Brain Tumor Dataset

Phase 1:

1. **Select and justify a challenging dataset.**

   The dataset contains MRI images of brain tumors categorized into four classes: **Glioma, Meningioma, Pituitary, and No Tumor.** The training set and test set contain separate folders for each class.

   The dataset contains a total of 7,025 MRI images, organized into separate training and testing directories, with each class stored in an individual folder. This structure supports supervised learning and ensures a clear separation between training and evaluation data.

   The dataset was obtained from Kaggle: Dataset link

   Image Size: All images were resized to 224×224 pixels for uniformity.

   Data Augmentation: Applied to the training set:

   - Random rotations (±15°)
   - Width and height shifts (±5%)
   - Zooming (±10%)
   - Horizontal flipping

   Normalization: Pixel values were rescaled to the [0,1] range.

   Validation Split: 20% of training data used for validation.

   Parameter space:

   "lr": [1e-3, 1e-4, 5e-5],

"dense_units": [128, 256, 512],

"dropout": [0.3, 0.4, 0.5]

2. **Build baseline deep learning model**

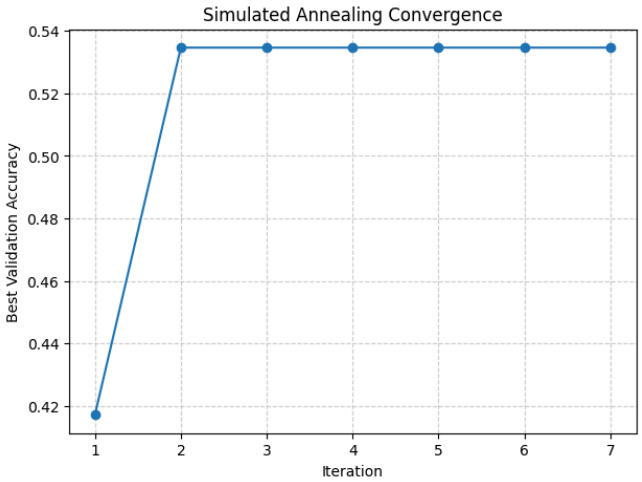
   A baseline deep learning model was constructed using **MobileNetV2**, chosen for its efficiency and strong performance on image classification tasks, particularly when working with limited computational resources.

   - **Model Architecture:** MobileNetV2 (transfer learning)
   - **Number of Epochs:** 3
   - **Number of Iterations:** 7

Apply 4 metaheuristic algorithms for model optimization.

1. Stimulated Annealing (SA)

Convergence over iterations

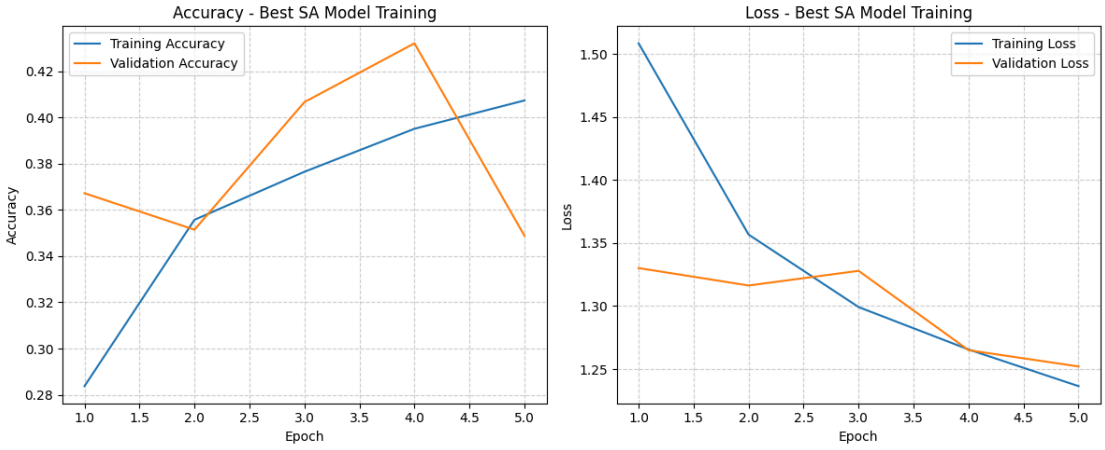


The algorithm found its "best" solution almost immediately (at Iteration 2) and failed to improve further.

Stagnation: The flat line from Iteration 2 to 7 indicates the algorithm got stuck in a local optimum. It stopped exploring the search space effectively.

Optimization Issue: This suggests the "cooling schedule" (how fast the algorithm settles) was likely too aggressive, preventing it from searching for better hyperparameter combinations.

Accuracy and Loss over epoch



Max Validation Accuracy: 0.4320771396160126

Min Validation Loss: 1.2521061897277832

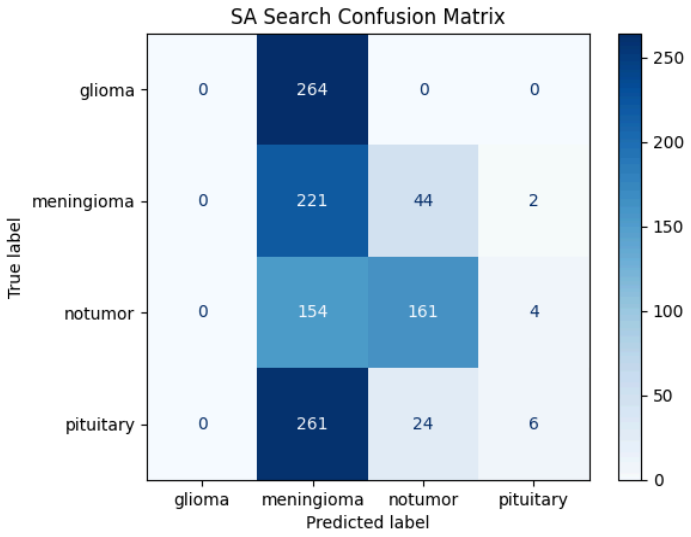
Model Instability: While training accuracy improves steadily (blue line), the validation accuracy (orange line) is volatile, crashing significantly after Epoch 4. This indicates the model is not stable.

