Artificial Intelligence Systems and Its Applications (AISA)
Volume 1, Issue 2, December 2025
http://journals.cognispectra.com/index.php/aisa/index
E-ISSN: 3110-0457

Multi-Objective Bio-Inspired Hyperparameter Optimization for Trustworthy Brain Tumor MRI Classification Using Calibration-Aware CNN Models

Kafitra Marna Ibrahim[1], Zaky Zaujan Jayaputra[2]

[1] Universitas Muhammadiyah Semarang, Semarang, Indonesia

[2] Universitas Islam Negeri Imam Bonjol Padang, Padang, Indonesia

Abstract: Automated brain tumor classification from magnetic resonance imaging (MRI) has become an essential component in advancing computer-aided diagnosis. However, many deep learning approaches prioritize accuracy alone while overlooking two key requirements for real-world medical deployment: the reliability of predicted confidence scores and the computational efficiency required for clinical integration. This study proposes a multi-objective bio-inspired hyperparameter optimization framework to produce convolutional neural network (CNN) models that are accurate, well-calibrated, and computationally efficient. The model is optimized using a Multi-Objective Particle Swarm Optimization (MOPSO) algorithm that jointly minimizes validation error, Expected Calibration Error (ECE), and inference latency. Experiments were conducted on a four-class Brain Tumor MRI dataset, and the optimized configuration achieved a test accuracy of 95 percent, an ECE of 1.48 percent, and a sub-millisecond inference latency of 0.88 milliseconds per sample. Grad-CAM visualizations further confirm that the model's decisions are guided by clinically relevant tumor regions. The results demonstrate that multi-objective hyperparameter optimization offers a robust pathway for developing trustworthy, efficient, and interpretable artificial intelligence systems for medical imaging applications.

Keywords: Brain Tumor Classification, Convolutional Neural Network, Multi-Objective Optimization, Particle Swarm Optimization, Calibration Error

Article History:

Received: 1 December 2025 Accepted: 27 December 2025 Published: 31 December 2025

Corresponding Author: Kafitra Marna Ibrahim, Email: kafitra.marna@gmail.com

DOI: 10.65917/aisa.v1i2.45

1 Introduction

Among neurological disorders, brain tumors are among the most life-threatening, and their early diagnosis is critical to improve treatment outcomes. Because of its excellent spatial and tissue-contrast resolution, MRI remains the most commonly used non-invasive modality for the identification of tumor type, extent, and progression [1]. More recently, deep learning has achieved impressive success in automating the classification of brain tumors, thus having the potential to alleviate the workload of radiologists while improving diagnostic consistency [2, 3]. Despite this progress, several challenges remain, which severely limit the clinical adoption of deep learning models in medical imaging.

Most of the existing works focus on maximizing classification accuracy while usually ignoring the requirements necessary for real deployment, such as model calibration, reliability of confidence estimates, and computational efficiency. Poor calibration can lead to overconfident yet incorrect predictions, which is not acceptable in safety-critical settings such as tumor diagnosis [4]. Furthermore, many state-of-the-art models are built using highly complex architectures with very costly computation and therefore could not be used in any time-sensitive clinical workflows or resource-limited devices [5]. The above-mentioned limits emphasize the necessity for optimization in a holistic manner where predictive performance needs to be balanced together with reliability and inference efficiency.

Bio-inspired optimization methods, particularly swarm intelligence approaches such as Particle Swarm Optimization (PSO), have shown strong capability in navigating high-dimensional, non-convex search spaces and are increasingly used in neural architecture and hyperparameter optimization [6-8]. However, prior works often optimize a single objective typically accuracy while ignoring competing goals such as calibration and latency. Such single-objective formulations cannot fully address the multi-faceted requirements of trustworthy medical artificial intelligence.

This study proposes a multi-objective bio-inspired hyperparameter optimization framework to improve the performance and trustworthiness of convolutional neural networks for the classification of brain tumors in MRI images. The proposed approach jointly minimizes a calibration-aware objective structure involving classification error, expected calibration error, and inference latency to ensure both the accuracy and trustworthiness of the model. A customized Multi-Objective Particle Swarm Optimization mechanism is proposed to navigate the search space effectively and determine Pareto-optimal hyperparameter configurations. The compactness of the underlying CNN architecture allows structural flexibility and computational efficiency while maintaining its discriminative capability. The performance of the proposed framework is evaluated on a comprehensive set of experiments involving accuracy assessment, calibration analysis, latency measurement, and Grad-CAM interpretability. Experimental results evidence

Artificial Intelligence Systems and Its Applications (AISA)
Volume 1, Issue 2, December 2025
http://journals.cognispectra.com/index.php/aisa/index
E-ISSN: 3110-0457

that the optimized model yields strong diagnostic performance with improved probability calibration and fast inference, which positions the model effectively for practical deployment in medical imaging. Overall, this work contributes to a robust pathway for developing dependable, clinically applicable artificial intelligence systems.

2 Related Work

Brain tumor classification from MRI has been widely explored across various methodological families, including classical machine learning, deep learning architectures, and bio-inspired optimization methods. Early approaches relied on handcrafted features and conventional classifiers, reaching reasonable accuracy but lacking scalability and robustness. More recent works employ CNN-based models or rely on transfer learning in order to improve feature representation, but most stress accuracy in isolation without considering calibration or computational efficiency. Swarm-based and evolutionary optimization methods have also been used for segmentation or the purpose of feature selection, but rarely in order to jointly optimize accuracy, reliability, and latency. Critical review of the above methods reveals that no prior studies incorporate multi-objective optimization by using Particle Swarm Optimization, especially for CNN-based multi-class brain tumor classification, nor include calibration error as an explicit target of optimization. Table 1 summarizes key representative studies and contrasts them against the capabilities required for trustworthy medical AI.

