Jurnal Masyarakat Informatika, 17(2), 2026
DOI:10.14710/jmasif.17.2.82619
ISSN: 2777-0648

Enhancing Brain Tumor Classification on Mildly Imbalanced Datasets Using Categorical Focal Cross-Entropy

Muhammad Naufal [*1)], Harun Al Azies [1)], Farrikh Alzami [2)], Novita Kurnia Ningrum [1)], Rivaldo Mersis Brilianto [3)], Pratidina Kusuma Dewi [4)]

[1)] *Informatics Engineering, Faculty of Computer Science, Universitas Dian Nuswantoro, Semarang, Indonesia*
[2)] *Information System, Faculty of Computer Science, Universitas Dian Nuswantoro, Semarang, Indonesia*
[3)] *School of Mechanical Engineering, Pusan National University, 2, Busandaehak-ro 63beon-gil Geumjeong-gu,Busan, Republic of Korea*
[4)] *Anugrah General Hospital, 51145, Pekalongan, Indonesia*
[*] *Corresponding author: m.naufal@dsn.dinus.ac.id*

Abstract

Brain tumor is a dangerous disease that requires accurate diagnosis, but medical datasets often suffer from class imbalance, resulting in bias in deep learning-based classification models. This study proposes using Categorical Focal Cross Entropy (CFCE) in a Convolutional Neural Network (CNN) to address this issue by comparing it with Categorical Cross-Entropy (CCE). CFCE is designed to emphasize minority class samples and hard-to-classify examples, thereby reducing the dominance of the majority class. Experiments were conducted on a brain tumor dataset with class imbalance, where the CNN model with CFCE achieved 84.57% accuracy, 83.57% precision, 85.27% recall, and 83.97% F1-score, outperforming the model with CCE (81.30% accuracy, 81.77% precision, 82.83% recall, and 81.15% F1-score). These results show that Focal Loss effectively improves the classification performance on imbalanced data, with a better ability to detect brain tumors, especially in the minority class. This study contributes to developing a more robust and reliable deep learning-based diagnosis system for medical applications.

Keywords : *Class Imbalance, Convolutional Neural Network, Focal Loss, Brain Tumor Classification*

1 Introduction

Brain tumors represent one of the deadliest types of cancer, characterized by high mortality rates and often poor prognoses, making early and accurate diagnosis crucial for patient survival. According to the Global Cancer Observatory (GCO), in 2022, there were approximately 321,731 new cases of brain tumors worldwide, with a staggering 248,500 deaths attributed to this condition, underscoring the urgent need for improved diagnostic approaches [1]. Recent advances in Artificial Intelligence (AI) and intense learning have revolutionized the medical field by enabling more precise early screening, classification, identification, monitoring, and evaluation of treatment outcomes [2-4].

Among the various deep learning architectures, Convolutional Neural Networks (CNNs) have emerged as a dominant approach for brain tumor classification [5], [6]. CNNs can extract complex hierarchical features from medical image data through convolutional, pooling, and fully connected layers, providing robust and automated diagnostic support [7], [8]. However, the effectiveness of CNNs is highly contingent upon the quality and quantity of training data. Medical imaging data, especially in the context of brain tumor diagnosis, presents several inherent challenges, including acquisition complexity, high inter-patient variability, limited sample sizes, and significant class imbalance [9].

Class imbalance poses a significant challenge in machine learning, particularly when the distribution of classes within a dataset is unequal. One class (e.g., malignant tumors) is often

Received: 21 February 2026; Revised: 21 July 2026; Accepted: 19 August 2026; Published: 9 September 2026

significantly underrepresented compared to another (e.g., benign tumors), potentially leading to biased model predictions. This occurs when certain classes are heavily underrepresented compared to others, leading to biased model performance. The severity of class imbalance can vary across different levels, ranging from mild and moderate to high and even extreme cases, where the disproportion between classes is exceptionally pronounced [10]. This imbalance can cause the learning process to become skewed toward the majority class, reducing the model's sensitivity to minority class samples, which are often the most clinically significant [9], [11]. Various data balancing strategies have been explored in prior research to address this. Traditional approaches, such as Random Undersampling (RUS) and Random Oversampling (ROSE), attempt to correct this imbalance by reducing the number of majority class samples or artificially increasing the number of minority class samples [12], [13]. However, these methods can disrupt the natural variance of the original data, potentially leading to overfitting or loss of critical information [13], [14].

A more sophisticated approach uses synthetic data generation techniques like the Synthetic Minority Over-sampling Technique (SMOTE). SMOTE generates new minority class samples by interpolating between existing samples based on feature proximity, which can help alleviate data imbalance without simply duplicating existing data [15]. However, SMOTE has limitations, particularly when applied to medical images, as it can produce samples with less realistic visual quality, potentially compromising model performance [16].

In contrast, deep learning-based approaches, such as Generative Adversarial Networks (GANs) and their derivatives, offer a more promising solution by generating high-quality synthetic data through adversarial training [17]. GANs consist of two competing neural networks a generator and a discriminator trained simultaneously in a zero-sum game to produce highly realistic synthetic data [18]. However, GANs also have significant drawbacks, including high computational complexity, long training times, and potential instability during the training process [19].