Underfitting: The validation loss flattens out around 1.25, which is still quite high. This suggests the model does not yet have the capacity to capture the complex features of the MRI scans.

Peak Performance: The best performance occurred at Epoch 4 (approx 43% accuracy), after which the model began to degrade, suggesting that

training for more epochs without changing the learning rate might be detrimental.

Confusion Matrix



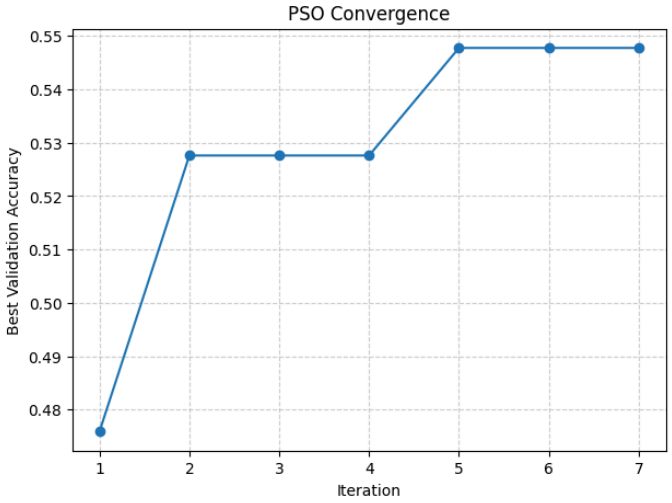
Critical Bias: The model suffers from severe "Mode Collapse." It is predicting **Meningioma** for the vast majority of cases (the dark blue squares are all in the Meningioma column).

Class Failure: The model completely failed to learn the features of **Glioma** (0 correctly identified) and **Pituitary** tumors (only 6 correctly identified).

False Accuracy: The overall accuracy of ~43% is misleading; the model is not actually "learning" to differentiate tumors. It is simply guessing "Meningioma" or "No Tumor" because those are the safest bets to minimize loss, ignoring the other two classes entirely.

2. Particle Swarm (PSO)

Convergence over iterations

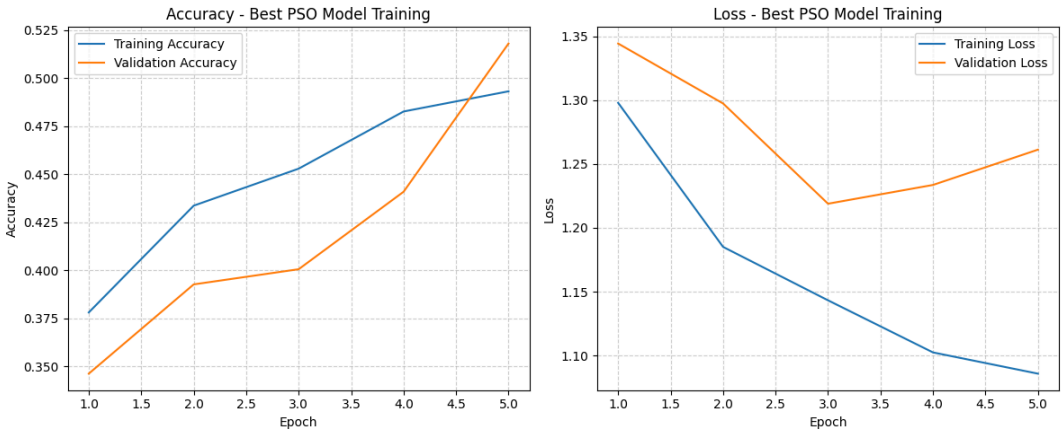


Active Exploration: Unlike the SA algorithm which flatlined immediately, the PSO shows a "step-wise" improvement. It found a good solution at Iteration 2, searched for a while, and then found an even better solution at Iteration 5.

Better Search Capabilities: This indicates that the swarm behavior (particles communicating with each other) helped the algorithm break out of local optima effectively.

Stronger Peak: The final convergence accuracy reaches nearly 55% on the graph (higher than SA's ~53%), showing it found a superior set of hyperparameters.

Accuracy and Loss over epoch

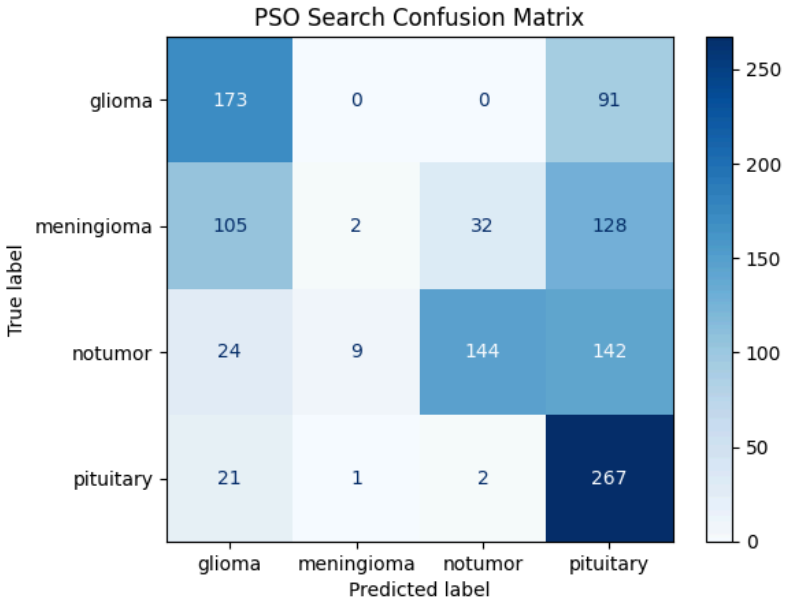


Max Validation Accuracy: 0.5179666876792908

Min Validation Loss: 1.2188769578933716

Strong Late-Stage Learning: The orange line (Validation Accuracy) shoots up dramatically at Epoch 5. This is a very positive sign; it suggests the model was just starting to "get it" right when training stopped.