Table 1: Overview of previous research regarding the prediction of Brain Tumor

Author	No. of Instances (Dataset)	Feature Selection	Parameter Tuning / Optimization	Models	Performance Metrics	Result
[9]	650 samples (DICOM)	GLCM (Texture Features)	-	PNN	Accuracy	95%
[10]	4600 images (Kaggle)	Automatic (CNN)	Adam Optimizer	MobileNetV2	Accuracy	88.77%
[11]	5,712 images (Augmented to 142,800)	Automatic (Deep CNN)	Data Augmentation	Deep CNN	Accuracy	91.5%
[12]	3,064 MRI scans	Automatic (YOLOv5 Backbone)	Batch: 64 LR: 0.00261	Standard YOLOv5	Precision	81.9%
					Recall	83%
					mAP	87%
		Automatic (YOLOv5 + NLNNs)	Batch: 64 LR: 0.00261	Improved YOLOv5	Precision	83.5%
					Recall	86%
					mAP	85.2%
[13]	6,000 MRI scans	Transfer Learning	Adam optimizer LR: 0.001	TTL Model	Accuracy	94.5%

			Batch: 32			
[14]	3,064 MRI scans	Ensemble Features	Epochs: 100 Batch: 20	Inception-v3 Ensemble	Accuracy	94.34%
[15]	2,870 MRI scans	GLCM + HOG + LBP	-	Fine KNN	Accuracy	91.1%
[16]	7,023 MRI scans	Automatic (2D CNN)	Adam optimizer LR: 0.001 Batch: 32	Three Layered CNN	Accuracy	89.79%
[17]	-	Automatic (CNN)	Adam optimizer Dropout: 0.5	LeNet Inspired Model	Accuracy	88%
[18]	3,064 MRI scans	Fisher Vector (FV)	Vocab size: 128	Fisher Vector	mAP	94.68%
[19]	3,264 MRI scans	Automatic (Custom)	Adam Optimizer Batch: 18 Epochs: 80	CNN	Accuracy	93.30%
[20]	7000 images, Tumor/No Tumor	Transfer Learning	Adam Optimizer, Dropout	VGG-16	Accuracy	94.98%
	Multiclass: Glioma, Meningioma, Pituitary	Transfer Learning	Adam Optimizer, Dropout	VGG-16	Accuracy	89.19%

3 Methods

This study proposes a bio-inspired hyperparameter optimization framework based on Multi-Objective Particle Swarm Optimization (MOPSO) to develop an efficient and trustworthy Convolutional Neural Network (CNN) for brain tumor classification. The framework is designed to simultaneously balance three conflicting objectives: maximizing classification accuracy, improving probability calibration to ensure reliable confidence estimates, and minimizing inference latency to support real-time clinical deployment. The overall workflow of the proposed method, including data preprocessing, hyperparameter optimization, model construction, and evaluation, is summarized in Figure 1.

Artificial Intelligence Systems and Its Applications (AISA)
Volume 1, Issue 2, December 2025
http://journals.cognispectra.com/index.php/aisa/index
E-ISSN: 3110-0457

