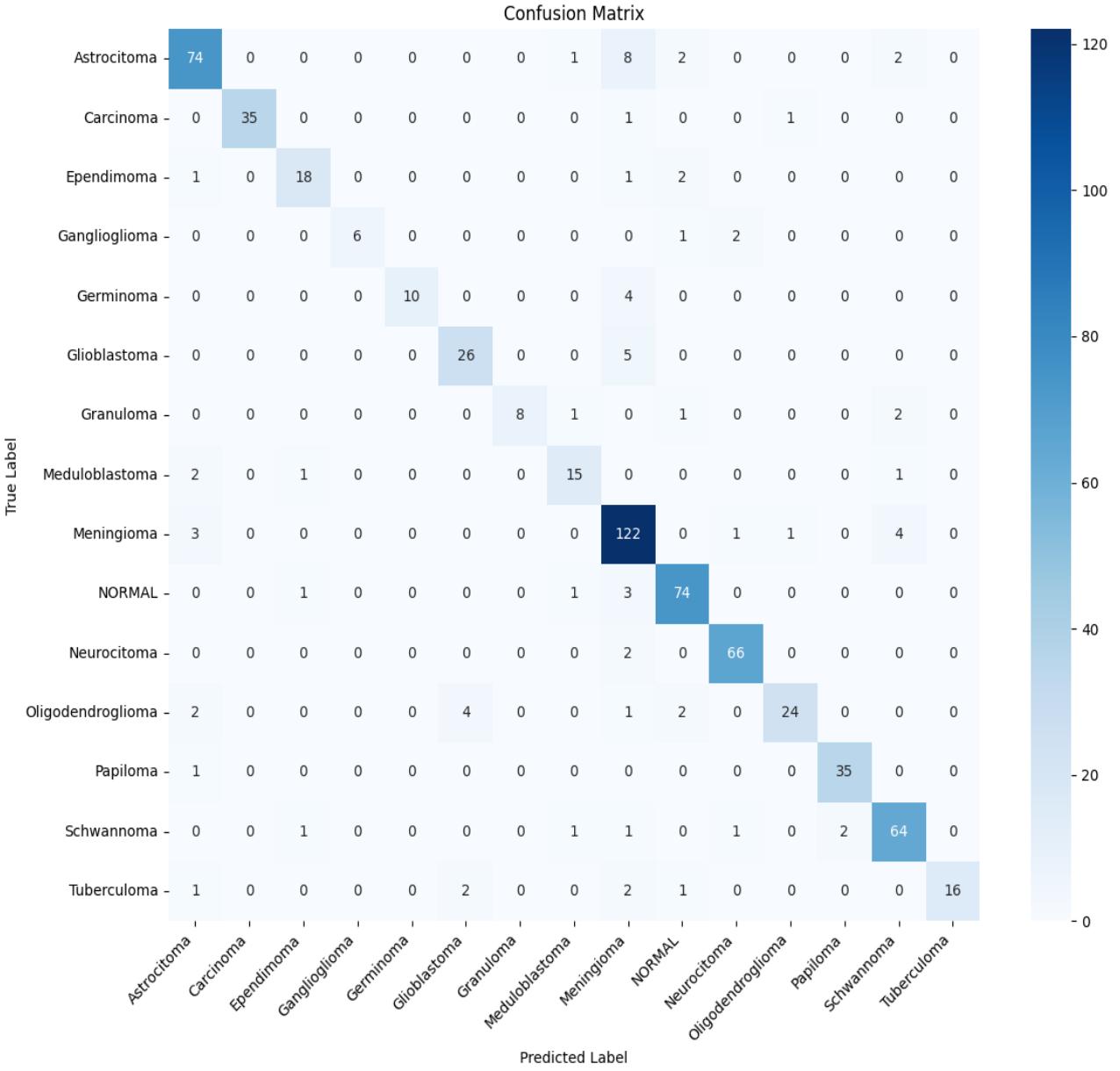

Fig. 3. Confusion matrix for baseline model.