Confusion Matrix



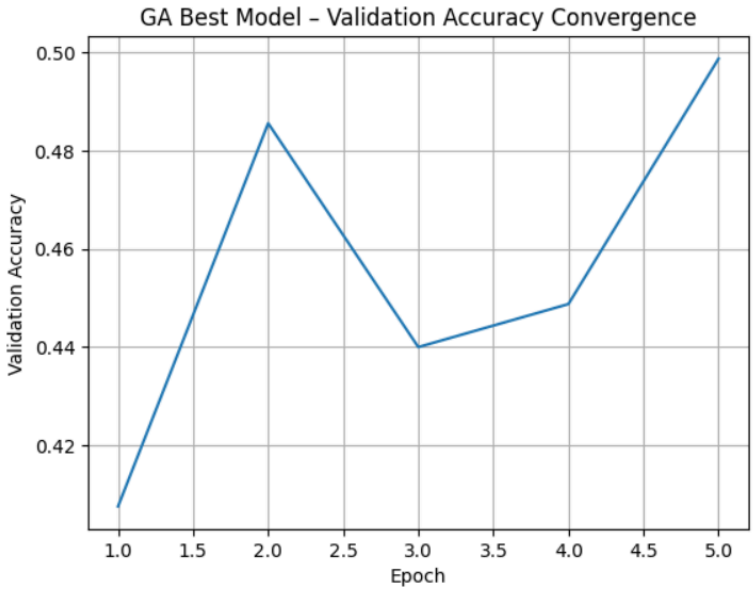
The "Meningioma" Problem: This is the model's biggest failure. It correctly identified only 2 Meningioma cases. It confuses them heavily with Gliomas (105 misclassified) and Pituitary tumors (128 misclassified).

Bias Shift: In the SA experiment, the model *only* predicted Meningioma. In this PSO experiment, it has swung the other way and now over-predicts Pituitary tumors (note the high numbers in the "Pituitary" column, particularly the 142 Notumor cases misclassified as Pituitary).

Genuine Learning: On the positive side, it learned Glioma (173 correct) and Pituitary (267 correct) quite well. This proves the model has the *capacity* to learn, but the class imbalance or feature similarity between Meningioma and others is confusing it.

3. Genetic Algorithm

Convergence over iterations

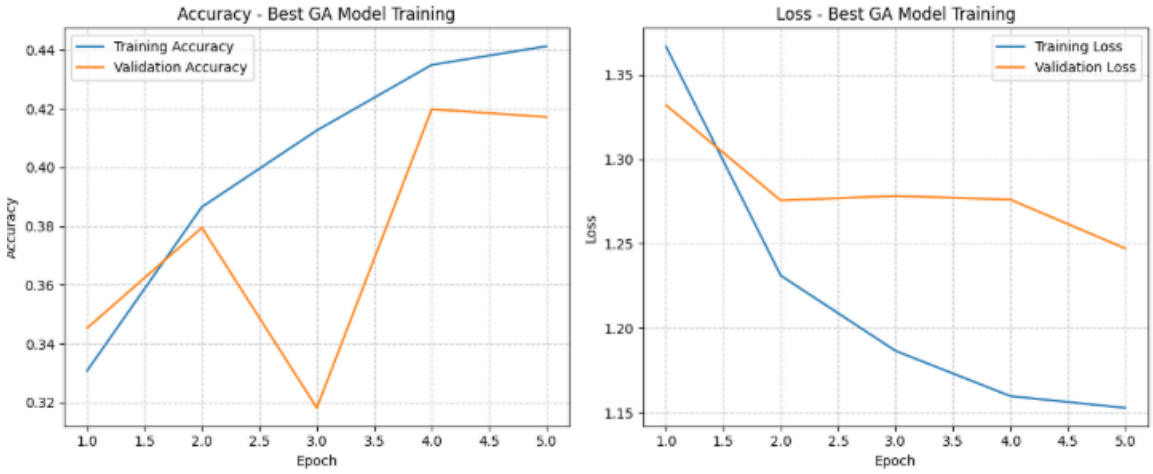


High Instability: Unlike the smooth learning curves in the PSO experiment, this graph is extremely jagged (zigzag pattern). The accuracy drops significantly at Epoch 3 before spiking at Epoch 4.

Unreliable Training: This volatility suggests the hyperparameters found by the GA result in a model with an unstable learning rate—it is "overshooting" the optimal weights during training.

Note on Graph Type: Unlike the previous "Convergence" graphs which showed improvement over Iterations (Generations), this graph appears to show the Epochs of the best model found. It indicates that even the "best" individual the GA found struggles to learn smoothly

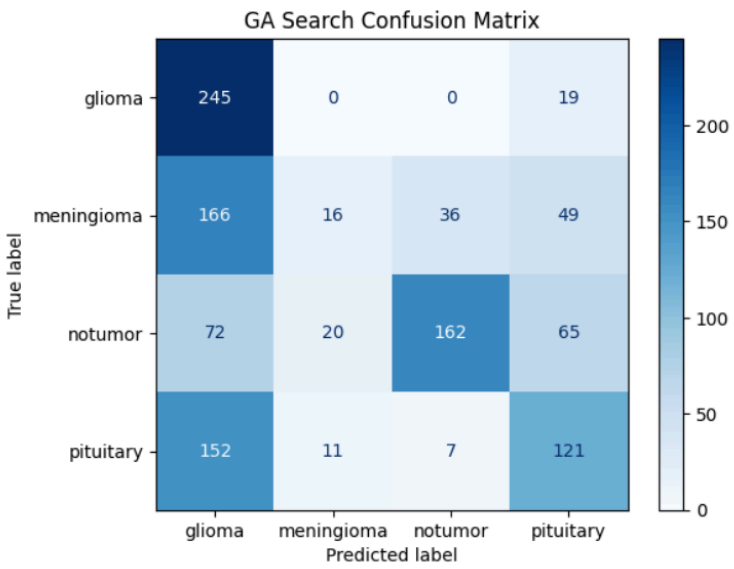
Accuracy and Loss over epoch



-

Stagnant Generalization: While the **Training Loss** (blue line on the right) goes down consistently, the **Validation Loss** (orange line) is almost completely flat after Epoch 2.
Failure to Learn Features: A flat validation loss means the model is not improving its understanding of unseen data at all. It effectively stopped learning meaningful patterns after the second epoch.
Overfitting Gap: By Epoch 4 and 5, the gap between training accuracy (improving) and validation accuracy (dropping/flat) widens, confirming the model is memorizing the training data rather than understanding the tumor types.