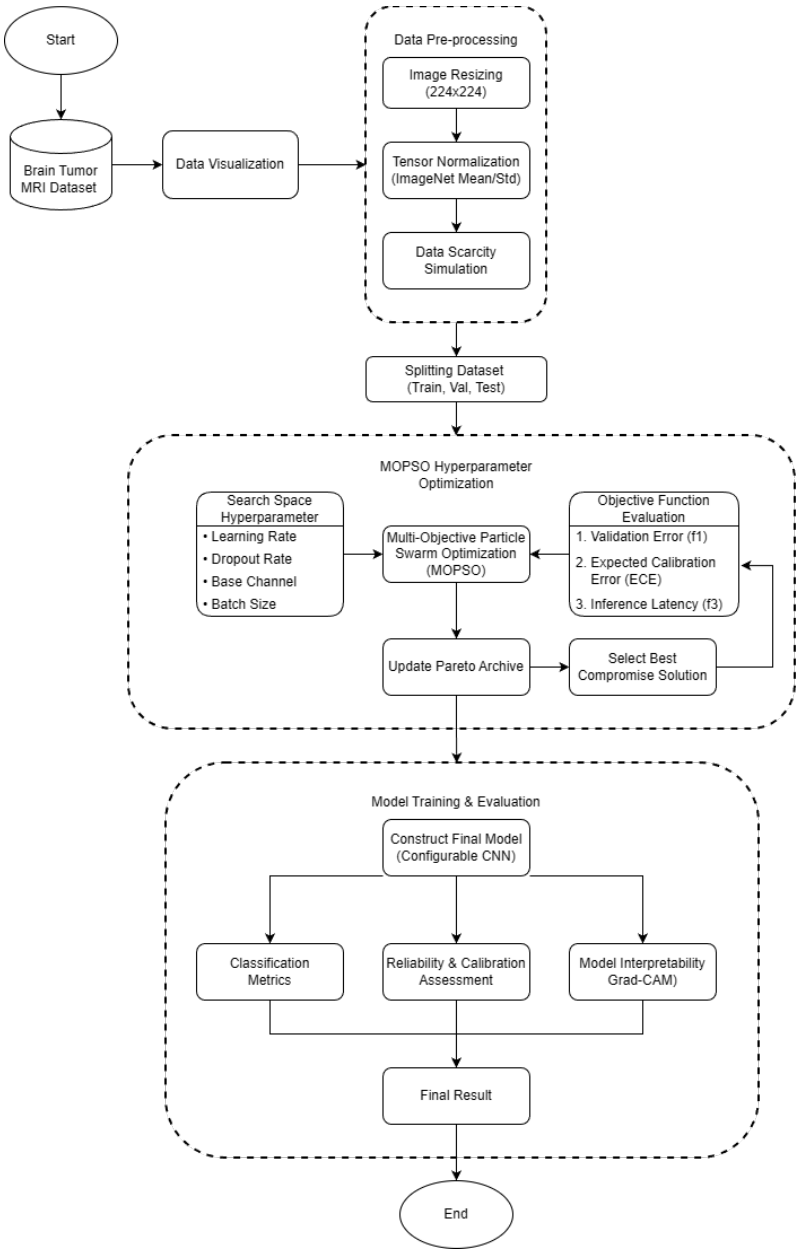

Figure 1: Workflow of the Proposed MOPSO-Optimized CNN Model.

3.1. Dataset and Preprocessing

The study employs the Brain Tumor MRI Dataset obtained from publicly accessible repositories (Masoud Nickparvar, Kaggle). The dataset comprises magnetic resonance imaging (MRI) scans that are categorized into four diagnostic classes: Glioma, Meningioma, Pituitary, and No Tumor. Prior to model training, several preprocessing procedures are applied to ensure uniformity and compatibility with the convolutional neural network (CNN) architecture.

All MRI images are first resized to a fixed spatial dimension of 224×224 pixels to standardize the input size and support efficient batch processing. Subsequently, pixel intensity values are normalized using the mean (μ) and standard deviation (σ) parameters from the ImageNet dataset, a widely adopted normalization scheme for deep learning–based image classification models. This normalization aims to stabilize gradient updates and accelerate model convergence. The normalized pixel value $x_{\text{norm}}$is computed using:

$$x_{\text{norm}} = \frac{x-\mu}{\sigma} \quad (1)$$

where $\mu = [0.485, 0.456, 0.406]$ and $\sigma = [0.229, 0.224, 0.225]$.

3.2 Configurable CNN Architecture

The proposed system is built upon a configurable convolutional neural network (CNN) architecture that enables flexible adjustment of hyperparameters during the optimization stage. The network is composed of three sequential convolutional blocks that progressively extract hierarchical feature representations from the input MRI scans. Each block integrates a two-dimensional convolutional layer with a 3×3 kernel, followed by batch normalization to enhance training stability by mitigating internal covariate shift. A Rectified Linear Unit (ReLU) activation function is subsequently applied to introduce nonlinearity, while a max-pooling operation with a 2×2 kernel performs spatial down-sampling to reduce computational complexity and emphasize the most salient features.

After the feature extraction stage, the output tensor is transformed into a one-dimensional representation through a flattening operation. To improve model generalization and reduce overfitting, a dropout layer is incorporated, where the dropout rate is treated as a tunable parameter optimized using the Multi-Objective Particle Swarm Optimization (MOPSO) algorithm. The final classification is performed using a fully connected linear layer that maps the learned features to the corresponding tumor categories.

3.3 Multi-Objective Optimization Strategy

To identify the optimal set of hyperparameters, this study employs a Multi-Objective Particle Swarm Optimization (MOPSO) framework. Within this framework, each particle $i$in the swarm encodes a candidate hyperparameter vector x_i, which may include parameters such as the learning rate, dropout rate, and the number of base convolutional channels. The swarm evolves iteratively, and at each iteration t, the position $x_i(t)$ and velocity $v_i(t)$ of each particle are updated following the standard Particle Swarm Optimization (PSO) formulation:

Artificial Intelligence Systems and Its Applications (AISA)
Volume 1, Issue 2, December 2025
http://journals.cognispectra.com/index.php/aisa/index
E-ISSN: 3110-0457

$$v_i(t+1) = w\, v_i(t) + c_1 r_1 (pbest_i - x_i(t)) + c_2 r_2 (gbest - x_i(t)) \quad (2)$$
$$x_i(t+1) = x_i(t) + v_i(t+1) \quad (3)$$

where $w$denotes the inertia weight, c_1 and c_2 are the cognitive and social acceleration coefficients, respectively, and r_1 and r_2 represent random vectors uniformly sampled from the interval $[0,1]$. Throughout the optimization process, all non-dominated solutions are archived in an external Pareto Repository, which functions as the global guide for the swarm and ensures adequate coverage of the Pareto front.