Beyond data-level solutions, algorithm-level strategies have also been proposed, including modifying the loss function used during training. One promising alternative is the Focal Loss Function, initially developed for object detection tasks but increasingly recognized for its effectiveness in handling class imbalance [20], [21]. Unlike the conventional Categorical Cross-Entropy (CCE) loss, which treats all samples equally [22], Focal Loss dynamically adjusts the weight of each sample based on its classification difficulty [21]. This approach reduces the influence of easily classified majority samples, allowing the model to focus more on learning the underrepresented minority class, thus mitigating the impact of class imbalance.

This study proposes integrating Focal Loss with Categorical Cross-Entropy (CCE) as Categorical Focal Cross-Entropy (CFCE) to enhance the performance of CCE within the CNN architecture for brain tumor classification. By comparing the performance of models trained with CCE and CFCE, the aim is to demonstrate the effectiveness of this approach in improving the sensitivity and overall classification accuracy for minority classes. The dataset employed in this study is of mild size, warranting an approach that balances model performance with the validity of outcomes. Given the medical nature of the data specifically, MRI (Magnetic Resonance Imaging) scans a paramount consideration is preserving data integrity and authenticity (data authenticity) to ensure that their representation accurately reflects real-world clinical conditions without excessive manipulation. By prioritizing minimal distortion of the original MRI dataset while effectively mitigating class imbalance, the selected method ensures that the resulting model remains both statistically robust and clinically interpretable.

This contribution is expected to enhance the reliability of CNN-based diagnostic systems in clinical practice, providing a more balanced and robust decision-making framework for medical professionals. Ultimately, this research aims to bridge the gap between deep learning advancements and practical clinical applications, supporting the development of more precise, reliable, and scalable diagnostic tools for brain tumor detection and classification under challenging real-world data conditions.

2 Research Methods

The initial phase of this research involves preparing a publicly available dataset, which serves as the foundation for training and evaluating the proposed deep learning model. This phase includes preprocessing steps to ensure the data is in a standardized format suitable for machine learning, such as resizing images, normalizing pixel values, and organizing data into structured directories. The data splitting process divides the dataset into distinct subsets for training, validation, and testing. This step is essential for achieving accurate and unbiased model evaluation, as it prevents data leakage and ensures that the model's performance is assessed on unseen data. Typically, the data is split into 70% for training, 15% for validation, and 15% for testing, maintaining the original class distribution to address potential class imbalance issues. Below is an our research flow showing in Figure 1.

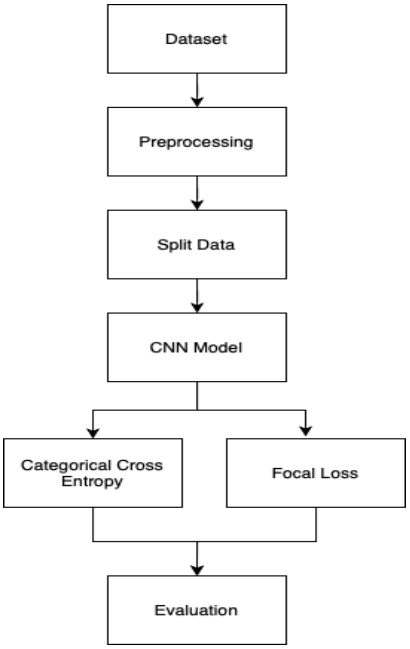


Figure 1 Research Flow

The training phase involves feeding the prepared data into a Convolutional Neural Network (CNN) architecture optimized using Categorical Cross-Entropy (CCE) and Categorical Focal Cross-Entropy (CFCE). This approach comprehensively evaluates the impact of different loss functions on model performance, particularly in scenarios with imbalanced class distributions. Finally, the evaluation phase assesses the model's predictive capabilities using various performance metrics, including accuracy, precision, recall, F1-score, and confusion matrix analysis, providing insights into the model's robustness and generalization ability [23]. As illustrated in Figure 1, this process represents the structured workflow adopted in this research to develop a reliable brain tumor classification model.

3 Results and Discussion

This section presents a comprehensive evaluation of the research results, examining each stage in detail. The analysis begins with preprocessing the image dataset, followed by a description of the CNN model architecture, including the selected layers, activation functions, and filter sizes. Next, the loss function is compared between the baseline Categorical Cross-Entropy (CCE) and the proposed method, Focal Loss, reformulated as Categorical Focal Cross-Entropy (CFCE). Finally, the model's performance is evaluated using a confusion matrix.

3.1 Data Preparation

The dataset used in this study is sourced from Figshare [24], consisting of a total of 3064 T1-CE MRI images taken from 233 patients with three types of tumors: 708 images of meningioma (82 patients), 1426 images of glioma (89 patients), and 930 images of pituitary tumors (62 patients). The dataset exhibits a class imbalance, with the minority-to-majority class ratio ranging from 1:1.53 to 1:2.1, indicating a mild imbalance characteristic [25]. The dataset consists of images with a resolution of 512x512 pixels, and the data is imbalanced. The images are then scaled down to 128x128 pixels. The dataset was randomly divided at the image level while maintaining the class distribution into training (70%), validation (15%), and test (15%) subsets. Although patient identifiers were available in the dataset, they were not used as grouping variables during data splitting in this study. Therefore, images originating from the same patient could potentially be distributed across different subsets. The training set was used to train the CNN models and update their parameters, while the validation set was used to monitor model performance during training. After the training process was completed, the final model was evaluated on the held-out test set to obtain the final classification performance. The test set was not used for model training or validation.

