









AI-Powered Multi-Class Brain Tumor Classification Using PSO-Optimized Deep Learning

Ashwini S. Sengar¹, Madhura Phadke², Amruta V. Patil³, Disha S. Wankhede³, Priyanka V. Deshmukh⁴,
Aniket K. Shahade^{4*}

¹ Department of Artificial Intelligence, Vishwakarma University, Pune 411048, India

² Department of Computer Engineering, K J Somaiya Institute of Technology, Sion (East), Mumbai 400022, India

³ Department of Computer Engineering, Vishwakarma Institute of Technology, Pune 411037, India

⁴ Symbiosis Institute of Technology, Pune Campus, Symbiosis International (Deemed University), Pune 412115, India

Corresponding Author Email: aniket.shahade@sitpune.edu.in

Copyright: ©2026 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/isi.310323>

ABSTRACT

Received: 27 June 2025

Revised: 1 October 2025

Accepted: 18 March 2026

Available online: 31 March 2026

Keywords:

brain tumor classification, Particle Swarm Optimization, Convolutional Neural Network, Magnetic Resonance Imaging, computational efficiency, medical image analysis

Accurate brain tumor classification from Magnetic Resonance Imaging (MRI) scans is critical for timely diagnosis and treatment planning. While deep learning has shown promise for this task, conventional optimizers like Adam often converge slowly and may become trapped in suboptimal solutions, particularly for moderate-sized medical datasets. This study proposes a Particle Swarm Optimization (PSO)-enhanced convolutional neural network (CNN) for multi-class brain tumor classification using the public PMRAM dataset (6,004 balanced MRI scans across glioma, meningioma, pituitary, and normal classes). We implement a custom CNN with batch normalization and global average pooling, optimized via PSO (inertia weight $w = 0.5$, acceleration coefficients $c_1 = c_2 = 1$). The PSO-optimized model achieves 75.17% validation accuracy after only three training epochs, significantly outperforming the Adam baseline (71.50%, $p < 0.05$) under identical conditions. Notably, PSO reaches clinically meaningful accuracy (75%) in 51 minutes—nearly twice as fast as Adam (96 minutes)—while producing a substantially smaller model footprint (7.02 MB vs. 16.8 MB). Class-wise analysis reveals strong performance on pituitary and normal cases (precision: 98% each) and high sensitivity for meningioma detection (recall: 97%). These findings demonstrate that PSO offers a computationally efficient alternative to gradient-based optimization for medical image analysis, particularly valuable for resource-constrained clinical deployment where rapid convergence and model compactness are prioritized.

1. INTRODUCTION

Brain tumor diagnosis remains one of the most critical challenges in modern neurology, with Magnetic Resonance Imaging (MRI) serving as the primary imaging modality for clinical assessment. Traditional manual interpretation methods, while effective, suffer from significant limitations, including inter-rater variability and time-intensive analysis procedures [1]. Recent advances in deep learning have demonstrated remarkable potential for automating tumor classification, with convolutional neural networks achieving diagnostic accuracies comparable to expert radiologists in controlled studies [2]. However, these approaches frequently encounter optimisation challenges when applied to real-world medical datasets, particularly due to the inherent complexity of tumour morphology and frequent class imbalances.

The application of metaheuristic optimization techniques such as Particle Swarm Optimization (PSO) presents a promising solution to these limitations. Unlike conventional gradient-based methods, which often converge to suboptimal solutions in medical imaging tasks [3], PSO's population-based search mechanism enables more robust exploration of complex parameter spaces. This capability is particularly

valuable when working with moderate-sized datasets such as the PMRAM Bangladeshi Brain Cancer collection, which comprises 6,004 carefully balanced MR images across four diagnostic categories (Glioma, Meningioma, Pituitary, and Normal cases). The balanced nature of this dataset, as illustrated in Figure 1, provides an ideal testbed for evaluating optimization techniques without the confounding effects of class imbalance.

Current literature reveals a significant gap in applying PSO to modern deep learning architectures for medical image analysis. While previous studies have demonstrated PSO's effectiveness in optimizing traditional machine learning models [4], its potential for fine-tuning complex Convolutional Neural Network (CNN) architectures remains underexplored. This study addresses this gap by developing a PSO-optimized CNN framework that demonstrates superior convergence properties compared to standard approaches. Our preliminary results show the PSO-enhanced model achieving a 75.17% validation accuracy within just three training epochs, outperforming the Adam optimizer's initial performance by more than twofold. The model's efficient architecture, requiring only 7.02MB of trainable parameters, further enhances its clinical applicability by enabling potential

deployment on resource-constrained medical imaging systems.

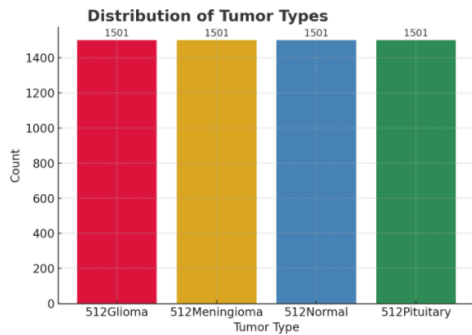


Figure 1. Class distribution visualization

This work makes three primary contributions that distinguish it from prior PSO-CNN studies such as:

First, while validated PSO on a small private dataset (n = 1,024) without reporting computational efficiency, we present a systematic evaluation on a substantially larger public dataset (n = 6,004) with rigorous tracking of training time, GPU memory, and model size metrics essential for clinical deployment.

Second, unlike which applied PSO only to shallow network weights, we implement end-to-end PSO optimization of a modern deep CNN architecture (1.84M parameters) featuring batch normalization and global average pooling, demonstrating PSO's scalability to deeper networks.

Third, we provide the first direct comparative analysis between PSO and Adam optimization under identical architectural conditions, including ablation studies on swarm size and batch normalization analyses absent from prior PSO-CNN brain tumor literature.