The optimization simultaneously minimizes three objective functions, each representing a distinct performance aspect of the CNN model. Unlike standard MOPSO implementations that typically focus solely on error rates, our framework modifies the fitness evaluation step to explicitly include Expected Calibration Error (ECE) as a critical minimization objective (Equation 5). This modification forces the swarm to navigate towards solutions that are not only accurate but also probabilistically reliable. The first objective is the validation error, expressed as:

$$f_1 = 1 - \text{Accuracy}_{\text{val}} \quad (4)$$

which encourages the model to achieve high predictive accuracy on unseen validation samples. The second objective aims to improve probabilistic reliability by minimizing the Expected Calibration Error (ECE). Model outputs are partitioned into Mconfidence bins, and calibration quality is assessed by comparing the accuracy and mean confidence of each bin. The ECE is defined as:

$$f_2 = \text{ECE} = \sum_{m=1}^{M} \frac{|B_m|}{N} \left| \text{acc}(B_m) - \text{conf}(B_m) \right| \quad (5)$$

where $|B_m|$ denotes the number of samples in bin m, N is the total number of samples, $\text{acc}(B_m)$ is the empirical accuracy of the bin, and $\text{conf}(B_m)$ is its average predicted confidence. Minimizing this objective enhances the trustworthiness of the model's confidence estimates, which is crucial for medical decision support applications.

The third objective addresses computational performance by measuring inference latency, defined as the average time required to process an individual MRI input during evaluation. Latency is computed as:

$$f_3 = \frac{1}{K} \sum_{k=1}^{K} (t_{\text{end}}^{(k)} - t_{\text{start}}^{(k)}) \quad (6)$$

where K represents the number of batches used for latency profiling, and $t_{\text{start}}^{(k)}$ and $t_{\text{end}}^{(k)}$ correspond to the timestamps recorded before and after the forward pass of batch k. This objective ensures that the optimized model remains suitable for real-time or near real-time clinical deployment.

To support the MOPSO optimization process, the search space of tunable hyperparameters is summarized in Table 2.

Table 2: Hyperparameter Search Space

Hyperparameter	Type	Range / Values	Description
Learning Rate	Continuous	$10^{-5} - 10^{-2}$	Controls the step size of weight updates.
Dropout Rate	Continuous	$0.0 - 0.5$	Prevents overfitting by randomly deactivating neurons.
Base Channels	Integer	$16 - 64$	Determines the number of filters in the initial convolutional layer.
Batch Size	Discrete	{8,16,32,64}	Number of samples processed per iteration.

3.4 Experimental Setup

All experiments were carried out in the Google Colab environment equipped with an NVIDIA T4 GPU, providing the computational resources necessary for training and evaluating the proposed model. The implementation was developed using Python 3 and the PyTorch deep learning framework, which offers efficient tensor operations and flexible model customization suitable for the optimization workflow.

The final CNN configuration corresponds to the set of hyperparameters selected from the Pareto-optimal solutions obtained through the MOPSO process. Model training employs the Adam optimizer to adaptively adjust learning rates during gradient-based updates and utilizes the Cross-Entropy Loss function as the objective for multi-class tumor classification. This setup ensures that the learning process is both stable and aligned with the hyperparameter values determined through multi-objective optimization.

4 Results and Discussion

4.1. Multi-Objective Optimization Dynamics

The proposed Multi-Objective Particle Swarm Optimization (MOPSO) framework was employed to navigate the complex hyperparameter search space defined in Chapter 3. Over the course of ten optimization iterations, the swarm systematically explored the trade-offs among the three conflicting objectives: Validation Error (f_1), Expected Calibration Error (f_2), and Inference Latency (f_3). The optimization logs demonstrate a clear progression in swarm dynamics throughout the search process.

Artificial Intelligence Systems and Its Applications (AISA)
Volume 1, Issue 2, December 2025
http://journals.cognispectra.com/index.php/aisa/index
E-ISSN: 3110-0457

During the early stage of optimization (iterations 1 to 3), particle performance exhibited substantial variance, indicating a strong exploratory behavior. Validation errors during this phase ranged from 29.8 percent down to 14.4 percent, while calibration errors frequently exceeded 10 percent. This period reflects the swarm's initial investigation of diverse regions of the hyperparameter space, which is typical of the exploratory phase where the algorithm probes broadly before moving toward more promising areas.

By iteration 6, the swarm began to converge as particles clustered around competitive solution regions. At this stage, the optimization process identified potential Pareto-optimal configurations that balanced model complexity, predictive accuracy, and computational efficiency. This transition marked the shift from exploration to exploitation and signaled a more refined search toward optimal decision boundaries.

The final Pareto front presented several non-dominated solutions that illustrated the inherent trade-offs among the three objectives. Certain configurations favored accuracy and achieved validation errors of approximately 12.2 percent, but these models exhibited higher inference latency, often exceeding 1.2 milliseconds per sample, which is typical for architectures with larger convolutional channel depths. In contrast, solutions optimized for computational efficiency achieved extremely low latency, around 0.28 milliseconds per sample, particularly those with lightweight architectures such as 16 base channels. However, these configurations tended to produce higher validation errors above 17 percent. Between these extremes, the optimization uncovered a distinct knee region of the Pareto front representing a balanced compromise among accuracy, calibration, and latency.

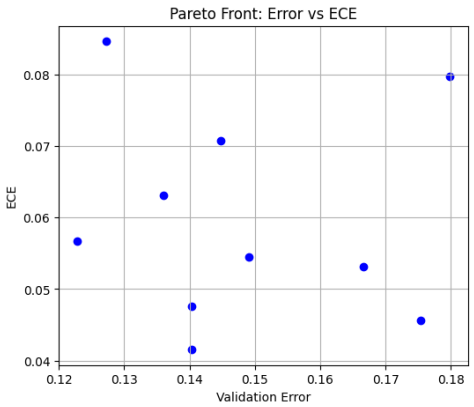

Figure 2: Pareto Front: Validation Error vs Expected Calibration Error (ECE).