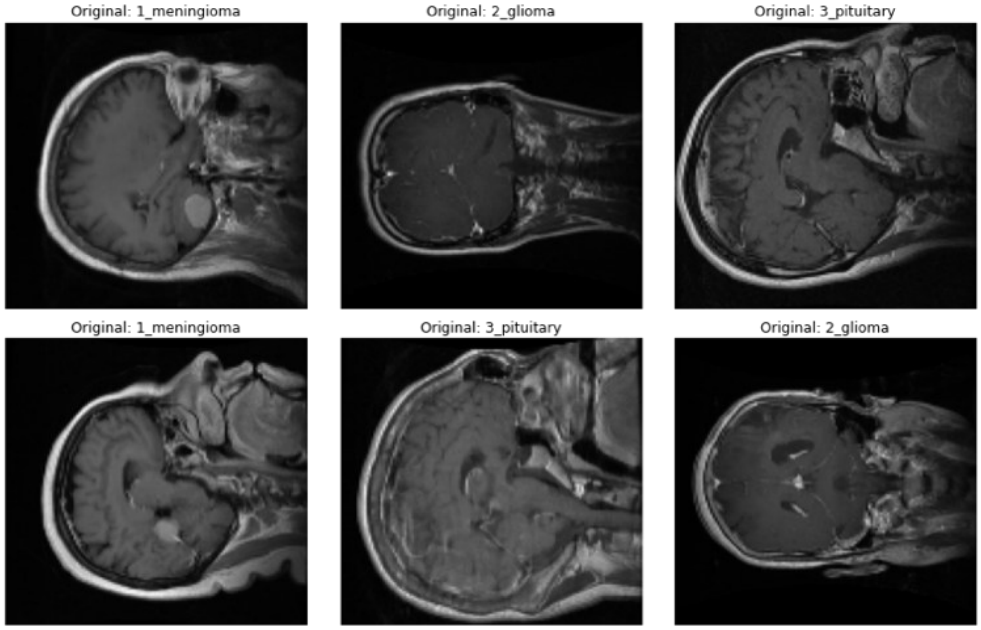


Figure 2 Sample Images from the T1-CE Brain Tumor MRI Dataset

3.2. Convolutional Neural Network Architecture

The dataset is trained using the training and validation sets with the CNN algorithm. The CNN architecture used for this study is built with the following specifications. A vanilla CNN architecture is used due to its low complexity [26], assuming it will perform well with architectures like ImageNet. The architecture consists of 32, 64, and 128 convolutional layers, followed by max-pooling layers to reduce the feature dimensions generated by each convolution. Afterward, a global max-pooling layer is used, followed by a fully connected layer with 64 neurons and a dropout of 0.5 before entering the classification layer. The classification layer uses softmax activation with three output classes. The total number of parameters in this architecture is 101.699. The batch size was set to 32, and the model was trained for 20 epochs. The CNN architecture adopted from the previous study [27] is presented in Table 1.

Table 1 CNN Architecture for Brain Tumor Classification

Layers	Configuration	Output Shape	Activation	Parameters
Conv2D	32 filters, 3 × 3 kernel	126 × 126 × 32	ReLu	896
MaxPooling2D	2 × 2 pool	63 × 63 × 32	-	0
Conv2D	64 filters, 3 × 3 kernel	61 × 61 × 64	ReLu	18.496
MaxPooling2D	2 × 2 pool	30 × 30 × 64	-	0
Conv2D	128 filters, 3 × 3 kernel	28 × 28 × 128	ReLu	73.856
MaxPooling2D	2 × 2 pool	14 × 14 × 128	-	0
GlobalMaxPooling2D	-	128	-	0
Dense	64	64	L2 regularization, ReLu	8.256
Dropout	0.5	64	-	0
Dense	3	3	Softmax	195
Total				101.699

Table 1 presents the architecture of the CNN model used in this study. The model consists of three convolutional blocks followed by GlobalMaxPooling2D and two fully connected layers. The convolutional layers use 32, 64, and 128 filters with a 3 × 3 kernel and ReLU activation. The extracted features are aggregated using GlobalMaxPooling2D and passed to a 64-unit dense layer with ReLU activation and L2 kernel regularization. A dropout layer with a rate of 0.5 is applied before the final three-unit softmax layer. The complete architecture and parameter configuration are presented in Table 1. The model contains 101,699 trainable parameters.