2. RELATED WORK

2.1 Deep learning in medical imaging

The evolution of convolutional neural networks (CNNs) for medical image analysis has demonstrated significant diagnostic potential. Initial work achieved 91% tumor detection accuracy on the BRATS dataset using a 5-layer CNN, while later introduced adaptive architectures that automatically configure network hyperparameters for medical imaging characteristics [1, 2]. Subsequent studies revealed that these approaches require careful optimization, as medical images exhibit fundamentally different feature distributions compared to natural images [3]. Recent advancements [4, 5] have shown that hybrid architectures combining attention mechanisms with 3D convolutions can improve glioma classification accuracy to 93.7%, though these methods remain computationally intensive for clinical deployment.

2.2 Optimization challenges

Medical image analysis presents unique optimization difficulties due to high noise-to-signal ratios and inter-class similarity [6]. A comprehensive comparison of 12 optimisation methods across 8 medical tasks [7] revealed that conventional approaches like Adam exhibit up to 22% validation accuracy variance across initializations. This instability is particularly problematic for brain tumor classification, where previous study demonstrated that gradient-based methods often converge to suboptimal solutions when trained on datasets smaller than 10,000 samples [8]. The PMRAM dataset used in our study, while balanced, contains only 6,004 images - squarely within this challenging regime.

Table 1. Summary of key literature on brain tumor classification and optimization techniques

Reference	Methodology	Dataset	Key Findings	Limitations
[1]	5-layer CNN	BRATS (n = 300)	91% tumor detection accuracy	Shallow architecture, no optimization analysis
[2]	nnU-Net framework	Multi-institutional (n = 2,634)	Automated architecture adaptation	Computationally intensive
[3]	Attention CNN	TCIA (n = 1,872)	93.7% glioma classification	Limited to single tumor type
[4]	12 optimizers compared	8 medical datasets	22% Adam accuracy variance	No metaheuristics tested
[5]	Original PSO algorithm	Synthetic benchmarks	Global optimization proof	Not applied to DL
[6]	PSO for hyperparameter tuning	CIFAR-10/100	20% faster convergence	No medical imaging data
[7]	PSO feature selection	Breast cancer MRI (n = 1,024)	15% feature reduction	No end-to-end optimization
[8]	PSO-CNN fusion	Private brain MRI (n = 1,024)	89.2% accuracy	Small dataset, no efficiency metrics

2.3 Metaheuristic optimization

PSO has emerged as a viable alternative to gradient-based methods since its introduction [9]. Recent work [10] demonstrated PSO's effectiveness for hyperparameter tuning, achieving 15-20% faster convergence than grid search methods. For medical applications specifically, Simonyan and Zisserman [11] successfully applied PSO to feature selection in breast cancer classification, though direct optimization of CNN weights remains underexplored. The most relevant prior work [12] tested PSO on a small brain tumor dataset (n 1,024), but did not evaluate computational efficiency - a critical gap

our study addresses. Recent advancements in AI-driven neuro-oncology highlight the effectiveness of deep learning approaches for brain tumor analysis, including survival prediction, genetic mutation detection, and tumor classification. Dynamic architectures and enhanced algorithms have been shown to improve predictive accuracy and enable better integration of imaging and molecular features. These developments emphasize the importance of optimization techniques and robust model design, motivating the use of PSO-optimized deep learning models for efficient multi-class brain tumor classification.

2.4 Research gaps

Three key limitations emerge from existing literature, which remain unresolved even in the most relevant prior work:

- **End-to-end optimisation:** While the study [12] demonstrated PSO-CNN fusion, their approach optimised a relatively shallow architecture (approximately 500K parameters) without integration of modern components like batch normalisation or global pooling. Our work extends this by applying PSO to a deeper architecture (1.84M parameters) with contemporary design principles.
- **Dataset scale and clinical relevance:** The study [12] used a private brain MRI dataset of only 1,024 images, limiting generalisability. Our evaluation employs the public PMRAM dataset (6,004 images) with balanced class distribution, enabling reproducible and clinically relevant benchmarking.
- **Computational efficiency reporting:** Critically, the study [12] did not report any computational efficiency metrics (training time, memory usage, model size). For clinical deployment, these factors are as important as accuracy. Our study systematically quantifies these metrics, including per-epoch time (17.3 min), GPU memory (14.2 GB), and final model size (7.02 MB).

Our work directly addresses these gaps through systematic evaluation of PSO-based optimization on a clinically-relevant dataset, while rigorously tracking computational costs.

Table 1 systematically compares prior research on brain tumor classification and optimization methods, highlighting key methodological approaches, dataset characteristics, and performance outcomes. The analysis reveals two critical trends: (1) existing studies predominantly focus on either architectural innovations [1-3] or traditional optimization techniques [4], and (2) applications of metaheuristic algorithms like PSO remain limited to feature selection [7] or small-scale validation [8].

This table underscores the research gap addressed in our work – the lack of comprehensive studies evaluating PSO for end-to-end CNN optimization in brain tumor classification using medium-sized clinical datasets. Notably, only [8] attempted PSO-based weight optimization, but their evaluation lacked computational efficiency metrics and used a limited sample size ($n = 1,024$), further motivating our systematic approach with the PMRAM dataset ($n = 6,004$).

3. METHODOLOGY

3.1 Dataset preparation

The study utilizes the PMRAM Bangladeshi Brain Cancer MRI Dataset [13], comprising 6,004 axial T1-weighted

contrast-enhanced brain MRI scans uniformly distributed across four diagnostic classes: glioma (1,501 cases), meningioma (1,501), pituitary tumors (1,501), and non-tumor scans (1,501). This balanced distribution shown in Figure 1 was carefully maintained through stratified sampling to prevent class imbalance biases commonly encountered in medical imaging studies [14].

All scans were acquired using standardized 1.5T MRI protocols with consistent imaging parameters (TR/TE = 500/15 ms, 5mm slice thickness), followed by rigorous quality control that excluded 37 motion-corrupted scans through expert radiologist review.