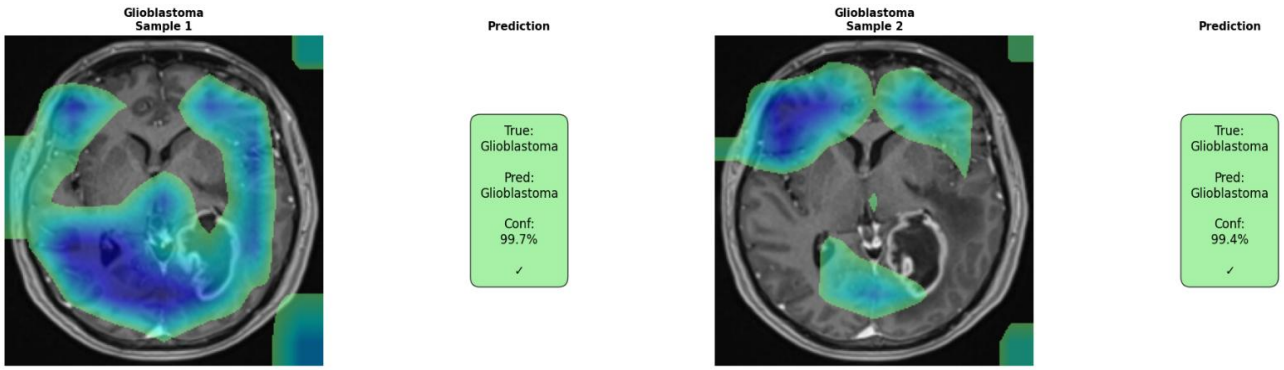

Fig. 4. Grad-CAM visualization for baseline model.

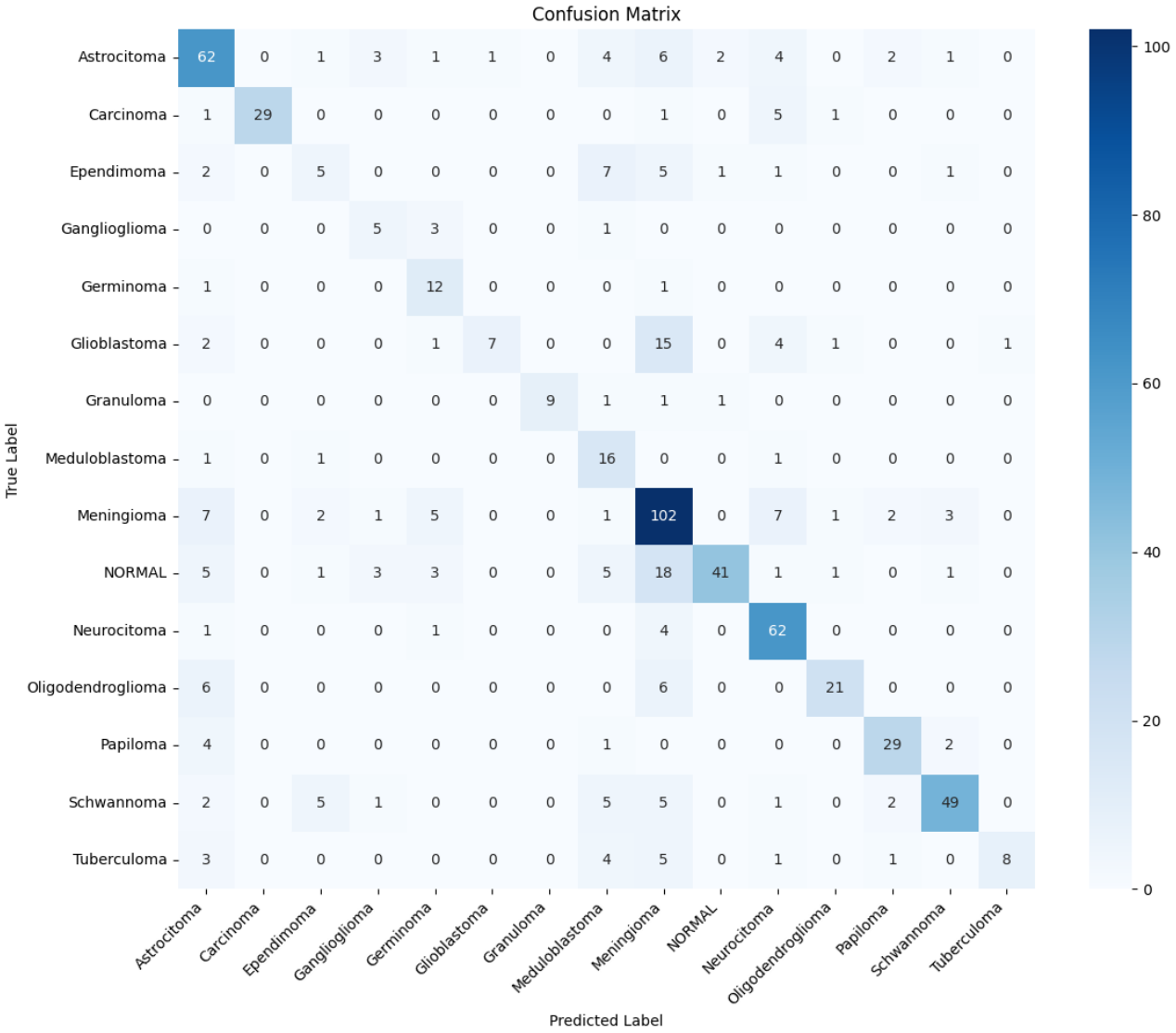

Fig. 5. Confusion matrix for CLAHE preprocessing.