Figure 2 illustrates the trade-off between Validation Error and ECE. As shown, points located toward the lower-left region of the plot represent more desirable configurations, whereas points

on the upper-right reflect inferior trade-offs. The spread of solutions confirms the conflicting nature of error minimization and calibration reliability.

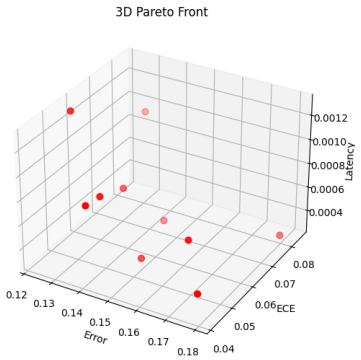

Figure 3: Pareto Front: Three-Dimensional Pareto Front of Validation Error, ECE, and Inference Latency.

A broader perspective on the optimization landscape is provided by the three-dimensional Pareto Front (Figure 3), in which Latency is incorporated as the third axis. This visualization highlights how the swarm discovers diverse solution clusters and clarifies the relative position of the knee region. It also shows that extremely low latency solutions tend to correspond with higher error values, while highly accurate configurations tend to incur additional computational cost.

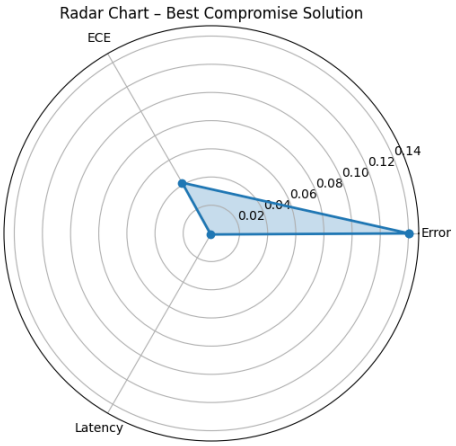

Figure 4. Radar Chart of the Best Compromise Solution.

To illustrate why the chosen model configuration represents the best compromise, a Radar Chart of the Best Compromise Solution (Figure 4) is provided. The radar plot displays the relative magnitudes of the three objectives for the selected model. The balanced shape of the polygon confirms that this configuration avoids extreme values in any objective dimension, making it suitable for real-time and high-reliability medical deployment.

The Best Compromise Solution selected from the Pareto knee region employed a learning rate of 3.49×10^{-5}, a dropout rate of 0.394, 32 base channels, and a batch size of 16. This configuration

Artificial Intelligence Systems and Its Applications (AISA)
Volume 1, Issue 2, December 2025
http://journals.cognispectra.com/index.php/aisa/index
E-ISSN: 3110-0457

produced a validation error of 14.0 percent and an Expected Calibration Error of 4.1 percent during the optimization phase, while maintaining an efficient inference latency of 0.88 milliseconds per sample. These results demonstrate the effectiveness of MOPSO in identifying high-quality hyperparameter configurations that reconcile the competing demands of predictive performance, reliability, and computational speed.

4.2. Diagnostic Performance Analysis

Following the optimization phase, the final CNN model was retrained using the optimal hyperparameters obtained from the MOPSO procedure. The model was then evaluated on the held-out test set consisting of 1,311 MRI scans. Quantitative results indicate that the optimized model achieved a Test Accuracy of 95 percent, demonstrating strong overall predictive capability across the four tumor categories.

Class-wise performance metrics are summarized in Table 3. The model exhibited consistently high precision, recall, and F1-Score across all classes, reflecting reliable discriminative behavior. In particular, the model achieved an F1-Score of 0.99 for both Pituitary tumors and No Tumor cases. The strong performance on the No Tumor category is clinically meaningful because it reduces the likelihood of false positives, which are particularly undesirable in diagnostic screening scenarios.

Table 3: Classification Report of the Optimized Model

Class	Precision	Recall	F1-Score	Support
Glioma	91%	93%	92%	300
Meningioma	92%	87%	89%	306
No Tumor	97%	100%	99%	405
Pituitary	99%	99%	99%	300
Overall	95%	95%	95%	1311

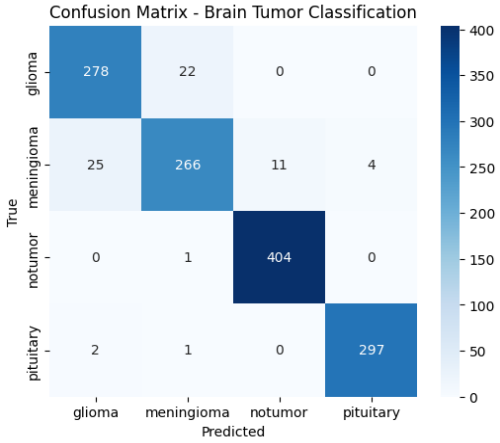

Figure 5: Confussion Matrix.

The confusion matrix, illustrated in Figure 5, provides further insight into model behavior. The most notable source of error occurs within the Meningioma category, which recorded a recall of 0.87. A portion of Meningioma samples was misclassified as Glioma. This pattern of misclassification aligns with established radiological challenges, as both tumor types can appear visually similar, particularly when presenting as extra-axial masses with enhancing margins in specific MRI sequences.

Despite this challenge, the confusion matrix maintains strong diagonal dominance, indicating that the classifier generalizes well across categories. The relatively small number of off-diagonal entries further supports the robustness of the optimized model in distinguishing between brain tumor types in diverse imaging conditions.