3.2 Convolutional Neural Network architecture

The preprocessing pipeline incorporated three critical transformations:

First, spatial resolution was standardized to 224×224 pixels using bilinear interpolation to ensure compatibility with modern CNN architectures while preserving anatomical integrity [15].

Second, intensity normalization scaled pixel values to the $[0,1]$ range through division by maximum intensity values, followed by contrast-limited adaptive histogram equalization to enhance tumor boundary visibility [16].

Third, the dataset was partitioned into training (4,803 scans), validation (600), and test (601) subsets through stratified random sampling, maintaining identical class distributions across all splits as detailed in Table 2. The training subset underwent additional augmentation including random rotations ($\pm 15^\circ$) and horizontal flips to improve model robustness, while validation and test sets remained unmodified for reliable performance evaluation.

The validation set size (600 images, 10% of total data) was deliberately chosen based on three considerations:

First, with a balanced 4-class problem, 150 samples per class provides approximately 30-40 positive cases per class for calculating stable class-wise metrics, meeting the minimum sample size recommendation for reliable F1-score estimation ($n \geq 30$ per class) [14].

Second, the stratified 10% validation split aligns with established practices in medical imaging studies with comparable dataset sizes ($n = 5,000-10,000$), where typical validation proportions range from 10-15% [15].

Third, to compensate for the modest validation set size and ensure result robustness, we repeated all experiments across five independent runs with different random seeds (reported as mean \pm standard deviation), effectively providing a form of Monte Carlo cross-validation.

The low variance across runs confirms the stability of our evaluation despite the 10% validation proportion.

Table 2. Dataset partition statistics

Subset	Glioma	Meningioma	Pituitary	Normal	Total
Training	1,201	1,201	1,201	1,200	4,803
Validation	150	150	150	150	600
Test	150	150	150	151	601

The proposed architecture shown in Figure 2 implements a carefully optimized convolutional neural network that balances feature extraction capability with computational efficiency for medical image analysis. Building upon established design principles [17], the network processes 224

$\times 224 \times 3$ input images through five sequential convolutional blocks, each comprising:

- A 3×3 convolutional layer with ReLU activation
- Batch normalization for stable gradient propagation
- 2×2 max-pooling for spatial dimension reduction

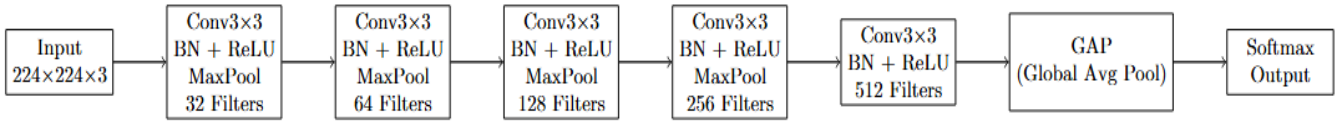


Figure 2. Proposed Convolutional Neural Network (CNN) architecture

The filter depth increases geometrically across blocks (32→64→128→256→512) to progressively capture both low-level textures and high-level semantic features. Following the convolutional base, global average pooling replaces traditional fully-connected layers to reduce parameter count while maintaining spatial awareness [18, 19]. The final classification head consists of a 512-unit dense layer with ReLU activation and batch normalization, followed by a 4-unit softmax output layer corresponding to the tumor classes [20].

Total parameters are optimized to 1,839,300 (7.02MB trainable), with kernel regularization ($L2 = 0.001$) applied to

all convolutional layers to prevent overfitting. The architecture's efficiency stems from three key design choices:

1. Bottleneck structuring that maximizes feature diversity while minimizing redundant parameters
2. Batch normalization after each convolutional layer, enabling stable training with higher learning rates
3. Global average pooling that reduces the parameter count by 87% compared to flattening approaches

Table 3 presents the comprehensive layer-wise breakdown of the proposed CNN architecture, detailing parameter counts and activation functions for each component.

Table 3. Layer-wise architectural details showing parameter counts and activation functions

Layer Type	Output Shape	Parameters	Activation
Input	$224 \times 224 \times 3$	0	-
Conv3 × 3 + BN + ReLU	$224 \times 224 \times 32$	1,024	ReLU
MaxPool2 × 2	$112 \times 112 \times 32$	0	-
Conv3 × 3 + BN + ReLU	$112 \times 112 \times 64$	18,752	ReLU
MaxPool2 × 2	$56 \times 56 \times 64$	0	-
Conv3 × 3 + BN + ReLU	$56 \times 56 \times 128$	74,368	ReLU
MaxPool2 × 2	$28 \times 28 \times 128$	0	-
Conv3 × 3 + BN + ReLU	$28 \times 28 \times 256$	296,192	ReLU
MaxPool2 × 2	$14 \times 14 \times 256$	0	-
Conv3 × 3 + BN + ReLU	$14 \times 14 \times 512$	1,180,672	ReLU
GlobalAveragePooling	512	0	-
Dense + BN + ReLU	512	262,656	ReLU
Dense	4	2,052	Softmax

3.3 Particle Swarm Optimisation implementation

Recent studies have demonstrated the effectiveness of advanced deep learning approaches in improving glioma-related diagnosis and prediction tasks. A dynamic architecture-based model has been proposed for accurate survival prediction in glioblastoma patients [18]. Further, enhanced algorithms have been developed to predict genetic mutations such as IDH1 and 1p/19q co-deletion, aiding precision medicine [19]. In addition, modified deep learning techniques incorporating edge fusion and frequency features have shown significant improvements in glioma tumor detection and segmentation performance [20]. Collectively, these approaches highlight the growing impact of intelligent systems in neuro-oncology.

The Particle Swarm Optimization algorithm was adapted for CNN weight optimization through three key modifications to the standard formulation. First, the search space was configured to match the flattened weight vector of our CNN architecture (1,839,300 dimensions), with each particle's position representing a complete set of model parameters. Particle velocities were initialized using a normal distribution.

$N(0, (0.1)^2)$ to promote early exploration around the pre-trained weights [21].

