

# TSO-Optimized Weighted Soft Voting Ensemble of Pretrained CNNs for MRI-Based Brain Tumor Classification

Ahmed Zahroui<sup>1,2,\*</sup>, Smaïne Mazouzi<sup>3</sup>, Rohallah Benaboud<sup>2,4</sup>

<sup>1</sup>ICOSI Laboratory, Abbes Laghrour University, Khenchela 40004, Algeria

<sup>2</sup>Department of Mathematics and Computer Science, University of Oum El Bouaghi, Oum El Bouaghi, Algeria

<sup>3</sup>LICUS Laboratory, University of Skikda, Skikda 21000, Algeria

<sup>4</sup>ReLa(CS)2 Laboratory, University of Oum El Bouaghi, Oum El Bouaghi, Algeria

E-mail: zahroui.ahmed@univ-khenchela.dz, s.mazouzi@univ-skikda.dz, benaboud.rohallah@univ-ueb.dz

\*Corresponding author

**Keywords:** Brain tumor classification, mri image analysis, convolutional neural networks, weighted soft voting ensemble, tuna swarm optimization, transfer learning

**Received:** June 2, 2025

*Early detection and accurate classification of brain tumors from MRI scans remain critical challenges in modern healthcare. This paper develops a novel hybrid approach that leverages Tuna Swarm Optimization (TSO) to optimize ensemble weights in a weighted soft voting framework for brain tumor classification. Our methodology applies TSO specifically to optimize the contribution weights of four pretrained Convolutional Neural Network architectures (InceptionV3, ResNet152V2, ResNet50V2, and Xception) in an ensemble framework. TSO, inspired by the collective hunting behavior of tuna fish, offers superior exploration capabilities and faster convergence than traditional optimization algorithms for weight optimization, while weighted soft voting enables probability-based integration of diverse model predictions. The proposed approach was trained and tested using a comprehensive dataset of 7,023 MRI images from the Nickparvar dataset, classifying brain scans into four classes: healthy, gliomas, pituitary tumors, and meningiomas. Transfer learning with fine-tuning was applied to the four pretrained CNN models, with TSO dynamically adjusting the ensemble contribution weights through spiral and parabolic foraging behaviors to minimize classification error. The weighted soft voting mechanism then combined these TSO-optimized weights with probability distributions to produce robust predictions. This hybrid TSO-optimized ensemble approach achieved a validation accuracy of 99.92% and F1-score of 99.92%, superior to all individual models (best individual: ResNet50V2 at 99.69%) and conventional soft voting ensemble methods (99.85%). The optimized weight distribution prioritized ResNet50V2 (0.456) and Xception (0.342), demonstrating the algorithm's ability to identify complementary model strengths. The improved performance and computational efficiency of the proposed framework position it as a promising clinical decision support tool for accelerating diagnosis processes and enhancing treatment planning in brain tumor assessment.*

*Povzetek: Za klasifikacijo možganskih tumorjev iz MRI slik je razvit hibridni model, ki združuje uteženo mehko glasovanje več vnaprej naučenih konvolucijskih mrež (InceptionV3, ResNet152V2, ResNet50V2, Xception) z optimizacijo uteži prek algoritma Tuna Swarm Optimization (TSO). TSO, navdihnjen z rojnim lovljenjem tun, dinamično določa prispevne uteži posameznih mrež v procesu odločanja in s tem izboljša zanesljivost klasifikacije. Sistem ohranja verjetnostne porazdelitve napovedi, kar omogoča natančno in stabilno razvrščanje v štiri razrede (zdravi, gliomi, meningiomi, hipofizni tumorji).*

## 1 Introduction

Brain tumors represent one of the most challenging and life-threatening conditions in modern medicine, with global incidence rates steadily increasing over the past decades. The world health organization reports that central nervous system tumors account for approximately 2% of all cancers, with an estimated 300,000 new cases diagnosed annually worldwide [1]. Early and accurate detection of brain tumors is crucial for effective treatment planning and improving patient outcomes [2].