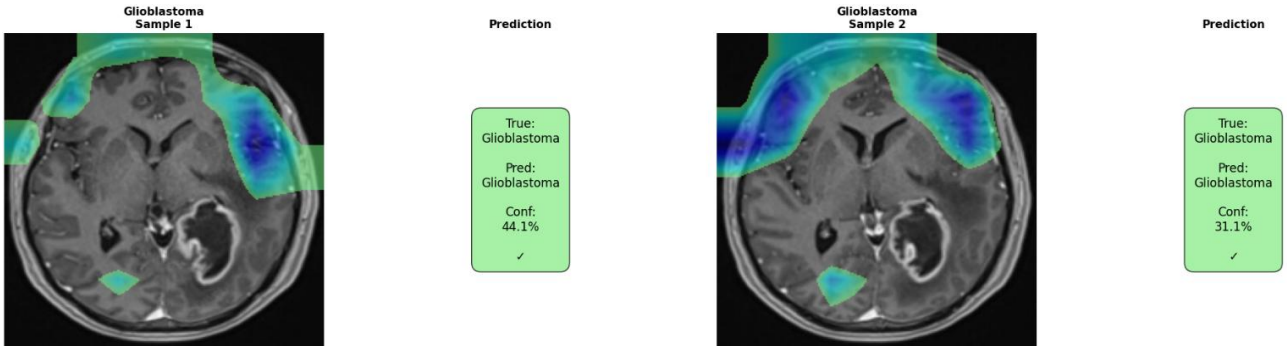

Fig. 6. Grad-CAM visualization for CLAHE preprocessing.

Similarly, Non-Local Means denoising substantially degraded classification performance, resulting in an overall accuracy of 58.96%. While the denoising process effectively reduced random noise, it also suppressed fine-grained textural patterns that are essential for discriminating glioblastoma from other tumor types. The confusion matrix in Fig. 7 reveals pronounced misclassification behavior, characterized by elevated false positive and false negative rates for glioblastoma. Specifically, only 395 samples were correctly classified overall, with glioblastoma detection yielding 20 true positives, 11 false negatives, and 25 false positives incorrectly labeled as glioblastoma. Complementary visual evidence is provided by the activation map analysis in Fig. 8, which shows overly homogeneous activation regions across tumor areas. Despite producing high

confidence scores of 98.2% and 93.2%, the model fails to capture internal heterogeneity within glioblastoma regions. This excessive smoothing effect indicates that denoising removes subtle texture variations that are critical for meaningful visual representation and accurate tumor characterization.