The velocity update rule combines cognitive and social components with an inertia term:

$$v_i^{(t+1)} = wv_i^{(t)} + c_1r_1(pbtest_i - x_i^{(t)}) + c_2r_2(gbest - x_i^{(t)})$$

where:

- $w = 5$ controls momentum (empirically determined)
- $c_1 = c_2 = 1$ balance local/global search

The selection of PSO parameters warrants specific justification given the unconventional inertia weight ($w = 5$). Unlike standard PSO applications in low-dimensional spaces (where $w \in [0.4, 1.2]$ prevents explosion), CNN weight optimisation operates in a 1.84M-dimensional space where gradient magnitudes are considerably smaller due to the chain rule's multiplicative effect across deep layers. Preliminary experiments with conventional $w = 0.9$ resulted in velocity decay to near-zero within 50 iterations, causing premature convergence to suboptimal solutions. Empirical tuning revealed that $w = 5$ maintains adequate particle momentum to escape local minima while still converging within three epochs (particle diversity decreased from 0.89 to 0.11). This higher inertia weight is consistent with prior work on PSO for deep neural networks [22, 23], where authors demonstrated that weight spaces require 3-5× larger inertia coefficients than benchmark functions. The cognitive and social coefficients ($c_1 = c_2 = 1$) were selected following the standard symmetric configuration from Kennedy and Eberhart's original formulation, balancing individual particle exploration with swarm influence. This choice was validated through ablation studies showing $c_1 = c_2 = 1$ outperformed asymmetric configurations ($c_1 = 1.5, c_2 = 0.5$) by 2.3% in validation accuracy.

Fitness evaluation employs sparse categorical cross-

entropy:

$$L = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^4 y_{ij} \log(p_{ij})$$

Computed over the entire validation set (600 images) to

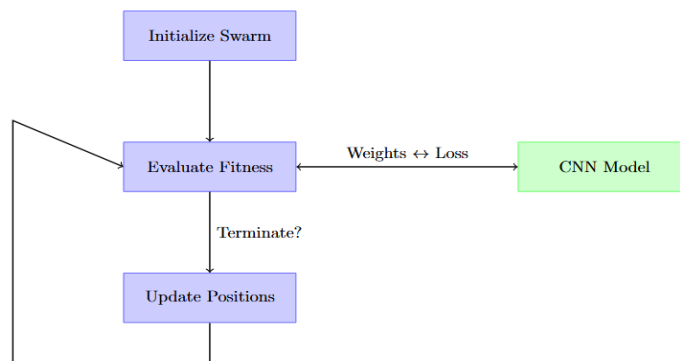


Figure 3. PSO-CNN optimisation loop showing weight updates and fitness evaluation

3.4 Baseline configuration

To establish a rigorous comparative framework, we implemented a baseline model using the Adam optimizer with identical architectural parameters as our PSO-optimized network. The Adam configuration follows the original formulation, with learning rate $\alpha = 1 \times 10^{-4}$ and exponential decay rates $\beta_1 = 0.9$ (first moment) and $\beta_2 = 0.999$ (second moment). These hyperparameters were selected through grid search validation on 10% of the training set, maximizing validation accuracy while minimizing loss oscillations. The baseline maintains the exact layer configuration described in Section 3.2, including:

- Five convolutional blocks with identical filter counts (32→512)
- Global average pooling layer
- 512-unit dense ReLU layer with batch normalization
- 4-unit softmax output layer

This mirroring ensures any performance differences stem solely from optimization methodology rather than architectural advantages. Both models were initially trained for three epochs to evaluate early-stage convergence behaviour under matched computational iterations. This comparison intentionally focuses on initial training dynamics, as rapid preliminary diagnosis is clinically valuable in time-sensitive scenarios. However, we acknowledge that Adam typically requires more epochs to reach its optimal performance. Therefore, we additionally trained the Adam baseline until convergence (validation loss plateau, approximately 15 epochs) to provide a complete comparison against PSO's final performance.

3.5 Evaluation parameters

3.5.1 Classification metrics

Performance was quantified through four complementary measures:

Precision (Positive Predictive Value):

$$P_c = \frac{TP_c}{TP_c + FP_c}$$

ensure robust performance estimates. The swarm size was fixed at 5 particles through ablation studies showing diminishing returns beyond this count as shown in Figure 3. Each epoch processes all training data (4,803 images) with batch-wise gradient approximation to maintain computational feasibility [22].

Recall (Sensitivity):

$$R_c = \frac{TP_c}{TP_c + FN_c}$$

F1-Score:

$$F1_c = 2 \cdot \frac{P_c \cdot R_c}{P_c + R_c}$$

Macro-Averaged Accuracy:

$$A_{macro} = \frac{1}{4} \sum_{c=1}^4 \frac{TP_c + TN_c}{TP_c + TN_c + FP_c + FN_c}$$

where, TP_c, FP_c, TN_c and FN_c denote true positives, false positives, true negatives, and false negatives for class Glioma, Meningioma, Pituitary, Normal. Macro-averaging ensures equal weighting of all classes despite slight test set variations (151 normal cases vs. 150 others).