3.3. Categorical Focal Cross Entropy

Categorical Focal Cross Entropy (CFCE) is an extension of Categorical Cross Entropy (CCE) designed to address class imbalance by assigning higher weights to misclassified or uncertain samples through a focusing parameter (γ) [28]. While CCE treats all samples uniformly, CFCE dynamically downweights the contribution of well-classified samples (high-probability predictions) and emphasizes those with higher uncertainty or misclassification, thereby improving model performance on minority classes or imbalanced datasets. This makes CFCE particularly effective in scenarios such as rare object detection or medical classification, where dominant classes may otherwise bias the learning process. The adaptive weighting mechanism enhances focus on challenging samples, leading to more robust generalization in skewed data distributions. The CFCE is defined by the following formula.

$$CE(y, \hat{y}) = \sum_{i=1}^{C} y_i \, log(\hat{y}_i) \quad (1)$$

Where $CE(y, \hat{y})$ represents categorical cross entropy loss**.** y_i is the true label (0 or 1 from each class) derived from one-hot encoding and $\hat{y}_i$ is the predicted probability for each class i with C representing the number of classes. Focal Cross Entropy is defined as [29]:

$$FL(p_t) = \alpha(1 - (p_t))^{\gamma} \times CE(y, \hat{y}) \quad (2)$$

$FL(p_t)$ is the categorical focal cross entropy, α balances the positive and negative values, and the γ parameter adjusts the level of down-weighting for easy samples. When γ = 0, FL becomes equivalent to CE, but increasing γ amplifies the modulation effect, to calculate p_t describe in equation (3) and in this experiment parametter used is α = 0.25 and γ = 2.0. These parameter values were used as the predefined configuration of CFCE and were not determined through an ablation study.

$$p_t = \begin{cases} p & if\ y = 1 \\ 1 - p & otherwise \end{cases} \quad (3)$$

3.4. Adam Optimizer

Adam is an adaptive optimization algorithm with several advantages for training neural network models. It automatically adjusts the learning rate for each parameter, making it suitable for various data and models. Adam is efficient in estimating gradients and squared gradients, effective for fluctuating data, and has adjustable settings. The number of learning rate (lr) used is 0.001, beta1 is 0.9, beta2 is 0.999, and epsilon value is 1e-07. The algorithm combines the first and second moments of the gradients to update parameters. The following equation (4-8) shows Adam's calculation [23], [27]:

$$m1 = beta1 \times m1 + (1 - beta1) \times gradient \quad (4)$$

$$m2 = beta2 \times m2 + (1 - beta2) \times gradient \quad (5)$$

$$m1_{hat} = m1 \div (1 - beta1^t) \quad (6)$$

$$m2_{hat} = m2 \div (1 - beta2^t) \quad (7)$$

$$p = p - lr \times \widehat{m1} \div \left(\sqrt{\widehat{m2} + epsilon}\right) \quad (8)$$

3.5. Modeling and Model Evaluation Results

The model's performance was evaluated by training the CNN on the training dataset and testing it using the validation data. The training results are shown in Figure 3 below.

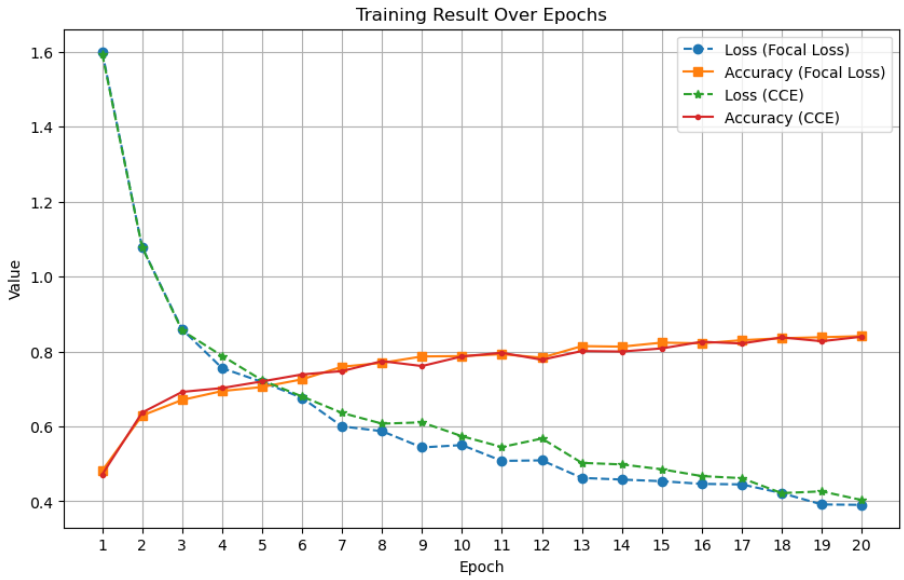


Figure 3 Model Training Results: Loss and Accuracy Comparison

Figure 3 shows that the model with Categorical Focal Cross Entropy (CFCE) achieved a lower loss of 0.390 compared to Categorical Cross Entropy (CCE) at 0.404, indicating that CFCE performs better in minimizing classification errors, especially for complex samples from minority classes. Regarding training accuracy, Categorical Focal Cross Entropy also outperforms CCE, with an accuracy of 84.10% compared to 83.96%, highlighting its effectiveness in addressing class imbalance without compromising overall model performance. The consistent gap between loss and accuracy

further supports the conclusion that CFCE reduces bias towards the majority class while maintaining training stability. The validation results can be seen in Figure 5.