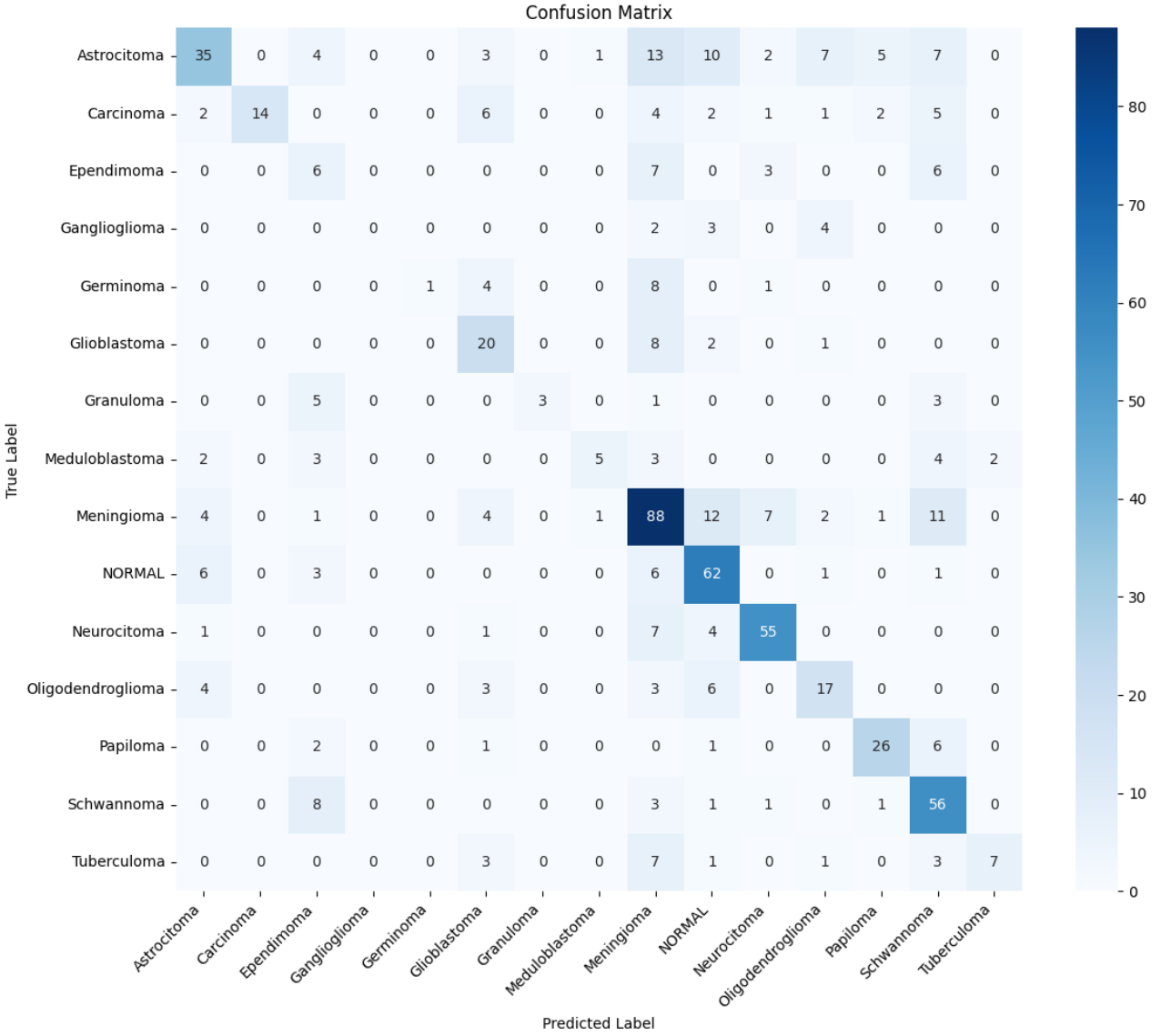

Fig. 7. Confusion matrix for denoising preprocessing.

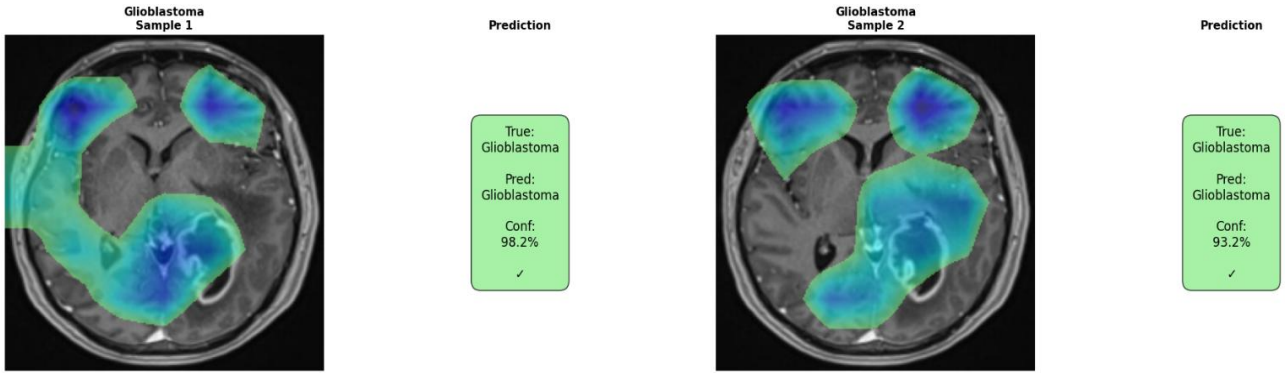

Fig. 8. Grad-CAM visualization for denoising preprocessing.

The failure of aggressive methods like skull stripping (40.15% accuracy) and denoising (58.96% accuracy) is attributed to the loss of anatomical context and fine-grained textures. Glioblastomas are characterized by infiltrative growth and internal heterogeneity (necrotic cores); denoising smooths these subtle textural cues, while

skull stripping removes the spatial context required for the CNN to localize the tumor relative to the brain boundaries. As shown in Fig. 9, the confusion matrix exhibits highly disorganized misclassification patterns, with only 269 correct predictions out of 670 test images. Glioblastoma detection was severely affected, producing 15 true positives, 16 false negatives, and 28 false positives.

The Grad-CAM analysis in Fig. 10 further explains this degradation, revealing model attention concentrated on artificial brain mask boundaries rather than pathological regions. These anomalous activation patterns indicate reliance on spurious edge features instead of tumor-specific characteristics, which is corroborated by reduced confidence scores of 64.3% and 58.0% by skull stripping.