4.3. Reliability and Calibration Assessment

A critical objective of this study was to ensure that the confidence scores produced by the model accurately reflect the true likelihood of correct classification. Standard deep learning classifiers are known to exhibit overconfidence, which can lead to misleading uncertainty estimates in clinical decision making. By incorporating the calibration-aware objective function f_2, the optimization process effectively reduced the mismatch between predicted confidence and empirical accuracy.

The final evaluation produced a Test Expected Calibration Error (ECE) of 0.0148, or 1.48 percent. This low ECE value indicates that when the model assigns a confidence of 90 percent to a prediction, the prediction is correct approximately 90 percent of the time. The Reliability Diagram (Calibration Curve) presented in Figure 6 further supports this finding. The plotted curve closely aligns with the ideal diagonal line $y = x$, signifying that the model maintains consistent alignment between confidence and actual predictive performance.

Artificial Intelligence Systems and Its Applications (AISA)
Volume 1, Issue 2, December 2025
http://journals.cognispectra.com/index.php/aisa/index
E-ISSN: 3110-0457

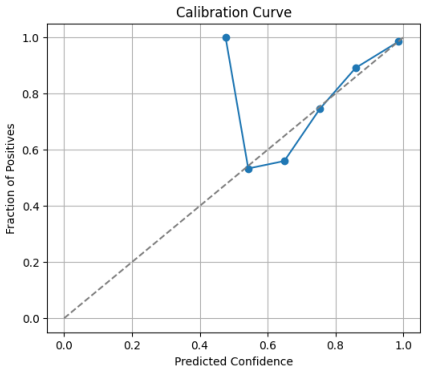

Figure 6: Calibration Curve.

This level of reliability is essential for the development of trustworthy artificial intelligence systems in healthcare. A well-calibrated model enables clinicians and radiologists to interpret the model's confidence estimates with greater assurance, particularly in scenarios where uncertainty plays a critical role in diagnostic decisions. As a result, the proposed calibration-aware optimization strategy contributes directly to the model's suitability for real-world diagnostic support applications.

4.4. Model Interpretability (Grad-CAM)

To validate the decision-making process of the CNN, Gradient-weighted Class Activation Mapping (Grad-CAM) was employed. This technique provides a visual explanation by highlighting the spatial regions that contribute most to the model's predictions. As illustrated in Figure 4, the Grad-CAM heatmaps for correctly classified samples demonstrate that the model consistently directs its attention toward the hyperintense tumor regions within the brain parenchyma. These regions correspond to the pathological areas that radiologists typically examine during clinical assessment.

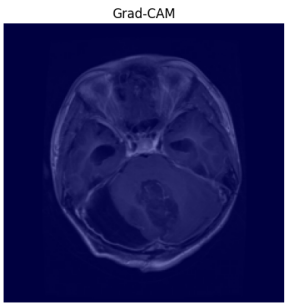

Figure 7: Grad-CAM.

The interpretability results further show that the model effectively disregards irrelevant structures such as background noise and cranial bone features, indicating that its predictions are driven by meaningful and disease-relevant patterns rather than spurious artifacts. This qualitative evidence supports the claim that the model satisfies the explainability requirements necessary for safe and trustworthy medical AI applications.

4.5. Computational Efficiency

Beyond accuracy, calibration, and interpretability, computational performance was an essential optimization objective in this study. The MOPSO framework successfully identified a model configuration that delivers high predictive capability while maintaining low inference latency. The selected architecture operates with an average inference latency of 0.88×10^{-3}seconds per sample, which corresponds to less than one millisecond on the test hardware.

This sub-millisecond inference time indicates that the model can process more than 1,000 MRI slices per second. Such computational efficiency is critical for real-time or near real-time deployment in clinical workflows. The model's speed allows it to be integrated seamlessly into Picture Archiving and Communication Systems (PACS) without introducing delays, ensuring that diagnostic pipelines remain efficient and responsive. The combination of rapid inference and high diagnostic accuracy enhances the practicality of the system for clinical use, particularly in high-throughput environments.

4.6. Comparative Performance Against Previous Studies

The performance of the proposed MOPSO-optimized CNN model demonstrates significant improvements when compared with prior brain tumor MRI classification studies. After undergoing multi-objective optimization, the final model achieved a test accuracy of 95 percent, placing it among the highest performing deep learning approaches in this domain. Beyond accuracy, the model also achieved an Expected Calibration Error of 1.48 percent and a sub-millisecond inference latency, providing a balanced combination of accuracy, reliability, and computational efficiency that earlier works seldom addressed simultaneously.

Previous studies in brain tumor classification reported a wide range of accuracy results depending on feature extraction techniques, model architectures, and dataset sizes. Traditional approaches that relied on handcrafted features, such as GLCM combined with Probabilistic Neural Networks, attained an accuracy of 95 percent on relatively small datasets consisting of 650 DICOM samples [9]. Deep learning methods trained on larger MRI collections generally achieved moderate improvements, such as 88.77 percent accuracy using MobileNetV2 [10], 91.5 percent using augmented Deep CNN models [11], and 93.30 percent using custom CNN architectures [19].