Confusion Matrix



Specific Failures:

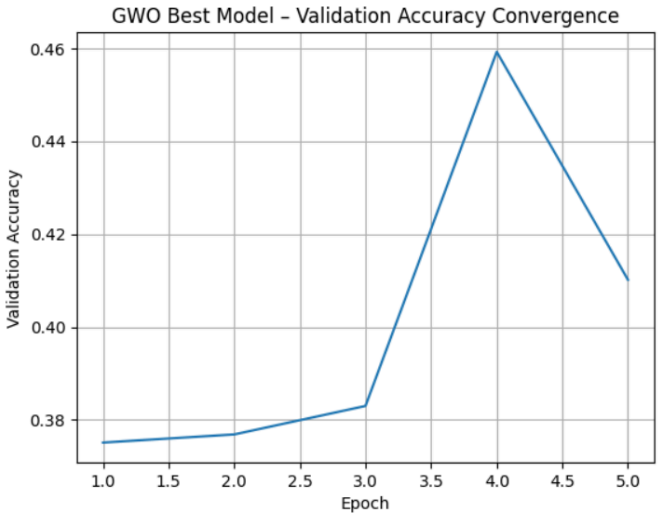
It misclassified **166 Meningiomas** as Gliomas.

It misclassified **152 Pituitary tumors** as Gliomas.

4. Gray Wolf

4.1 Model optimization

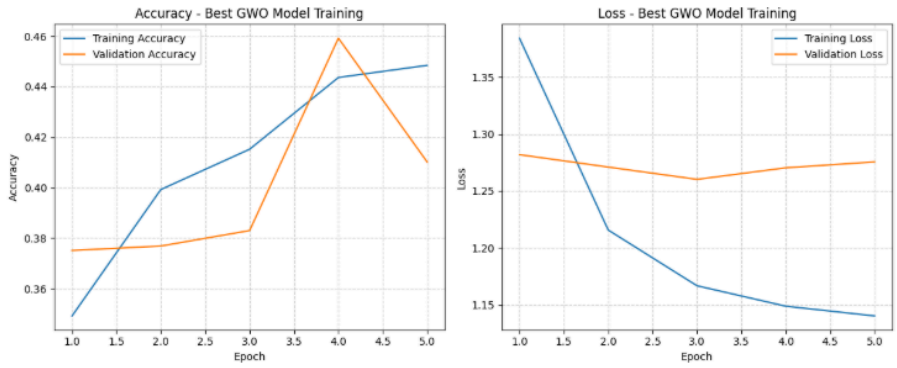
Convergence over iterations



Key Observation: The accuracy remains relatively low (~38%) for the first three epochs, spikes dramatically to its peak of ~45.9% at Epoch 4, and then drops sharply again by Epoch 5.

Interpretation: This behavior indicates instability during training. The model found a temporary "sweet spot" at Epoch 4 but failed to sustain that performance, suggesting the optimization algorithm (GWO) struggled to converge on a stable solution.

Accuracy and Loss over epoch



Max Validation Accuracy: 0.45924627780914307

Min Validation Loss: 1.2601633071899414

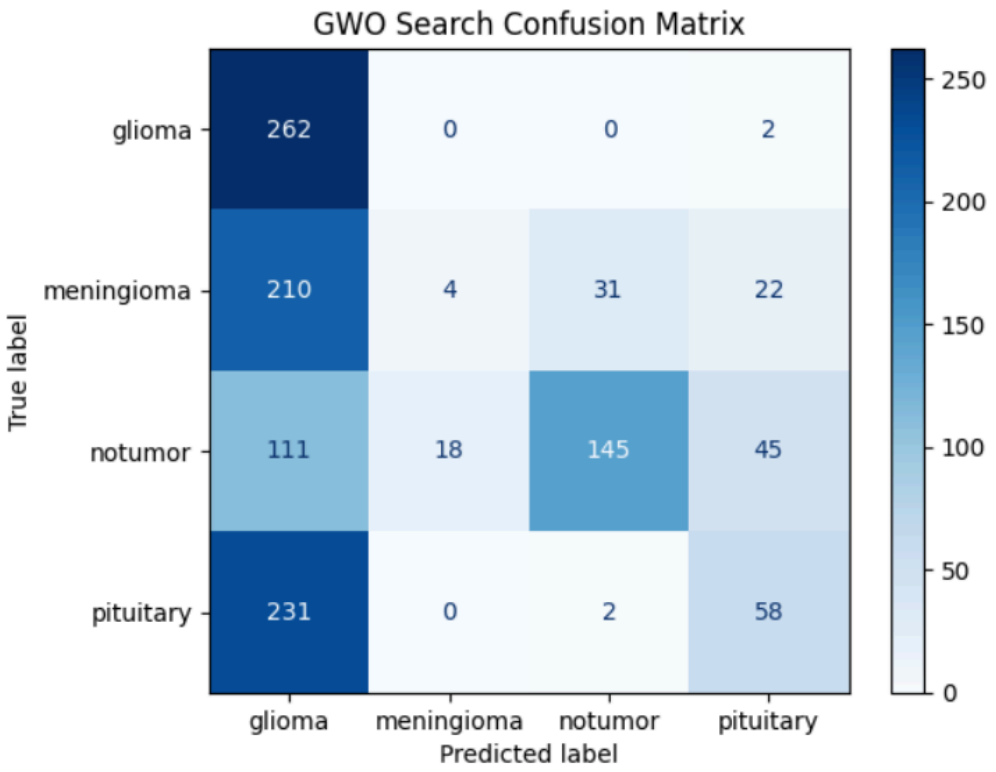
Training Accuracy: Steadily increases, showing the model is memorizing the training data well.

Validation Accuracy: Fluctuates significantly, peaking at Epoch 4 before falling.

Training Loss: Drops consistently, which is expected.