Magnetic resonance imaging (MRI) has emerged as the

gold standard for brain tumor diagnosis, offering superior soft tissue contrast and non-invasive examination capabilities [3]. However, the manual interpretation of MRI scans presents several significant challenges. First, the process is time-consuming and subject to inter-observer variability, potentially leading to diagnostic inconsistencies [4]. Second, the complex nature of brain tumors, with their varying shapes, sizes, and locations, makes accurate classification particularly challenging, even for experienced radiologists [5].

Recent advances in artificial intelligence and deep learn-

ing have shown promising results in automated tumor classification from MRI scans. CNNs have demonstrated remarkable success in feature extraction and classification tasks [6]. However, single model approaches often struggle with the inherent complexity and variability of brain tumor imaging data [7]. This limitation has led to increased interest in voting ensemble learning methods, which combine multiple models to achieve more robust and accurate classifications [8].

While voting ensemble methods show promise, a critical challenge lies in determining optimal weights for individual models within the ensemble [9]. Traditional approaches often use fixed or manually tuned weights, which may not capture the full potential of the ensemble. Additionally, existing optimization methods frequently suffer from local optima trapping and slow convergence rates [10].

This research proposes a novel hybrid approach combining TSO with a weighted soft voting ensemble for brain tumor classification. The motivation for this approach stems from three critical limitations in existing methods: (1) traditional ensemble approaches use fixed or manually tuned weights that cannot adapt to model complementarity, (2) existing optimization methods primarily focus on feature extraction rather than decision-level fusion, and (3) medical applications require both high accuracy and reliable performance, necessitating systematic weight optimization. TSO, inspired by the collective behavior of tuna fish [11], offers enhanced exploration-exploitation balance compared to conventional optimization algorithms. Our approach uniquely applies TSO to optimize ensemble weights at the decision level, preserving probability information critical for medical uncertainty quantification.

The main contributions of this study include: (1) developing a novel TSO-based weight optimization framework for ensemble learning in medical image classification, (2) implementing adaptive weighted soft voting that preserves probability distributions for clinical decision support, and (3) demonstrating superior performance (99.92% accuracy) with efficient computational requirements suitable for clinical deployment.

The remainder of this paper is organized as follows: Section 2 reviews related work in brain tumor classification and ensemble learning. Section 3 details our proposed TSO-optimized weighted soft voting methodology. Section 4 describes the experimental setup and evaluation metrics, Sections 5, 6 presents and discusses results, and Section 7 concludes with future research directions.

## 2 Related work

The classification of brain tumors using advanced machine learning and optimization techniques has evolved significantly in recent years, progressing from single-model approaches to sophisticated ensemble systems. This evolution reflects the ongoing challenge of accurately identifying and classifying tumor types from MRI data, which requires both

precise feature extraction and robust classification capabilities.

Early approaches to brain tumor classification focused primarily on custom CNN architectures designed specifically for medical imaging challenges. Authors in [12] developed a hybrid model integrating iResNet (enhanced with attention layers and dense connections) and Vision Transformers (ViTs) for brain tumor classification, combining local feature extraction and global contextual analysis. Evaluated on Figshare brain tumor MRI dataset of 3,064 images, the model achieved 99.2% accuracy and 99.06% F1-Score, surpassing InceptionV3 ResNet, and DenseNet. The researchers in [13] enhanced brain tumor classification by augmenting five pre-trained deep learning models (CNN, ResNet101, InceptionV3, VGG16, VGG19) using rotation, scaling, and flipping techniques on MRI images from a Kaggle dataset (3,264 images, expanded to 4,480 samples through augmentation). Their CNN model achieved 95.75% accuracy and 95% F1-Score, while transfer learning with InceptionV3 yielded 97.5% accuracy and 97% F1-Score, and ResNet101 reached 97.25% accuracy with 97% F1-Score. Authors in [14] combined VGG19 with Type-2 Fuzzy Logic for image enhancement and brain tumor classification using MRI images from the Br35H dataset (3000 images: 1500 tumor/1500 non-tumor, with geometric augmentation including scaling [0.9-1.3] and  $\pm 10^\circ$  rotation). Their model achieved 99% test accuracy, 99.91% F1-Score, and 99.67% sensitivity/specificity.