Artificial Intelligence Systems and Its Applications (AISA)
Volume 1, Issue 2, December 2025
http://journals.cognispectra.com/index.php/aisa/index
E-ISSN: 3110-0457

Object detection based approaches like YOLOv5 have been evaluated using precision, recall, and mAP metrics rather than pure accuracy scores. These models achieved precision and recall values between 81 and 86 percent, with mAP values ranging from 85 to 87 percent, illustrating strong localization performance but leaving room for improvement in classification tasks [12]. Transfer learning techniques also contributed notable performance gains, including a TTL-based model reaching 94.5 percent accuracy [13] and VGG-16 achieving up to 94.98 percent in binary classification and 89.19 percent in multiclass settings [20].

Compared with these studies, the proposed model demonstrates competitive, and in many cases superior, classification performance while simultaneously enhancing model calibration and inference speed. These improvements stem from the use of a multi-objective optimization strategy, which tunes hyperparameters not only for accuracy but also for reliability and computational efficiency. Consequently, the model is better aligned with clinical deployment requirements, where both trustworthiness and real-time performance are essential.

Table 4: Comparison of the Proposed Model and Previous Studies

Author	Dataset Size	Method	Metric	Result
[9]	650 DICOM	PNN + GLCM	Accuracy	95%
[10]	4600 images	MobileNetV2	Accuracy	88.77%
[11]	5712 images (augmented 142k)	Deep CNN	Accuracy	91.5%
[12]	3064 MRI scans	YOLOv5	Precision	81.9%
			Recall	83%
			mAP	87%
		YOLOv5 + NLNNs	Precision	83.5%
			Recall	86%
			mAP	85.2%
[13]	6000 scans	TTL Transfer Learning	Accuracy	94.5%
[14]	3064 scans	Inception-v3 Ensemble	Accuracy	94.34%
[15]	2870 scans	GLCM + HOG + LBP (KNN)	Accuracy	91.1%
[16]	7023 scans	Three-Layer CNN	Accuracy	89.79%
[17]	-	LeNet-based CNN	Accuracy	88%
[18]	3064 scans	Fisher Vector	mAP	94.68%
[19]	3264 scans	Custom CNN	Accuracy	93.30%
[20]	7000 scans	VGG-16 (binary)	Accuracy	94.98%
	Multiclass	VGG-16	Accuracy	89.19%
Proposed Method	1311 scans (test set)	MOPSO-Optimized CNN	Accuracy	**95%**
			ECE	**1.48%**
			Latency	**0.88 ms**

5 Conclusion

This study presented a Multi-Objective Particle Swarm Optimization (MOPSO)-driven framework for optimizing a Convolutional Neural Network (CNN) applied to brain tumor classification using MRI scans. By simultaneously optimizing validation error, calibration reliability, and inference latency, the

proposed approach offers a comprehensive performance improvement that addresses practical demands in clinical diagnostics. The resulting model achieved high predictive accuracy combined with exceptionally low calibration error and sub-millisecond inference latency, demonstrating its suitability for real-time deployment.

One of the key contributions of this work lies in incorporating calibration awareness into the optimization process, enabling the model to produce probability estimates that accurately reflect predictive correctness. Additionally, the integration of latency as an optimization objective ensures efficient computational performance, which is often overlooked in deep learning studies for medical imaging. The Grad-CAM interpretability analysis further validated that model predictions are guided by clinically relevant tumor regions, enhancing trustworthiness and transparency.

Comparison with existing literature shows that the proposed method matches or exceeds the performance of state-of-the-art CNN and transfer learning models while additionally offering improved calibration and computational efficiency. These advancements make the model highly promising for integration into clinical workflows, such as real-time screening systems or PACS-based diagnostic support tools.

Despite these strengths, several areas remain open for exploration. Future research may include extending the model to multi-modal MRI scans (such as T1, T2, and FLAIR), incorporating volumetric 3D CNN architectures, or integrating hybrid optimization techniques such as NSGA-II or MOEA/D for broader performance comparison. The incorporation of uncertainty quantification techniques, beyond ECE, could further support its clinical adoption in high-risk decision-making environments.

While the proposed model demonstrates strong performance on the evaluated dataset, the limitation of external validity inherent in single-dataset studies is acknowledged. Variations in MRI acquisition parameters and scanner types across different medical centers may affect model generalizability. Therefore, future research will prioritize validating this framework on multi-center datasets to ensure robustness across diverse clinical environments.

Overall, the proposed MOPSO-optimized CNN presents a strong foundation for reliable, efficient, and interpretable medical image classification, contributing valuable advancements to the development of trustworthy artificial intelligence in healthcare.

Funding Information

”The authors declare no funding was received for this study.”

Conflict of Interest Statement

”The authors declare no conflicts of interest.”

Ethical Approval

”This study did not involve human or animal subjects.”

Data Availability

The dataset used in this study is publicly available and can be accessed through the Brain Tumor MRI Dataset repository at https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset. No additional datasets were generated or analyzed, and all data utilized in this research are openly accessible. Therefore, data sharing beyond the referenced repository is not applicable.

References

[1] F. J. Dorfner, J. B. Patel, J. Kalpathy-Cramer, E. R. Gerstner, and C. P. Bridge, “A review of deep learning for brain tumor analysis in MRI,” Dec. 01, 2025, *Nature Research*. doi: 10.1038/s41698-024-00789-2.

[2] G. Litjens *et al.*, “A survey on deep learning in medical image analysis,” Dec. 01, 2017, *Elsevier B.V.* doi: 10.1016/j.media.2017.07.005.