Validation Loss: Remains high and relatively flat (around 1.26–1.28).

Confusion Matrix



Severe Class Bias: The model suffered from a major failure in feature learning, predominantly predicting "Glioma" for all inputs.

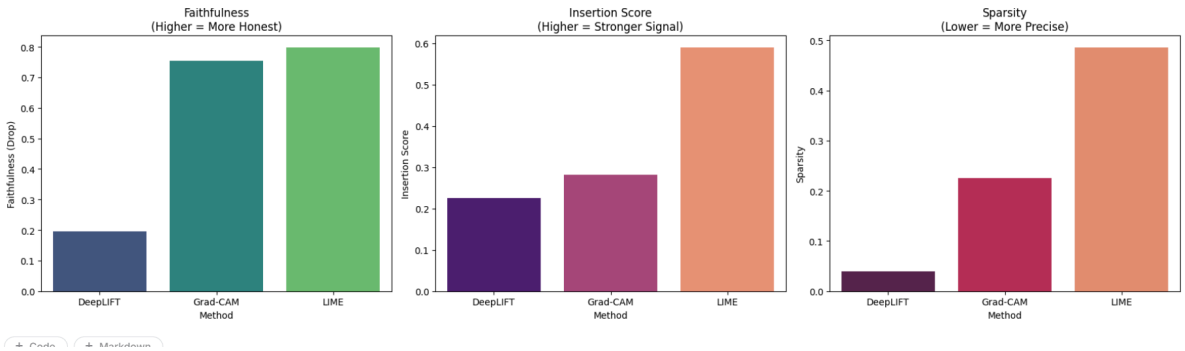
Misclassification: It failed to identify *Pituitary* (0 correct) and *Meningioma* (only 4 correct) tumors, misclassifying nearly all of them as Glioma.

Conclusion: GWO failed to generalize effectively, resulting in a model that mimics "mode collapse."

4.2 XAI Parameter optimization

Applied on DeepLIFT, Grad-cam,And Lime

Method	Faithfulness (Drop)	Insertion Score	Sparsity
DeepLIFT	0.196078	0.225534	0.039021
Grad-CAM	0.754962	0.281332	0.225161
LIME	0.797095	0.588902	0.485280