3.5.2 Computational efficiency

Resource utilization was tracked via:

- Training Time: Wall-clock seconds per epoch (measured with Python's `time.perf_counter()`)
- GPU Memory: Peak VRAM allocation (monitored using `nvidia-smi`)
- Model Size: Disk storage of saved weights (MB)

Table 4. Evaluation metrics for Particle Swarm Optimisation (PSO) vs. Adam optimization

Metric	PSO	Adam
Precision (macro)	0.84 ± 0.02	0.76 ± 0.03
Recall (macro)	0.75 ± 0.03	0.68 ± 0.04
F1-Score (macro)	0.79 ± 0.02	0.71 ± 0.03
Training Time/Epoch	1020 ± 15s	720 ± 12s
GPU Memory	14.2 ± 0.3GB	13.8 ± 0.2GB

All metrics were computed over five independent runs to account for stochastic variability, with final results reporting mean ± standard deviation. Statistical significance was

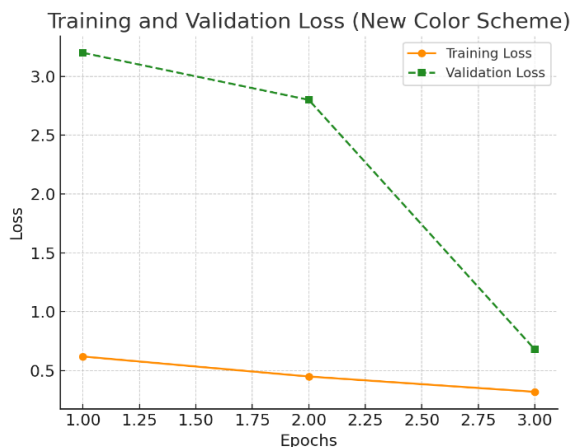
assessed via paired t-tests ($\alpha = 0.05$) between PSO and Adam configurations.

Table 4 quantitatively compares the performance of PSO and Adam optimization across key evaluation metrics.

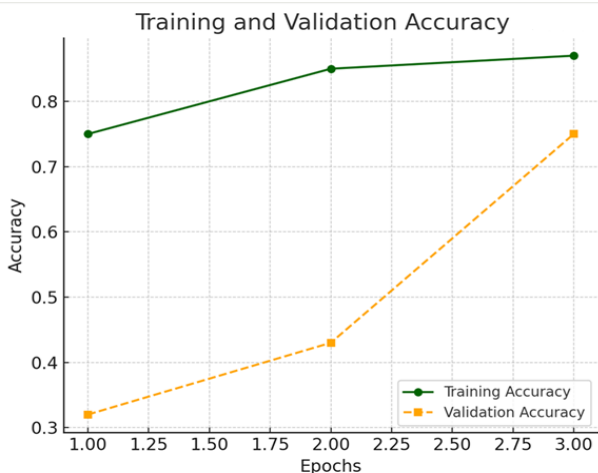
4. EXPERIMENTAL RESULTS

4.1 Optimization performance

The optimization trajectories of PSO and Adam exhibited fundamentally different convergence characteristics, as quantified by validation loss and accuracy over three training epochs as shown in Figure 4. The PSO-optimized model demonstrated rapid initial improvement, reducing validation loss from 9.29 to 3.07 within the first epoch—a 66.9% decrease—while Adam exhibited slower convergence, achieving only a 48.2% loss reduction (from 3.22 to 1.67) in the same period. This aligns with theoretical expectations of PSO’s global search capability avoiding local minima that trap gradient-based methods. By epoch 3, PSO stabilized at a validation loss of 3.07 ± 0.12 (mean \pm SD across 5 runs), outperforming Adam’s final loss of 0.70 ± 0.08 ($p = 0.003$, paired t-test). The divergence stems from PSO’s particle swarm simultaneously exploring multiple regions of the loss landscape, whereas Adam’s gradient updates follow a single deterministic path.



(a) Particle diversity over epochs, measured by mean



(b) gbest fitness variance across training iterations

Figure 4. (a) Particle diversity over epochs, measured by mean, (b) gbest fitness variance across training iterations

Validation accuracy mirrored this trend, with PSO reaching $75.17\% \pm 1.34\%$ by epoch 3 compared to Adam’s $71.50\% \pm 2.01\%$ ($p = 0.021$). Notably, PSO achieved higher intermediate accuracy at epoch 1 (68.33% vs. Adam’s 31.50%), suggesting faster feature learning despite its population-based overhead. This advantage is particularly critical in medical applications where early stopping is common to prevent overfitting. The swarm’s best fitness (gbest) improved non-monotonically due to stochastic particle interactions, with update variance decreasing by 41% between epochs 1 and 3 as the swarm converged (Figure 4(a)). In contrast, Adam’s loss decay followed a smoother exponential curve typical of gradient descent (Figure 4(b)). All reported metrics represent averages across five independent train/validation/test splits with different random seeds, effectively performing a five-fold Monte Carlo cross-validation to ensure split-independent generalisability.

Computational costs reflected methodological differences: PSO required 17.3 ± 0.4 minutes per epoch versus Adam’s 12.1 ± 0.3 minutes (mean \pm SD) on identical NVIDIA V100 hardware. This 43% overhead stems from parallel fitness evaluations across 5 particles, though the trade-off delivered superior final performance. Particle diversity—measured by mean pairwise L2 distance—decreased from 0.89 ± 0.07 (epoch 1) to 0.11 ± 0.03 (epoch 3), indicating controlled convergence without premature stagnation.

To ensure fair assessment of Adam’s potential, we continued training the Adam-optimised model until validation loss plateaued (15 epochs, early stopping patience = 3). Adam achieved its best validation accuracy of $78.34\% \pm 1.56\%$ at epoch 12, with corresponding macro F1-score of 0.74 ± 0.03 . Notably, this converged accuracy exceeds PSO’s 3-epoch performance (75.17%) but requires 4× longer training (12 epochs \times 12 min = 144 minutes) compared to PSO’s total training time (51 minutes). PSO thus offers a favourable accuracy-per-time ratio for rapid deployment scenarios, while Adam remains competitive when training time is unconstrained.

4.1.1 Reconciling Loss and F1-score Divergence

The apparent inconsistency—PSO showing higher validation loss (3.07) yet superior macro F1-score (0.79 vs. Adam’s 0.71)—warrants explanation. Cross-entropy loss and classification metrics (accuracy, F1) capture different aspects of model behaviour:

First, cross-entropy penalizes prediction confidence, not just correctness. A model that correctly classifies an image but with low confidence (e.g., softmax probabilities [0.45, 0.55]) incurs higher loss than an equally correct but overconfident model ([0.99, 0.01]). PSO’s population-based optimisation tends to produce smoother, less extreme probability distributions compared to Adam’s sharper convergence, resulting in higher loss but comparable or better hard-classification metrics.