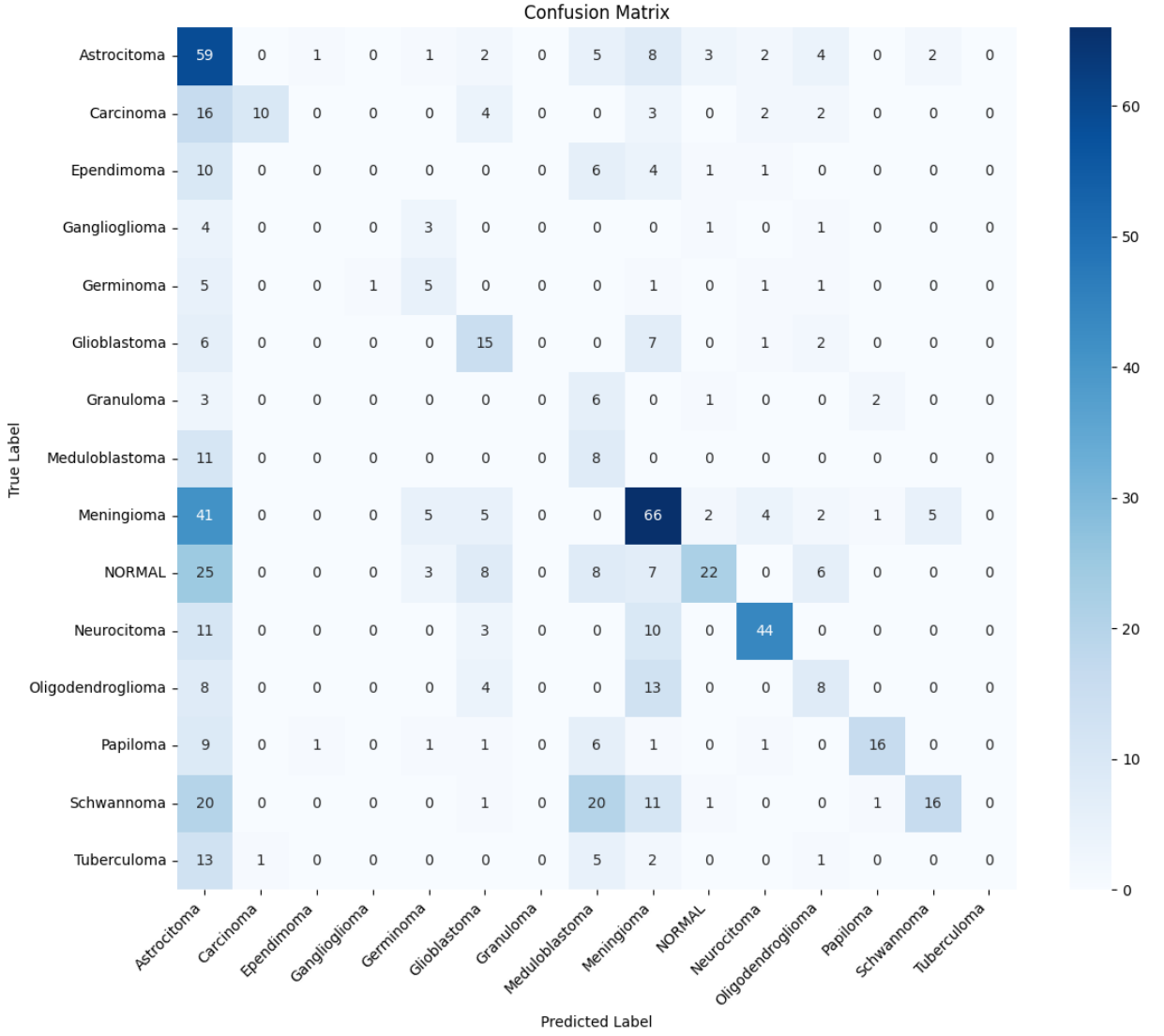

Fig. 9. Confusion matrix for skull stripping preprocessing.

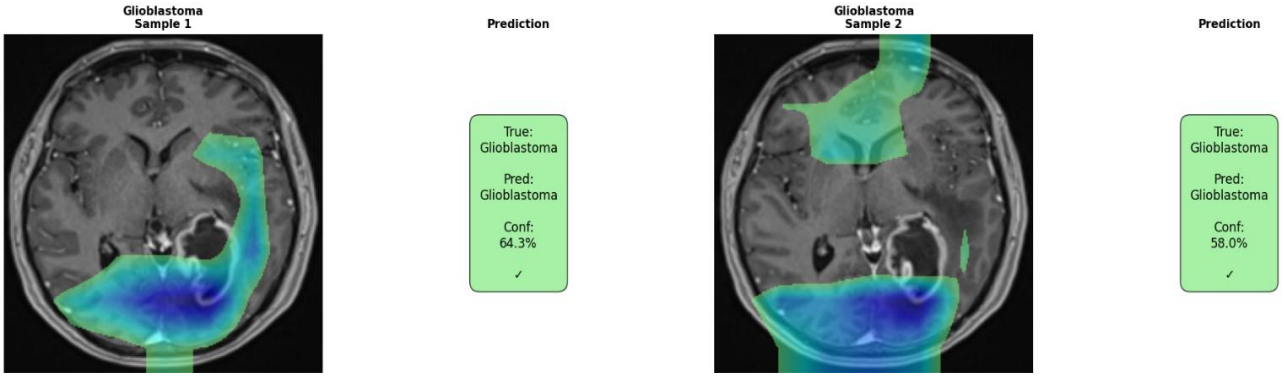

Fig.10. Grad-CAM visualization for skull stripping preprocessing.