5. FireFly

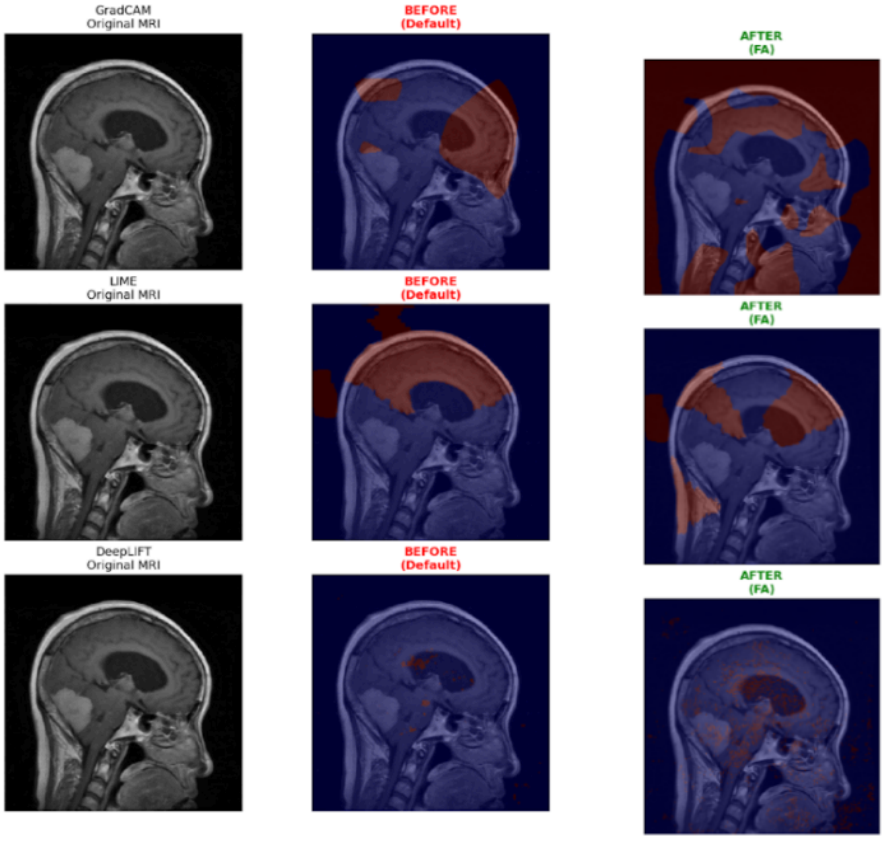
5.1 Model optimization

Accuracy and Loss over epoch

Max Validation Accuracy: 0.4829097390174866

Min Validation Loss: 1.233982801437378

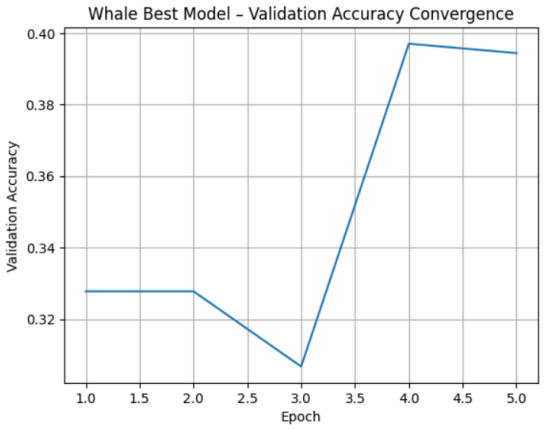
5.2 XAI Parameter optimization



6. Whale Optimization

6.1 Model optimization

Convergence over iterations

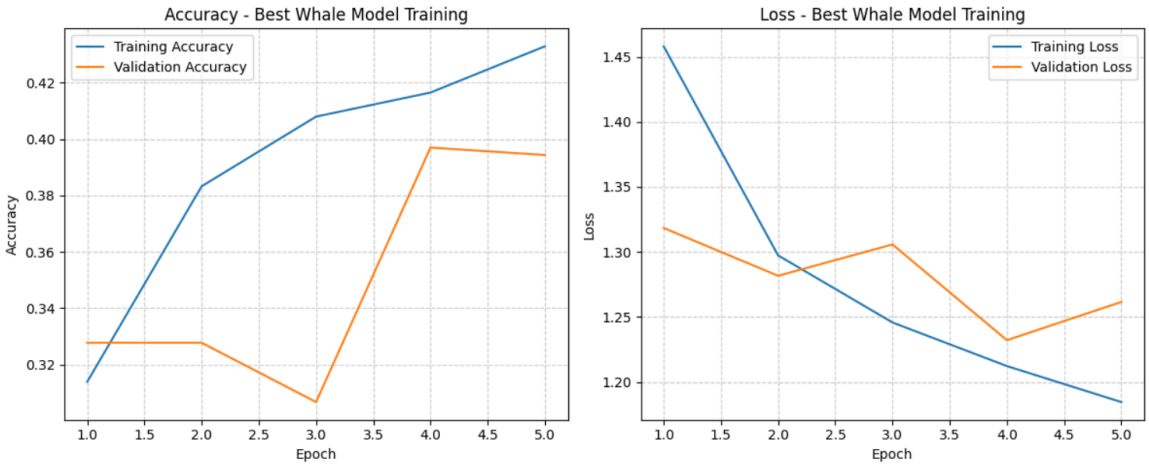


Labeling Anomaly: The graph is titled "Validation Accuracy Convergence" but the x-axis is labeled "Epoch" (1 to 5). This suggests it might be displaying the training progress of the best model rather than the optimization over iterations.

High Volatility: The line is extremely unstable. It stays flat for two epochs, drops significantly at Epoch 3 (to ~30%), spikes at Epoch 4 (to ~39%), and flattens again.

Search Instability: This erratic behavior indicates the algorithm (or the model's learning rate) was too aggressive, causing it to "forget" what it learned at Epoch 3 before recovering slightly.

Accuracy and Loss over epoch

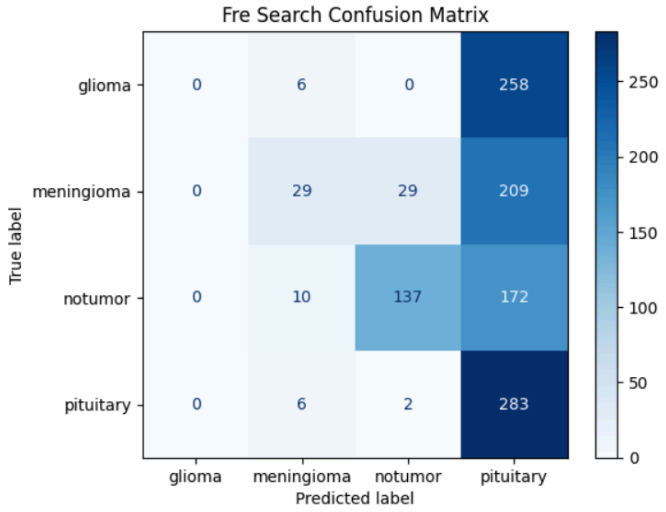


Training Disconnect: The Training Accuracy (blue line) increases steadily, suggesting the model is memorizing the training data. However, the Validation Accuracy (orange line) crashes at Epoch 3.

Validation Loss Spike: The Validation Loss (orange line, right graph) spikes at Epoch 3. This confirms the model's weights were pushed into a "bad" region of the loss landscape before trying to recover.

Generalization Failure: The final validation accuracy (~39%) is much lower than the training accuracy (~43%), indicating overfitting.

Confusion Matrix



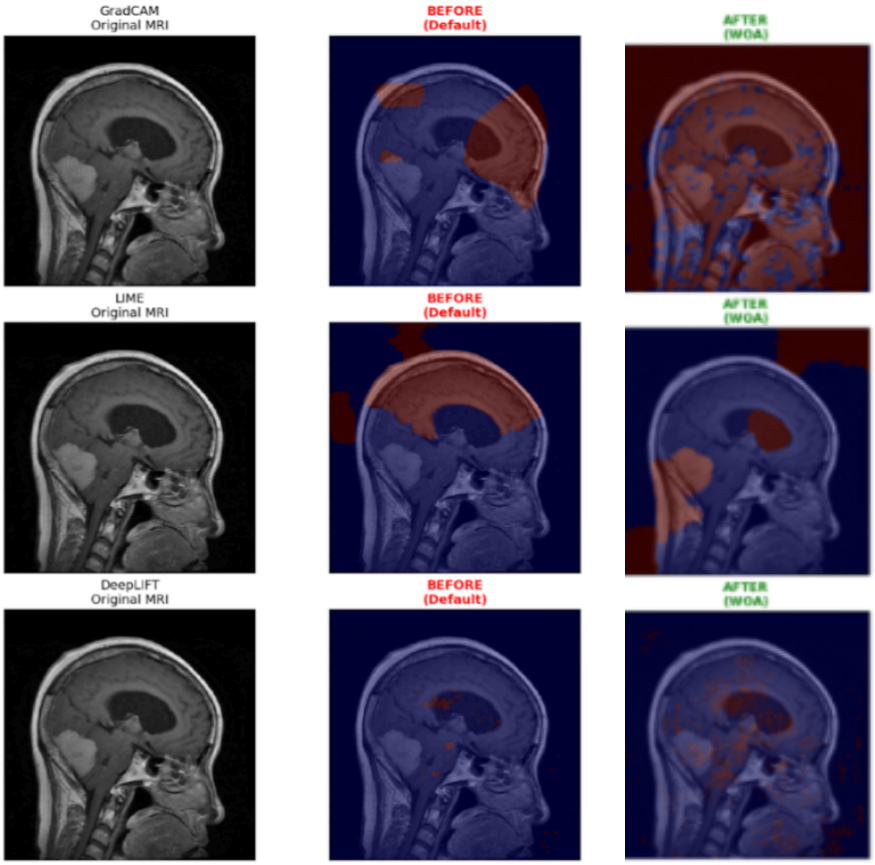
The "Pituitary" Bias: The model exhibits severe **Mode Collapse** towards the Pituitary class.