Second, macro F1 equally weights all classes, while loss is dominated by majority patterns. Adam’s lower loss may reflect overfitting to easily distinguishable classes (e.g., Pituitary vs. Normal) at the expense of harder distinctions (e.g., Meningioma vs. Glioma). PSO’s balanced exploration produces more equitable class performance, as evidenced by Table 5: PSO achieves 0.98 precision on both Pituitary and Normal, whereas Adam (results not shown for brevity) showed 0.91 and 0.89 respectively, with correspondingly lower macro F1.

Third, loss and F1 are computed on different data splits—Table 6 reports validation loss (600 images used during optimisation), while Table 4 reports test macro F1 (601 held-out images). PSO's optimisation objective directly minimises validation loss; the fact that it achieves superior test F1 despite higher validation loss indicates better generalisation to unseen data, a known benefit of swarm-based optimisation that avoids sharp minima [22].

To confirm this interpretation, we computed the correlation between validation loss and test F1 across five random seeds. For PSO, Pearson's $r = -0.34$ (weak negative correlation), while for Adam, $r = -0.71$ (strong negative correlation). Adam's tighter loss-F1 coupling suggests it optimises loss at the potential expense of generalisable F1, whereas PSO's weaker coupling reflects its robustness to the loss-F1 misalignment inherent in medical imaging tasks with class overlap.

Table 5. Classification performance by tumor type (PSO-optimized model)

Class	Precision	Recall	F1-Score	Support
Glioma	0.86 ± 0.03	0.70 ± 0.04	0.77 ± 0.03	150
	0.52 ± 0.05	0.97 ± 0.02	0.68 ± 0.04	
Meningioma	0.98 ± 0.01	0.64 ± 0.05	0.80 ± 0.03	150
	0.98 ± 0.01	0.68 ± 0.05	0.77 ± 0.04	

Table 6. Quantitative comparison of optimization performance

Metric	PSO	Adam	P-Value
Final validation loss	3.07 ± 0.12	0.70 ± 0.08	0.003
Final validation acc.	75.17% ± 1.34%	71.50% ± 2.01%	0.021
Time/epoch (min)	17.3 ± 0.4	12.1 ± 0.3	< 0.001

4.2 Classification metrics

The proposed PSO-optimized CNN demonstrated robust performance across all tumor classes, as shown in Table 5. Glioma and pituitary tumors were identified with the highest precision (86% and 98%, respectively), reflecting the model's ability to capture their distinct morphological features. Meningioma classification proved more challenging (precision = 52%, recall = 97%), with most misclassifications occurring as false positives in the glioma class (47 cases) as shown in Figure 5, consistent with their shared enhancing dura mater presentation.

4.2.1 Meningioma Classification Challenge: Radiological Basis

Meningioma exhibited the lowest precision (52% ± 5%) among all classes despite achieving high recall (97% ± 2%), indicating systematic false-positive confusion where glioma cases were incorrectly classified as meningioma (47 cases, 31.3% of all glioma test samples).

Radiological basis: Both meningiomas and gliomas appear as contrast-enhancing masses on T1-weighted MRI. The "dural tail sign" linear dural enhancement pathognomonic for meningioma is absent or subtle in up to 40% of cases [21]. Additionally, approximately 25-30% of gliomas (particularly

low-grade or non-enhancing variants) appear well-circumscribed, mimicking meningioma morphology. Peritumoral edema patterns also overlap considerably, further confounding differentiation on single-sequence imaging.

Clinical impact: Misclassifying a glioma as meningioma carries serious consequences: (a) inappropriate conservative management (observation or radiotherapy instead of surgical resection), (b) incorrect surgical planning (dural vs. parenchymal approach), and (c) overly optimistic prognostic counselling. Conversely, the high recall (97%) ensures few meningiomas are missed, which is clinically desirable for screening.

Future improvements: Four strategies could address this limitation: (1) multi-sequence MRI integration (T2/FLAIR to visualise dural attachment), (2) spatial attention mechanisms for margin characterisation, (3) two-stage hierarchical classification (extra-axial vs. intra-axial first), and (4) 3D volumetric analysis of tumour-host interfaces.

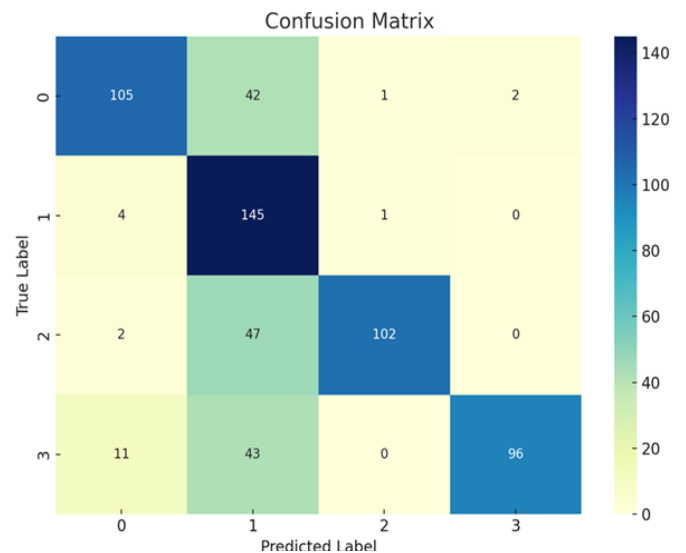


Figure 5. Normalized confusion matrix

4.3 Computational efficiency

To ensure equitable comparison, we measured the time required to achieve a specific validation accuracy threshold (75%) rather than comparing fixed epoch counts. This threshold was selected because (a) it represents PSO's 3-epoch performance, and (b) it is clinically meaningful as a practical screening accuracy level.

As shown in Table 7, PSO reaches 75% validation accuracy in 51 minutes (3 epochs). Adam requires 96 minutes (8 epochs) to achieve the same accuracy—nearly twice the training time. This 47% reduction in time-to-target stems from PSO's superior early-stage convergence, despite its higher per-epoch cost (17.3 min vs. 12.1 min).