The validation results, shown in Figure 4, confirm that the model with CFCE consistently outperforms the CCE approach in various aspects. The validation loss for CFCE is 0.390, significantly lower than CCE's 0.444, indicating better generalization and handling of overfitting. The validation accuracy is 84.10%, surpassing CCE's 81.30%, which further supports that CFCE is more effective in handling class imbalance during validation. The significant difference in both loss and accuracy suggests that this approach reduces classification errors and improves overall prediction accuracy, especially for more complex cases in the validation set. Next, the model was tested on the test data, and the confusion matrices are shown in Figure 5.

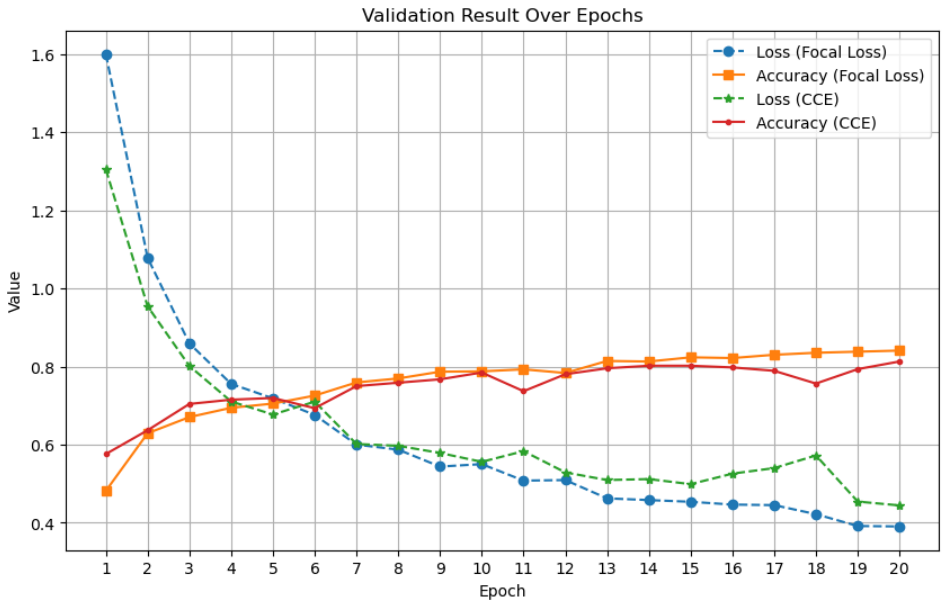
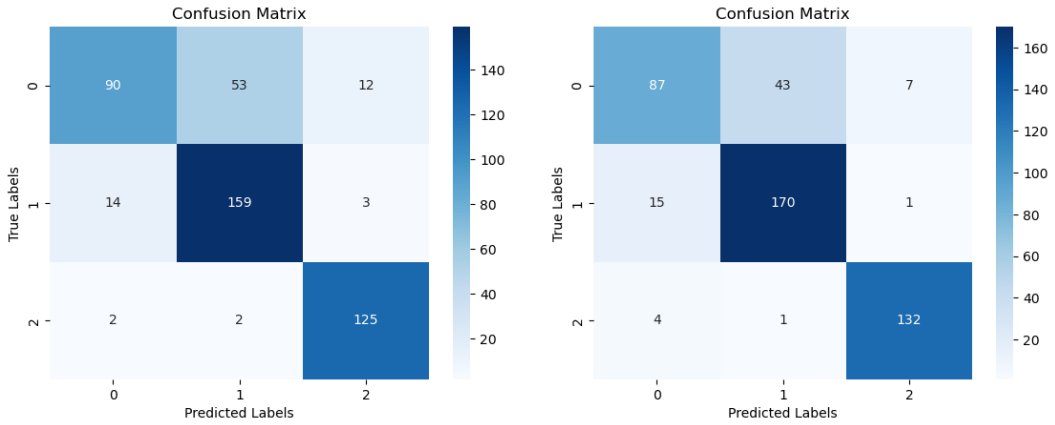


Figure 4. Model Validation Results: Loss and Accuracy Comparison



(a) (b)

Figure 5. Confusion Matrics: (a) CCE, (b) CFCE

Figure 5 presents the confusion matrics (CM) for the model's performance on the test data. The values in the CM are used to evaluate accuracy, precision, recall, and F1 score. Accuracy represents how well the model predicts correctly, both for positive and negative classes. Precision measures how accurate the model is in identifying positive classes, while recall evaluates how well the model identifies all positive classes [30].

Table 2 Model Performance Evaluation on Test Data with Class Imbalance Handling Techniques

Algorithm	Accuracy	Precision	Recall	F1-score
CNN	81.30	81.77	82.83	81.15
CNN - Class Weighted [31]	81.74	82.75	84.35	81.80
CNN - CFCE	**84.57**	**83.57**	**85.27**	**83.97**

Bold values indicate the best performance for each evaluation metric among the compared models

Based on the results in Table 2, the use of Categorical Focal Cross-Entropy (CFCE) in the CNN architecture shows a performance improvement compared to Categorical Cross-Entropy (CCE). Overall, the CNN-CFCE achieved an accuracy of 84.57%, surpassing CNN-CCE's 81.30%. CNN-CFCE also resulted in higher precision (83.57%), recall (85.27%), and F1-score (83.97%) compared to CCE (81.77%, 82.83%, and 81.15%, respectively). These results demonstrate that CFCE is more effective at addressing class imbalance, as it reduces the learning on the majority class and focuses on complex (hard) samples, thus improving classification performance on imbalanced datasets.