D. *Intensity Normalization as an Optimal Strategy*

Intensity normalization consistently outperformed all other preprocessing techniques, achieving an overall accuracy of 89.70%, with average precision, recall, and F1-Score of 90.84%, 85.07%, and 87.86%, respectively. In glioblastoma detection, the model correctly identified 25 out of 31 cases, resulting in a sensitivity of 80.65% and a precision of 83.33%. The confusion matrix shown in Fig. 11 exhibits the strongest diagonal dominance among all evaluated scenarios, indicating reliable and balanced classification across tumor categories. Visual interpretability analysis further supports these findings. As illustrated in Fig. 12, Grad-CAM reveals precise and well-localized attention on tumor regions, capturing heterogeneous internal structures characteristic of glioblastoma. The activation maps display sharply defined tumor boundaries and internal heterogeneity, including solid enhancing regions, necrotic cores, and infiltrative margins, with very high confidence scores of 99.9% and 99.6%.

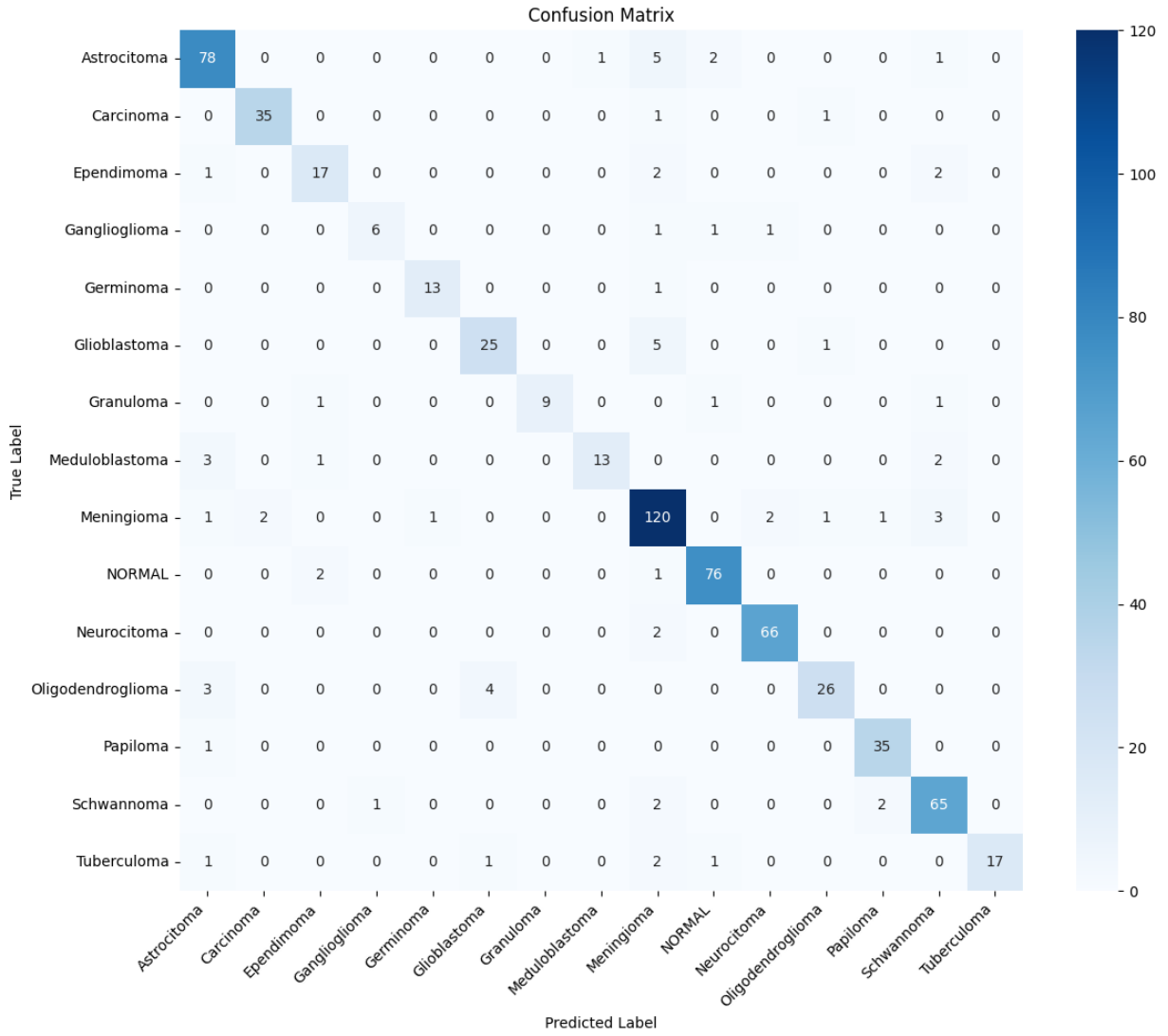

Fig. 11. Confusion matrix for intensity normalization.