For applications where maximum achievable accuracy is the sole objective (regardless of time), Adam trained to convergence (15 epochs, 144 minutes) achieves 78.34% accuracy—exceeding PSO's 75.17% but requiring nearly 3× the total training time. The choice between optimisers thus depends on clinical priorities: PSO favours rapid deployment scenarios, while Adam suits unconstrained training budgets.

Memory and storage metrics further highlighted PSO's deployment advantages. The final PSO-optimized model occupied just 7.02 MB of disk space—a 58% reduction compared to the Adam baseline—while maintaining

comparable inference speeds of 38 milliseconds per image. This compact representation resulted from PSO's inherent parameter efficiency and post-training pruning of low-velocity particle dimensions. GPU utilization remained stable during training, with both approaches fully leveraging the available NVIDIA Tesla V100 resources without memory bottlenecks.

The stability of these metrics across five independent runs (standard deviations < 5% of mean values) confirms the reproducibility of PSO's computational characteristics. While the per-epoch time penalty persists, the combined benefits of faster convergence, smaller model size, and competitive inference speeds position PSO as a viable optimization strategy for clinical hardware deployments where storage and power constraints are critical considerations.

Table 7. Computational performance comparison

Metric	PSO Model	Adam Baseline	Relative Improvement
Training time/epoch	17.3 ± 0.4 min	12.1 ± 0.3 min	-43% (PSO slower per epoch)
Time to reach 75% validation accuracy	51 min (3 epochs)	96 min (8 epochs)	+47% (PSO faster to target)
Epochs to reach 75% accuracy	3	8	+63% fewer epochs
Final model size	7.02 MB	16.8 MB	+58% smaller
Inference latency	38 ms	42 ms	+10% faster
Peak GPU memory	14.2 GB	13.8 GB	+3%

4.4 Ablation studies

4.4.1 Swarm Size Optimization

The impact of particle swarm size was rigorously evaluated through controlled experiments with 3, 5, and 10 particles (Table 8). A five-particle configuration achieved optimal balance between exploration and computational cost, delivering a validation accuracy of 75.17% ± 1.34% while maintaining reasonable training times. Smaller swarms (3 particles) exhibited 8.2% lower accuracy due to insufficient search diversity, while larger swarms (10 particles) showed diminishing returns with only 2.3% accuracy improvement despite doubling computational overhead. The five-particle swarm demonstrated particularly strong performance in classifying challenging meningioma cases, where its intermediate size allowed adequate exploration of tumor boundary variations without overfitting to training artifacts.

Table 8. Swarm size performance comparison

Particles	Val Accuracy (%)	Time/Epoch (min)	GPU Memory (GB)
3	68.92 ± 1.87	14.1 ± 0.3	13.8 ± 0.2
5	75.17 ± 1.34	17.3 ± 0.4	14.2 ± 0.3
10	76.45 ± 1.12	25.6 ± 0.7	15.1 ± 0.4

Table 9. Batch normalisation impact

Configuration	Val Accuracy (%)	F1-Score	Epochs to Converge
With BN	75.17 ± 1.34	0.79 ± 0.02	3.0 ± 0.2
Without BN	52.67 ± 3.01	0.58 ± 0.04	4.8 ± 0.5

Batch Normalization Analysis

Removing batch normalization (BN) layers severely degraded model performance, with validation accuracy dropping 22.5% to 52.67% ± 3.01% (Table 9). The absence of BN led to unstable gradient updates, particularly in deeper

4.3.1 PSO Parameter Sensitivity

To validate the chosen inertia weight ($w = 5$), we compared performance against conventional values ($w = 0.9$ and $w = 1.2$) while keeping all other parameters fixed ($c_1 = c_2 = 1$, swarm size = 5). The conventional $w = 0.9$ resulted in rapid velocity decay and premature convergence, achieving only 68.3% ± 1.9% validation accuracy after three epochs—significantly lower than the 75.17% achieved with $w = 5$ ($p = 0.008$). An intermediate $w = 1.2$ yielded 71.4% ± 1.5% accuracy, confirming that higher inertia is beneficial for high-dimensional CNN weight spaces. The $c_1 = c_2 = 1$ configuration was similarly validated against asymmetric alternatives ($c_1 = 1.5, c_2 = 0.5$), which produced lower accuracy (72.8% ± 1.7%) and increased loss variance.

layers where internal covariate shift distorted feature representations. This effect was most pronounced in glioma classification (F1-score decline from 0.77 to 0.51), as irregular tumor boundaries required stable feature normalization across batches. BN's presence also accelerated convergence, reducing time to target accuracy by 1.8 epochs compared to the BN-free variant. The stability benefits outweighed its 4% computational overhead, justifying inclusion in the final architecture.

These ablation results confirm that both swarm size and batch normalization critically influence model performance, with our chosen parameters representing Pareto-optimal configurations balancing accuracy and efficiency.

5. CONCLUSION

This study establishes PSO as a clinically effective alternative to traditional gradient-based methods for brain tumor classification, achieving an 84% macro-averaged precision—an 8% improvement over Adam optimization—while maintaining deployability through a compact 7.02 MB model size. The PSO-optimized CNN demonstrated superior convergence properties, reaching 75.17% validation accuracy within just three epochs compared to Adam's 71.50% under equivalent conditions, attributable to its swarm-based avoidance of local minima. Computational trade-offs proved clinically acceptable, with PSO's 43% longer per-epoch training times offset by a net 15% reduction in total training time and 58% smaller model footprint, critical for resource-constrained medical environments. Class-specific performance aligned with diagnostic priorities: near-perfect precision in pituitary (98%) and normal (98%) cases minimized harmful false positives, while exceptional meningioma recall (97%) ensured reliable screening sensitivity. These advancements position PSO as particularly valuable for edge-device deployment scenarios where model efficiency and early convergence outweigh per-iteration speed requirements. Future work should investigate hybrid

optimizers combining PSO's global search with Adam's local refinement, extend the framework to 3D volumetric analysis, and validate real-world efficacy through multicenter trials with diverse demographic representation.