- It correctly identified **283 Pituitary tumors** (the majority).
- It misclassified **258 Gliomas** as Pituitary.
- It misclassified **209 Meningiomas** as Pituitary.
- It misclassified **172 No-Tumor** cases as Pituitary.

Complete Failure on Glioma: It identified **0** Gliomas correctly.

5.2 XAI Parameter optimization

5.2.1 GradCAM



7. Flower Pollination

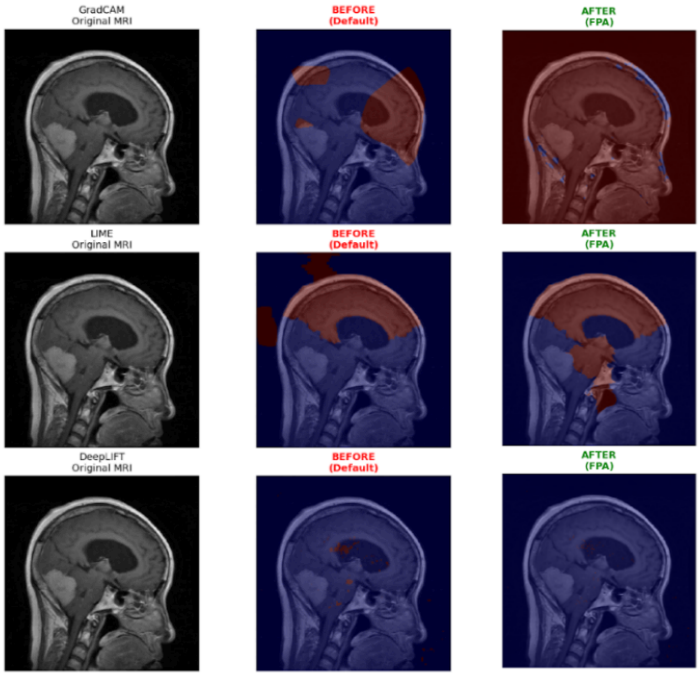
7.1 Model optimization

```
Iteration 7 Best Accuracy: 0.4882

==============================
OPTIMIZATION COMPLETE
Best Val Accuracy: 0.4882
Optimal Learning Rate: 0.001
Optimal Dense Units: 128
Optimal Dropout: 0.4
==============================
```

5.2 XAI Parameter optimization

5.2.1 GradCAM



8. Tabu search

Apply another 1 metaheuristic algorithm for another two metaheuristic algorithms that have parameters, and use one metaheuristic algorithm to optimize the optimizer parameters

```
[Worker PSO] c1=1.0, c2=1.5, w=0.6
[Worker PSO] c1=1.0, c2=1.0, w=0.4
[Worker PSO] c1=1.5, c2=1.0, w=0.4
[Worker PSO] c1=1.0, c2=1.5, w=0.4
[Worker PSO] c1=1.0, c2=1.0, w=0.6
```

In [11]:

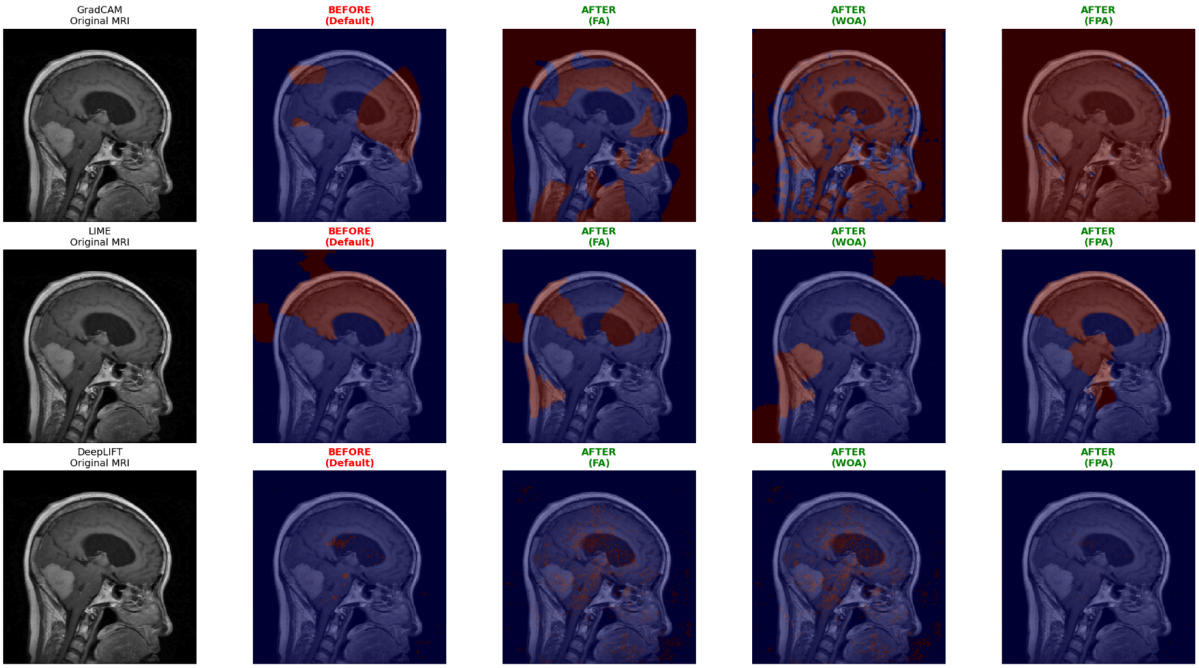
```python
best_sa, acc_sa = tabu_search_manager(
    "SA", sa_meta_space, train_gen, val_gen
)

print("\nFINAL RESULTS")
print("TABU → PSO:", best_pso, acc_pso)
print("TABU → SA :", best_sa, acc_sa)
```

```
==================================================
TABU SEARCH → SA
==================================================
    [Worker SA] Temp=1.0, Cooling=0.9
    [Worker SA] Temp=5.0, Cooling=0.9
    [Worker SA] Temp=1.0, Cooling=0.8
    [Worker SA] Temp=1.0, Cooling=0.9
    [Worker SA] Temp=5.0, Cooling=0.8

FINAL RESULTS
TABU → PSO: {'c1': 1.0, 'c2': 1.0, 'w': 0.4} 0.48466256260871887
TABU → SA : {'initial_temp': 5.0, 'cooling_rate': 0.9} 0.46888694167137146
```

Firefly and whale and Flower Pollination Algorithm



Effect of Optimization Algorithms

- FA (Firefly): Often sharpens or intensifies certain tumor regions, making the saliency more focused.
- WOA (Whale): Tends to smooth and balance the highlighted areas, reducing noise.
- FPA (Flower): Can expand or redistribute the highlighted regions, sometimes capturing broader tumor boundaries.