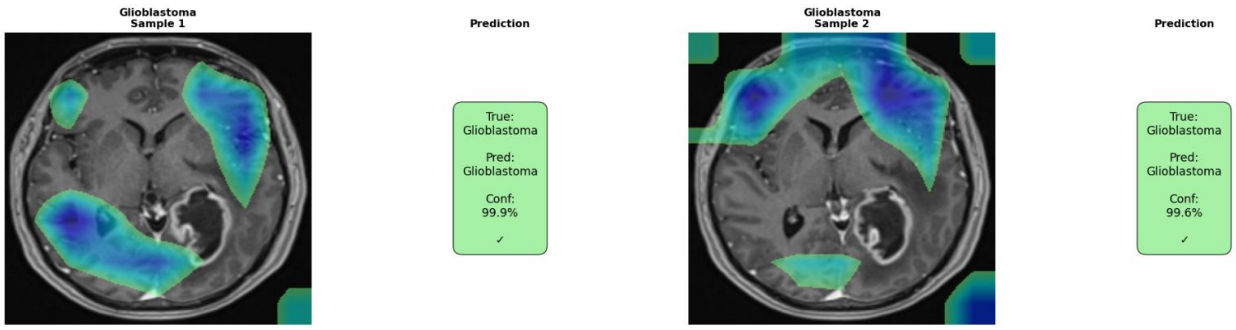

Fig. 12. Grad-CAM visualization for intensity normalization.

The superior performance of intensity normalization can be attributed to its ability to mitigate scanner-dependent intensity variability while preserving morphological and textural features. Unlike CLAHE or denoising, this global linear transformation maintains compatibility with ImageNet pre-trained weights by preserving relative intensity relationships. Intensity normalization yields the most stable and discriminative feature representations. Therefore, Min-Max Intensity Normalization was chosen because it provides a global linear transformation that standardizes the intensity range without distorting the relative intensity relationships or the morphological feature hierarchies required for ImageNet pre-trained weights.

E. Comparative Analysis and Clinical Implications

The comparative results demonstrate that preprocessing strategies designed to enhance human visual perception do not necessarily translate to improved deep learning performance. Aggressive transformations introduced artifacts or removed contextual information, ultimately degrading model reliability.

From a clinical perspective, the intensity normalization-based model shows potential as a decision-support system rather than an autonomous diagnostic tool. While its GBM sensitivity of 80.65% is promising, the remaining false negatives necessitate radiologist oversight. Importantly, the high specificity minimizes unnecessary false alarms, supporting practical clinical deployment.

Although this study focuses on 2D slices for computational efficiency, the findings serve as a foundation for 3D volumetric analysis, where intensity normalization would likely remain the most compatible strategy for deep learning. Future work should focus on model-aware preprocessing and multi-institutional validation for clinical deployment.

V. Conclusion

This study demonstrates that MRI preprocessing is a critical design factor in deep learning. While aggressive techniques often improve visual quality for humans, they can degrade model performance by distorting intensity distributions and suppressing discriminative textures. Global intensity normalization is the most effective strategy, yielding the highest accuracy while maintaining alignment between learned representations and pathological regions. Future work should focus on model-aware preprocessing and multi-institutional validation for clinical deployment.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

N.F. contributed to data curation, methodology, model development, experiments, analysis, and manuscript drafting. L.M. contributed to conceptualization, supervision, methodology, and manuscript review and revision. R.S.P. contributed to technical supervision, deep learning and transfer learning strategies, and validation of experimental results. N.W. contributed to conceptualization, data validation, and manuscript refinement. All authors approved the final version of the manuscript.

Acknowledgment

The authors wish to thank the Directorate General of Research and Development, Ministry of Higher Education, Science, and Technology, Republic of Indonesia, for the support provided during the completion of this research.

Funding

This research was funded by the Directorate General of Research and Development, Ministry of Higher Education, Science, and Technology, under the Fiscal Year 2025 through the Research Grant Program "Regular Fundamental Research", based on the DIPA Grant No. SP DIPA-139.04.1.693320/2025

References

[1] D. N. Louis, A. Perry, P. Wesseling *et al*., "The 2021 who classification of tumors of the central nervous system: A summary," *Neuro Oncol*, vol. 23, no. 8, pp. 1231–1251, 2021. doi: 10.1093/NEUONC/NOAB106

[2] R. Stupp, W. P. Mason, M. J. Van-Den-Bent *et al*., "Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma," *New England Journal of Medicine*, vol. 352, no. 10, pp. 987–996, 2005. doi: 10.1056/NEJMoa043330