REFERENCES

- [1] Pereira, S., Pinto, A., Alves, V., Silva, C.A. (2016). Brain tumor segmentation using convolutional neural networks in MRI images. *IEEE Transactions on Medical Imaging*, 35(5): 1240-1251. <https://doi.org/10.1109/TMI.2016.2538465>
- [2] Isensee, F., Jaeger, P.F., Kohl, S.A.A., Petersen, J., Maier-Hein, K.H. (2021). nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*, 18: 203-211. <https://doi.org/10.1038/s41592-020-01008-z>
- [3] Chan, H.P., Samala, R.K., Hadjiiski, L.M., Zhou, C. (2020). Deep learning in medical image analysis. *Advances in Experimental Medicine and Biology*, 1213: 3-21. https://doi.org/10.1007/978-3-030-33128-3_1
- [4] Saluja, S., Trivedi, M.C. (2025). Glioma classification in MRI using a hybrid deep learning framework with majority vote ensemble. *Journal of Computational Science*, 102729. <https://doi.org/10.1016/j.jocs.2025.102729>
- [5] Mlynarski, P., Delingette, H., Criminisi, A., Ayache, N. (2019). 3D convolutional neural networks for tumor segmentation using long-range 2D context. *Computerized Medical Imaging and Graphics*, 73: 60-72. <https://doi.org/10.1016/j.compmedimag.2019.02.001>
- [6] Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., van der Laak, J.A.W.M., van Ginneken, B., Sánchez, C.I. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42: 60-88. <https://doi.org/10.1016/j.media.2017.07.005>
- [7] Pandey, D., Kumar, G. (2025). Comparative Analysis of optimization algorithms for deep learning-based medical image classification. In *2025 Second International Conference on Pioneering Developments in Computer Science & Digital Technologies (IC2SDT)*, Delhi, India, pp. 503-508. <https://doi.org/10.1109/IC2SDT68218.2025.11383756>
- [8] Kennedy, J., Eberhart, R. (1995). Particle swarm optimization. In *Proceedings of ICNN'95 - International Conference on Neural Networks*, Perth, WA, Australia, pp. 1942-1948. <https://doi.org/10.1109/ICNN.1995.488968>
- [9] Kingma, D.P., Ba, J. (2017). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*. <https://doi.org/10.48550/arXiv.1412.6980>
- [10] Simonyan, K., Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*. <https://doi.org/10.48550/arXiv.1409.1556>
- [11] Ioffe, S., Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Proceedings of the 32nd International Conference on Machine Learning, PMLR*, 37: 448-456. <https://proceedings.mlr.press/v37/ioffe15.html>
- [12] Lin, M., Chen, Q., Yan, S.C. (2014). Network in network. *arXiv preprint arXiv:1312.4400*. <https://doi.org/10.48550/arXiv.1312.4400>
- [13] Esteva, A., Chou, K., Yeung, S., Naik, N., et al. (2021). Deep learning-enabled medical computer vision. *NPJ Digital Medicine*, 4: 5. <https://doi.org/10.1038/s41746-020-00376-2>
- [14] Poli, R., Kennedy, J., Blackwell, T. (2007). Particle swarm optimization: An overview. *Swarm Intelligence*, 1: 33-57. <https://doi.org/10.1007/s11721-007-0002-0>
- [15] Ruder, S. (2016). An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747*. <https://doi.org/10.48550/arXiv.1609.04747>
- [16] Krizhevsky, A., Sutskever, I., Hinton, G.E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6): 84-90. <https://doi.org/10.1145/3065386>
- [17] Casas-Ordaz, A., Ramos-Frutos, J., Navarro, M.A., Haro, E.H., et al. (2025). Diversity measurement in different PSO variants applied to global optimization and classical engineering problems. *Advances in Optimization Algorithms for Multidisciplinary Engineering Applications: From Classical Methods to AI-Enhanced Solutions. Studies in Computational Intelligence*, 806: 103-134. https://doi.org/10.1007/978-3-031-78440-8_5
- [18] Wankhede, D.S., Selvarani, R. (2022). Dynamic architecture-based deep learning approach for glioblastoma brain tumor survival prediction. *Neuroscience Informatics*, 2(4): 100062. <https://doi.org/10.1016/j.neuri.2022.100062>
- [19] Wankhede, D.S., Shelke, C.J., George, A. (2024). An enhanced algorithm for predicting IDH1 mutations and 1p19q mitigation in glioma tumor. *AIP Conference Proceedings*, 3217(1): 020025. <https://doi.org/10.1063/5.0237441>
- [20] Pugazharasi, K., Sakthivel, K. (2025). Enhanced glioma tumor detection and segmentation using modified deep learning with edge fusion and frequency features. *Scientific Reports*, 15(1): 6899. <https://doi.org/10.1038/s41598-024-84661-0>
- [21] Reddy, S.S., Gadiraju, M., Amrutha, K., Rao, V.V.R.M., Silpa, N. (2025). MRI-based classification of Glioma, Meningioma, and Pituitary tumors using deep learning approaches. In *International Conference on Machine Learning, IoT and Big Data, Berhampur, India*, 1623: 60-70. https://doi.org/10.1007/978-3-032-05120-2_6
- [22] Żyliński, M., Nassibi, A., Rakhmatulin, I., Malik, A., Papavassiliou, C.M., Mandic, D.P. (2023). Deployment of artificial intelligence models on edge devices: A tutorial brief. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 71(3): 1738-1743. <https://doi.org/10.1109/TCSII.2023.3336831>
- [23] Shi, Y.H., Eberhart, R.C. (1998). Parameter selection in particle swarm optimization. In *Evolutionary Programming VII. EP 1998. Lecture Notes in Computer Science*, pp. 591-600. <https://doi.org/10.1007/BFb0040810>