In addition, one of the common strategies to handle class imbalance is the use of class weighting, as adopted in [31]. When applied to our dataset (CNN - Class Weighted), this approach achieved a slight performance improvement over the baseline CNN-CCE, with accuracy increasing to 81.74% and gains observed in precision (82.75%), recall (84.35%), and F1-score (81.80%). This indicates that class weighting can help balance the contribution of each class during training, mitigating bias toward the majority class, although in this experiment CFCE yielded a more substantial improvement.

To further examine the interpretability of the CNN model, Score-CAM was applied to visualize the image regions contributing to the model's predictions. Figure 6 presents Score-CAM visualizations for representative images from the meningioma, glioma, and pituitary tumor classes. The original MRI images are shown alongside their corresponding Score-CAM heatmaps, where warmer colors indicate regions with higher activation.

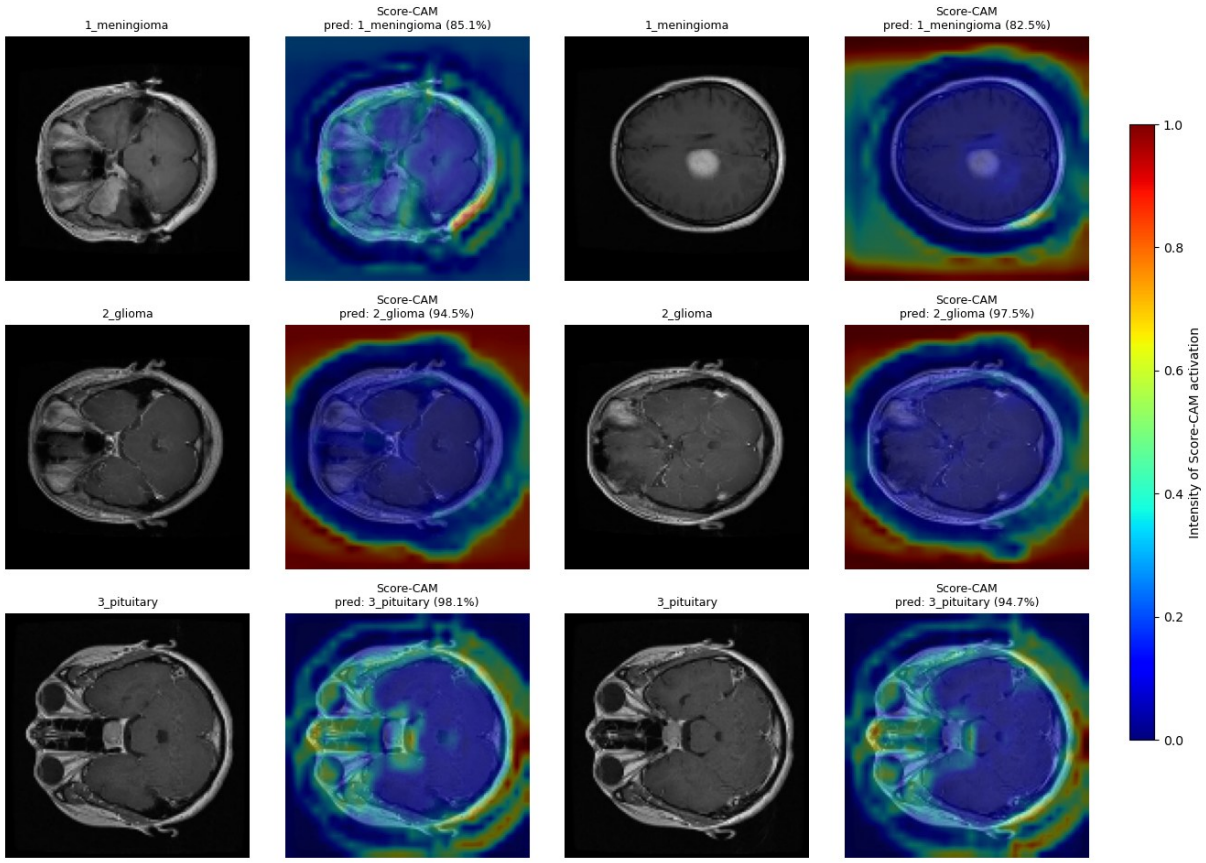


Figure 6. Score-CAM Visualizations of Representative Model

As shown in Figure 6, the Score-CAM visualizations, extracted from the last convolutional layer, indicate that the model correctly classified the representative images with relatively high confidence; however, the highlighted regions are not consistently concentrated on the tumor areas. In several cases, the activation extends to the surrounding brain structures, skull boundaries, and peripheral regions of the image. This suggests that the model may rely on broader image characteristics rather than exclusively focusing on the tumor region when making its predictions. Therefore, the high classification performance should not be interpreted as evidence that the model has learned clinically meaningful tumor localization. These findings highlight an important limitation of the current CNN model in terms of interpretability and indicate the need for further XAI and localization analysis in future studies.