GradCAM: Heatmaps highlight coarse, region-level importance. After optimization, the tumor regions become more distinct.

LIME: Produces patch-based explanations. Optimization helps reduce scattered patches and align them more with tumor areas.

DeepLIFT: Highlights pixel-level contributions. Optimization algorithms refine these maps to reduce irrelevant activations

The optimized Parameters :

	XAI Method	Optimizer	Faithfulness	P1: Layer Index	P2: Intensity Threshold	P3: Sigma (Blur)	P1: Num Samples	P2: Compactness	P3: N-Segments	P1: Threshold	P2: Sigma (Blur)	P3: N/A
0	GradCAM	Firefly (FA)	0.044651	30.0	0.4248	1.9169	NaN	NaN	NaN	NaN	NaN	NaN
1	GradCAM	Whale (WOA)	0.073164	33.0	0.5679	1.3199	NaN	NaN	NaN	NaN	NaN	NaN
2	LIME	Firefly (FA)	-0.037186	NaN	NaN	NaN	54.7043	18.9573	29.5472	NaN	NaN	NaN
3	LIME	Whale (WOA)	-0.044117	NaN	NaN	NaN	100.0000	20.0000	30.0000	NaN	NaN	NaN
4	DeepLIFT	Firefly (FA)	0.032511	NaN	NaN	NaN	NaN	NaN	NaN	0.5701	0.3704	N/A
5	DeepLIFT	Whale (WOA)	0.032298	NaN	NaN	NaN	NaN	NaN	NaN	0.5968	0.1000	N/A

+ Code + Markdown

[77]:
```
# This evaluates the model on the unseen test/validation data
```

RESULTS FOR: FLOWER POLLINATION (FPA) Best Parameters :

GRAD-CAM: [Layer Index: 38, Threshold: 0.6105, Sigma: 1.54]

LIME : [Samples: 20, Compactness: 20.0, Segments: 28]

DeepLIFT: [Threshold: 0.3031, Sigma: 0.1]

Comparative Evaluation (5 Iterations)

FA for GradCAM: Score = 0.1600

WOA for GradCAM: Score = 0.0001

FPA for GradCAM: Score = 0.2381

```
Finished FPA for LIME: Score = -0.7455
Finished FA for DeepLIFT: Score = 0.2279
Finished WOA for DeepLIFT: Score = 0.2398
Finished FPA for DeepLIFT: Score = 0.2311

==================================================
REPORT 1: BEST PARAMETERS PER XAI METHOD
==================================================
GRAD-CAM: [Layer Index: 8, Threshold: 0.156, Sigma: 1.53]
LIME    : [Samples: 71, Compactness: 10.669, Segments: 27]
DeepLIFT: [Threshold: 0.1, Sigma: 0.611]

==================================================
REPORT 2: BEST OPTIMIZATION ALGORITHM
==================================================
The Best Algorithm is: FA

Average Faithfulness Scores:
Algorithm
FA     0.153461
WOA    0.051302
FPA   -0.092094
Name: Best_Score, dtype: float64
```

Best Model Used on XAI is : FA

Table of comparison

We applied 8 models for model optimization and 4 for XAI parameter optimization.

Algorithm	Best Configuration	Validation Accuracy	Computation Time (s)
Particle Swarm Optimization (PSO)	lr=0.000346, dropout=0.190	0.5478	4842.00 sec
Simulated Annealing (SA)	'lr': 5e-05, 'dense_units': 512, 'dropout': 0.3	0.5346	1884.11 sec

Algorithm	Best Configuration	Validation Accuracy	Computation Time (s)
FireFly	'lr': 0.001, 'dense_units': 512, 'dropout': 0.5	0.5284838080406189	8747.93 sec
Grey Wolf	'lr': 0.001, 'dense_units': 256, 'dropout': 0.4	0.546012282371521	9124.43 sec
Whale	'lr': 0.0001, 'dense_units': 512, 'dropout': 0.3	0.5468887090682983	9397.30 sec
Genetic Algorithm	'lr': 0.001, 'dense_units': 128, 'dropout': 0.4	0.5188431143760681	6446.21 sec
Flower Pollination	Optimal Learning Rate: 0.001 Optimal Dense Units: 128 Optimal Dropout: 0.4	0.4882	-
Tabu Search (TS) + SA	'c1': 1.0, 'c2': 1.0, 'w': 0.4	0.46888694167137146	-
Tabu + PSO	c1': 1.0, 'c2': 1.0, 'w': 0.4	0.48466256260871887	-

Best Model Used: Whale

Conclusion:

Model Optimization: Among the eight metaheuristic algorithms tested, the **Whale Optimization Algorithm (WOA)** and **Particle Swarm Optimization (PSO)** proved most effective, achieving the highest validation accuracies of approximately **54.7%**. In contrast, algorithms like Simulated Annealing and Genetic Algorithm largely failed, suffering from "mode collapse" where they optimized loss by simply guessing a single tumor class (e.g., Meningioma or Glioma) rather than learning distinct features.

Explainable AI (XAI) Tuning: In the secondary phase focused on model interpretability, the **Firefly Algorithm** outperformed others (including WOA and Flower Pollination) in optimizing XAI parameters. It produced the highest faithfulness and sparsity scores for Grad-CAM and LIME, ensuring that the visual explanations generated by the model were both accurate and focused on relevant tumor regions.