[3] M. K. Pehlivanoğlu, A. G. Eker, İ. İnce, D. Karhan, and N. Duru, "Brain tumor classification using deep learning algorithms," in *Proc. 2023 8th International Conference on Computer Science and Engineering (UBMK)*, 2023, pp. 387–392. doi: 10.1109/UBMK59864.2023.10286764

[4] A. C. Yadav and M. H. Kolekar, "Deep feature-based approaches for brain tumor classification and segmentation in medical imaging," *Biomedical Signal Processing and Control*, vol. 117, 109603, 2026. doi: 10.1016/j.bspc.2026.109603

[5] P. Kora, C. P. Ooi, O. Faust *et al*., "Transfer learning techniques for medical image analysis: A review," *Biocybernetics and Biomedical Engineering*, vol. 42, no. 1, pp. 79–107, 2022. doi: 10.1016/j.bbe.2021.11.004

[6] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. the 36th International Conference on Machine Learning*, 2019, pp. 6105–6114. https://proceedings.mlr.press/v97/tan19a.html

[7] A. Hatamizadeh, Y. Tang, V. Nath *et al*., "UNETR: Transformers for 3D medical image segmentation," in *Proc. 2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2022, pp. 1748–1758. doi: 10.1109/WACV51458.2022.00181

[8] H. ZainEldin, S. A. Gamel, E. S. M. El-Kenawy *et al.*, "Brain tumor detection and classification using deep learning and sine-cosine fitness grey wolf optimization," *Bioengineering (Basel)*, vol. 10, no. 1, 18, 2022. doi: 10.3390/bioengineering10010018

[9] Z. Akkus, A. Galimzianova, A. Hoogi, D. L. Rubin, and B. J. Erickson, "Deep learning for brain MRI segmentation: State of the art and future directions," *J. Digit. Imaging*, vol. 30, no. 4, pp. 449–459, 2017. doi: 10.1007/s10278-017-9983-4

[10] N. J. Tustison, B. B. Avants, P. A. Cook *et al*., "N4ITK: Improved N3 bias correction," *IEEE Trans Med Imaging*, vol. 29, no. 6, pp. 1310–1320, 2010. doi: 10.1109/TMI.2010.2046908

[11] W. Jun and Z. Liyuan, "Brain tumor classification based on attention guided deep learning model," *Int. J. Comput. Intell. Syst*., vol. 15, no. 1, 35, 2022. doi: 10.1007/s44196-022-00090-9

[12] M. A. Khan, K. Muhammad, M. Sharif, T. Akram, and S. Kadry, "Intelligent fusion-assisted skin lesion localization and

classification for smart healthcare," *Neural Comput. Appl.*, vol. 36, no. 1, pp. 37–52, 2021. doi: 10.1007/s00521-021-06490-w

[13] A. Joseph, "A novel preprocessing model for multi modal brain MRI image classification for stroke prognosis," *Northeast Journal of Complex Systems (NEJCS)*, vol. 7, no. 1, 4, 2025. doi: 10.22191/nejcs/vol7/iss1/4

[14] J. Lao, Y. Chen, Z. C. Li *et al*., "A deep learning-based radiomics model for prediction of survival in Glioblastoma multiforme," *Sci. Rep.*, vol. 7, no. 1, 10353, 2017. doi: 10.1038/s41598-017-10649-8

[15] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. 2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 618–626. doi: 10.1109/ICCV.2017.74

[16] S. Aulia and D. Rahmat, "Brain tumor identification based on VGG-16 architecture and CLAHE method," *International Journal on Informatics Visualization*, vol. 6, no. 1, pp. 96–102, 2022. doi: 10.30630/joiv.6.1.864

[17] R. N. Jadhav and G. Sudhagar, "Classification of brain tumours in MRI images using VGG 19 algorithm," *Journal of Information Systems Engineering and Management*, vol. 10, no. 13s, pp. 601–609, 2025. doi: 10.52783/jisem.v10i13s.2114

[18] K. Contreras, J. Gutierrez-Rengifo, O. Casanova-Carvajal, A. L. Alvarez, P. E. Vélez-Varela, and A. L. Urbano-Bojorge, "Deep learning for glioblastoma multiforme detection from MRI: A statistical analysis for demographic bias," *Applied Sciences*, vol. 15, no. 11, 6274, 2025. doi: 10.3390/app15116274

[19] S. Kollem, "An efficient method for MRI brain tumor tissue segmentation and classification using an optimized support vector machine," *Multimed. Tools Appl.*, vol. 83, no. 26, pp. 68487–68519, 2024. doi: 10.1007/s11042-024-18233-9

[20] Fernando Feltrin. (2023). Brain Tumor MRI Images 44 Classes. [Online]. Available: https://www.kaggle.com/datasets/fernando2rad/brain-tumor-mri-images-44c