4 Conclusion

The research conducted using a CNN model evaluated the effectiveness of Categorical Focal Cross-Entropy (CFCE) compared with Categorical Cross-Entropy (CCE) for brain tumor classification on a mildly imbalanced dataset. The CNN model with CFCE achieved an accuracy of 84.57%, precision of 83.57%, recall of 85.27%, and F1-score of 83.97%, compared with 81.30%, 81.77%, 82.83%, and 81.15%, respectively, for the CNN model with CCE. These results indicate that the use of CFCE in the evaluated CNN configuration yielded higher classification performance than CCE on the investigated dataset. The success of CFCE can be attributed to its ability to address class imbalance, focus training on complex examples, and reduce bias towards the majority class, ultimately enhancing brain tumor detection. In addition, the implementation of class weighting, as proposed in [31] and applied to this dataset, also resulted in a modest improvement over the baseline CCE, indicating that balancing the contribution of each class during training can partially mitigate the negative effects of imbalance. However, the performance gap between class weighting and CFCE highlights that while both methods address imbalance, CFCE offers a more adaptive and targeted approach, making it particularly effective for medical imaging tasks where minority class detection is critical. For future research, a patient-wise data split is recommended to prevent data leakage between training and testing sets and to ensure a more robust and clinically realistic evaluation of the model's generalizability.

Bibliography

[1] The Global Cancer Observatory (GCO), “Global Cancer Data.” [Online]. Available: https://gco.iarc.fr/en.

[2] B. Bhinder, C. Gilvary, N. S. Madhukar, and O. Elemento, “Artificial Intelligence in Cancer Research and Precision Medicine,” *Cancer Discov.*, vol. 11, no. 4, pp. 900–915, Apr. 2021, doi: 10.1158/2159-8290.CD-21-0090.

[3] D. G. Poalelungi *et al.*, “Advancing Patient Care: How Artificial Intelligence Is Transforming Healthcare,” *J. Pers. Med.*, vol. 13, no. 8, p. 1214, Jul. 2023, doi: 10.3390/jpm13081214.

[4] W. Abbaoui, S. Retal, B. El Bhiri, N. Kharmoum, and S. Ziti, “Towards revolutionizing precision healthcare: A systematic literature review of artificial intelligence methods in precision medicine,” *Inform. Med. Unlocked*, vol. 46, p. 101475, 2024, doi: 10.1016/j.imu.2024.101475.

[5] A. A. Akinyelu, F. Zaccagna, J. T. Grist, M. Castelli, and L. Rundo, “Brain Tumor Diagnosis Using Machine Learning, Convolutional Neural Networks, Capsule Neural Networks and Vision Transformers, Applied to MRI: A Survey,” *J. Imaging*, vol. 8, no. 8, p. 205, Jul. 2022, doi: 10.3390/jimaging8080205.

[6] Y. Xie *et al.*, “Convolutional Neural Network Techniques for Brain Tumor Classification (from 2015 to 2022): Review, Challenges, and Future Perspectives,” *Diagnostics*, vol. 12, no. 8, p. 1850, Jul. 2022, doi: 10.3390/diagnostics12081850.