[3] T. S. Azzahra, J. J. Cerelia, F. Azhar, L. Nugraha, and A. A. Pravitasari, “MRI-Based Brain Tumor Classification Using Inception Resnet V2,” *ENTHUSIASTIC INTERNATIONAL JOURNAL OF APPLIED STATISTICS AND DATA SCIENCE*, vol. 3, pp. 163–175, Oct. 2023, doi: 10.20885/enthusiastic.vol3.iss2.art4.

[4] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, “On Calibration of Modern Neural Networks,” 2017. doi: 10.48550/arXiv.1706.04599.

[5] M. Latha, P. S. Kumar, R. R. Chandrika, T. R. Mahesh, V. V. Kumar, and S. Guluwadi, “Revolutionizing breast ultrasound diagnostics with EfficientNet-B7 and Explainable AI,” *BMC Med Imaging*, vol. 24, no. 1, Dec. 2024, doi: 10.1186/s12880-024-01404-3.

[6] J. Kennedy, “Particle Swarm Optimization,” in *Encyclopedia of Machine Learning*, C. Sammut and G. I. Webb, Eds., Boston, MA: Springer US, 2010, pp. 760–766. doi: 10.1007/978-0-387-30164-8_630.

[7] I. Malashin, V. Tynchenko, A. Gantimurov, V. Nelyub, and A. Borodulin, “A Multi-Objective Optimization of Neural Networks for Predicting the Physical Properties of Textile Polymer Composite Materials,” *Polymers (Basel)*, vol. 16, no. 12, Jun. 2024, doi: 10.3390/polym16121752.

[8] S. J. Hosseini Dehshiri, A. Yousefi Hanoomarvar, and M. Amiri, “Comparative performance of the NSGA-II and MOPSO algorithms and simulations for evaluating time–cost–quality–risk trade-off in multi-modal PERT networks,” *Soft comput*, vol. 27, no. 24, pp. 18651–18666, 2023, doi: 10.1007/s00500-023-09099-4.

[9] N. Varuna Shree and T. N. R. Kumar, “Identification and classification of brain tumor MRI images with feature extraction using DWT and probabilistic neural network,” *Brain Inform*, vol. 5, no. 1, pp. 23–30, Mar. 2018, doi: 10.1007/s40708-017-0075-5.

[10] N. A. Husna, D. Hendri, M. F. Wajdi, E. S. Ginting, and C. H. Pramesthi, "Implementation of Deep Learning for Brain Tumor Classification from Magnetic Resonance Imaging," *Public Research Journal of Engineering, Data Technology and Computer Science*, vol. 3, no. 1, pp. 31–41, Jul. 2025, doi: 10.57152/predatecs.v3i1.1570.

[11] S. Immaculate Joy, G. Sriram, and S. Sriram Venkatesan, "Deep CNN-Based Multi-Grade Brain Tumor Classification with Enhanced Data Augmentation," in *Procedia Computer Science*, Elsevier B.V., 2025, pp. 300–307. doi: 10.1016/j.procs.2025.03.205.

[12] A. Abdusalomov, M. Rakhimov, J. Karimberdiyev, G. Belalova, and Y. I. Cho, "Enhancing Automated Brain Tumor Detection Accuracy Using Artificial Intelligence Approaches for Healthcare Environments," *Bioengineering*, vol. 11, no. 6, Jun. 2024, doi: 10.3390/bioengineering11060627.

[13] R. W. Anwar, M. Abrar, and F. Ullah, "Transformative Transfer Learning for MRI Brain Tumor Precision: Innovative Insights," *IEEE Access*, vol. 13, pp. 31749–31761, 2025, doi: 10.1109/ACCESS.2025.3542451.

[14] N. Noreen, S. Palaniappan, A. Qayyum, I. Ahmad, and M. O. Alassafi, "Brain Tumor Classification Based on Fine-Tuned Models and the Ensemble Method," *Computers, Materials and Continua*, vol. 67, no. 3, pp. 3967–3982, Mar. 2021, doi: 10.32604/cmc.2021.014158.

[15] B. B. Pattanaik, K. Anitha, S. Rathore, P. Biswas, P. K. Sethy, and S. K. Behera, "Brain tumor magnetic resonance images classification based machine learning paradigms," *Wspolczesna Onkologia*, vol. 26, no. 4, pp. 268–274, 2022, doi: 10.5114/wo.2023.124612.

[16] P. M. Custodio, "Improving MRI Classification through Layered Convolutional Neural Networks Configuration," vol. 7, no. 1, pp. 38–44, 2025, doi: 10.59356/s.

[17] I. Wahlang *et al.*, "Brain Magnetic Resonance Imaging Classification Using Deep Learning Architectures with Gender and Age," *Sensors*, vol. 22, no. 5, Mar. 2022, doi: 10.3390/s22051766.

[18] J. Cheng *et al.*, "Retrieval of Brain Tumors by Adaptive Spatial Pooling and Fisher Vector Representation," *PLoS One*, vol. 11, no. 6, Jun. 2016, doi: 10.1371/journal.pone.0157112.

[19] M. I. Mahmud, M. Mamun, and A. Abdelgawad, "A Deep Analysis of Brain Tumor Detection from MR Images Using Deep Learning Networks," *Algorithms*, vol. 16, no. 4, Apr. 2023, doi: 10.3390/a16040176.

[20] H. Al Sukhni *et al.*, "Brain Tumor Detection: Integrating Machine Learning and Deep Learning for Robust Brain Tumor Classification," *Journal of Intelligent Systems and Internet of Things*, vol. 15, no. 1, pp. 1–16, 2025, doi: 10.54216/JISIoT.150101.