[7] D. R. Sarvamangala and R. V. Kulkarni, "Convolutional neural networks in medical image understanding: a survey," *Evol. Intell.*, vol. 15, no. 1, pp. 1–22, Mar. 2022, doi: 10.1007/s12065-020-00540-3.
[8] Yogita K. Desai, "Diagnosis of Medical Images Using Convolutional Neural Networks," *J. Electr. Syst.*, vol. 20, no. 6s, pp. 2371–2376, May 2024, doi: 10.52783/jes.3220.
[9] E. Tasci, Y. Zhuge, K. Camphausen, and A. V. Krauze, "Bias and Class Imbalance in Oncologic Data—Towards Inclusive and Transferrable AI in Large Scale Oncology Data Sets," *Cancers*, vol. 14, no. 12, p. 2897, Jun. 2022, doi: 10.3390/cancers14122897.
[10] P. Kaur and A. Gosain, "Empirical Assessment of Ensemble based Approaches to Classify Imbalanced Data in Binary Classification," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 3, 2019, doi: 10.14569/IJACSA.2019.0100307.
[11] R. Walsh and M. Tardy, "A Comparison of Techniques for Class Imbalance in Deep Learning Classification of Breast Cancer," *Diagnostics*, vol. 13, no. 1, p. 67, Dec. 2022, doi: 10.3390/diagnostics13010067.
[12] M. Khushi *et al.*, "A Comparative Performance Analysis of Data Resampling Methods on Imbalance Medical Data," *IEEE Access*, vol. 9, pp. 109960–109975, 2021, doi: 10.1109/ACCESS.2021.3102399.
[13] S. Bagui and K. Li, "Resampling imbalanced data for network intrusion detection datasets," *J. Big Data*, vol. 8, no. 1, p. 6, Dec. 2021, doi: 10.1186/s40537-020-00390-x.
[14] G. Varotto, G. Susi, L. Tassi, F. Gozzo, S. Franceschetti, and F. Panzica, "Comparison of Resampling Techniques for Imbalanced Datasets in Machine Learning: Application to Epileptogenic Zone Localization From Interictal Intracranial EEG Recordings in Patients With Focal Epilepsy," *Front. Neuroinformatics*, vol. 15, p. 715421, Nov. 2021, doi: 10.3389/fninf.2021.715421.
[15] Muljono, S. A. Wulandari, H. A. Azies, M. Naufal, W. A. Prasetyanto, and F. A. Zahra, "Breaking Boundaries in Diagnosis: Non-Invasive Anemia Detection Empowered by AI," *IEEE Access*, vol. 12, pp. 9292–9307, 2024, doi: 10.1109/ACCESS.2024.3353788.
[16] D. Dablain, B. Krawczyk, and N. V. Chawla, "DeepSMOTE: Fusing Deep Learning and SMOTE for Imbalanced Data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 9, pp. 6390–6404, Sep. 2023, doi: 10.1109/TNNLS.2021.3136503.
[17] Y. Chen *et al.*, "Generative Adversarial Networks in Medical Image augmentation: A review," *Comput. Biol. Med.*, vol. 144, p. 105382, May 2022, doi: 10.1016/j.compbiomed.2022.105382.
[18] A. Kebaili, J. Lapuyade-Lahorgue, and S. Ruan, "Deep Learning Approaches for Data Augmentation in Medical Imaging: A Review," *J. Imaging*, vol. 9, no. 4, p. 81, Apr. 2023, doi: 10.3390/jimaging9040081.
[19] "Survey on Generative Adversarial Behavior in Artificial Neural Tasks," *Iraqi J. Comput. Sci. Math.*, pp. 83–94, Mar. 2022, doi: 10.52866/ijcsm.2022.02.01.009.
[20] M. Mulyanto, M. Faisal, S. W. Prakosa, and J.-S. Leu, "Effectiveness of Focal Loss for Minority Classification in Network Intrusion Detection Systems," *Symmetry*, vol. 13, no. 1, p. 4, Dec. 2020, doi: 10.3390/sym13010004.
[21] K. Pasupa, S. Vatathanavaro, and S. Tungjitnob, "Convolutional neural networks based focal loss for class imbalance problem: a case study of canine red blood cells morphology classification," *J. Ambient Intell. Humaniz. Comput.*, vol. 14, no. 11, pp. 15259–15275, Nov. 2023, doi: 10.1007/s12652-020-01773-x.
[22] Y. Ho and S. Wookey, "The Real-World-Weight Cross-Entropy Loss Function: Modeling the Costs of Mislabeling," *IEEE Access*, vol. 8, pp. 4806–4813, 2020, doi: 10.1109/ACCESS.2019.2962617.
[23] F. Alzami, M. Naufal, H. A. Azies, S. Winarno, and M. A. Soeleman, "Time Distributed MobileNetV2 with Auto-CLAHE for Eye Region Drowsiness Detection in Low Light

Conditions," *Int. J. Adv. Comput. Sci. Appl.*, vol. 15, no. 11, 2024, doi: 10.14569/IJACSA.2024.0151146.
[24] J. Cheng, "brain tumor dataset." figshare, p. 879509079 Bytes, 2017. doi: 10.6084/M9.FIGSHARE.1512427.V5.
[25] "Hybrid Chicken Swarm Optimization (CSO) and Fuzzy Logic (FL) Model for Handling Imbalanced Datasets," *Int. J. Intell. Eng. Syst.*, vol. 14, no. 6, pp. 10–19, Dec. 2021, doi: 10.22266/ijies2021.1231.02.
[26] L. Tuggener, J. Schmidhuber, and T. Stadelmann, "Is it enough to optimize CNN architectures on ImageNet?," *Front. Comput. Sci.*, vol. 4, p. 1041703, Nov. 2022, doi: 10.3389/fcomp.2022.1041703.
[27] M. Naufal, H. Al Azies, and R. M. Brilianto, "Enhanced Brain Tumor Classification through Gamma Correction in Deep Learning," *SISTEMASI*, vol. 13, no. 6, p. 2348, Nov. 2024, doi: 10.32520/stmsi.v13i6.4474.
[28] L. Wang, C. Wang, Z. Sun, S. Cheng, and L. Guo, "Class Balanced Loss for Image Classification," *IEEE Access*, vol. 8, pp. 81142–81153, 2020, doi: 10.1109/ACCESS.2020.2991237.
[29] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal Loss for Dense Object Detection," Feb. 07, 2018, *arXiv*: arXiv:1708.02002. doi: 10.48550/arXiv.1708.02002.
[30] M. Muljono, P. N. Andono, S. A. Wulandari, H. A. Azies, and M. Naufal, "Tempo Recognition of Kendhang Instruments Using Hybrid Feature Extraction,", *Journal of Applied Science and Engineering*, vol. 27, no. 3, Mar. 2024, doi: 10.6180/jase.202403_27(3).0004.
[31] M. N. Razali, N. Arbaiy, P.-C. Lin, and S. Ismail, "Optimizing Multiclass Classification Using Convolutional Neural Networks with Class Weights and Early Stopping for Imbalanced Datasets," *Electronics*, vol. 14, no. 4, p. 705, Feb. 2025, doi: 10.3390/electronics14040705.