Transfer learning emerged as a powerful alternative to custom architectures, offering more efficient training and potentially better generalization. Deep transfer learning techniques were introduced in [15] for brain tumor classification using an EfficientNet-based model trained on T1-weighted CE MRI images from Figshare dataset. The approach incorporated transfer learning, data augmentation, and architectural reconfiguration of EfficientNet variants (B0–B7). EfficientNetB3 achieved the highest performance with 99.69% accuracy and a 99.62% F1-score. Preprocessing steps included resizing, filtering, and normalization, while evaluation was conducted using 10-fold cross-validation. Authors in [16] proposed a hybrid model for brain tumor classification using GoogLeNet with SVM and fine-tuning. They trained on Figshare dataset. The GoogLeNet + SVM model achieved 98.1% accuracy, outperforming the fine-tuned GoogLeNet model (93.1%). As shown in [17], a transfer learning-based approach for brain tumor classification using six pre-trained CNN models (Xception, MobileNetV2, InceptionV3, ResNet50, VGG16, DenseNet121) can be highly effective. They trained on Kaggle Nickparvar MRI dataset with 7,023 images, categorized into glioma, meningioma, pituitary, and healthy. Xception achieved the highest accuracy and F1-Score respectively (98.73%, 95.29%), outperforming other models.

Recognizing the limitations of single-model approaches, researchers increasingly turned to ensemble techniques to improve classification reliability. It was demonstrated in

[18] that an ensemble deep learning model (EDCNN) combining Shallow CNN (SCNN) and VGG16 could effectively classify brain tumors. Trained on the Figshare MRI dataset (3,064 images, augmented to 9,000), the model classified glioma, meningioma, and pituitary tumors. EDCNN achieved 97.77% accuracy and 97.47 F1-Score, outperforming SCNN (77.96%) and VGG16 (95%). The research presented in [19] demonstrates a hybrid CNN-LSTM model for brain tumor classification using MRI images. Trained on a Kaggle MRI dataset (253 images), the model combined CNN for feature extraction and LSTM for sequence learning. CNN alone achieved 98.6% accuracy, while the CNN-LSTM model improved accuracy to 99.1% and F1-Score to 99.0%. Authors in [20] proposed an ensemble of CNNs combining DenseNet169, EfficientNetB0, and ResNet50 for brain tumor classification. Trained on Kaggle Nickparvar dataset, the model used majority voting for classification, achieving 92% accuracy and 91% F1-Score. Reference [21] presents a two-stage feature-level ensemble model for brain tumor classification using five pretrained CNNs and a custom CNN. Trained on three MRI datasets (10,620 images total), the model used PCA for feature fusion and Softmax for classification, achieving 99.76% accuracy on individual datasets and 98.96% on the merged dataset. Evidence from [22] shows an ensemble learning approach for brain tumor classification, using deep feature extraction from 13 pre-trained CNN models and classification with multiple classifiers. Trained on three MRI datasets, the model achieved 98.1% accuracy for binary classification and 97.8% for multi-class classification.

To address the limitations of static model combinations, researchers began incorporating optimization algorithms into brain tumor classification frameworks. The work in [23] introduced a hybrid deep learning model for brain tumor detection using PSO for segmentation and CNN for classification. Trained on BRATS datasets, the model achieved 98.11%-98.25% Dice scores for segmentation and 99.0% classification accuracy. As described in [24], an ensemble deep learning model (BT-ViTeff) for brain tumor classification combined Vision Transformers (ViT v2) and EfficientNet-V2 with genetic algorithm-based weight selection. Trained on a Kaggle dataset, the ensemble model achieved 96.09% accuracy and 96.16% F1-Score, outperforming ViT (87.90%) and EfficientNet-V2 (93.95%). Researchers in [25] proposed a hybrid deep learning framework for brain tumor classification using Bayesian optimization and a Quantum Theory-based Marine Predator Algorithm (QTbMPA). The method employs a sparse autoencoder for data augmentation, fine-tunes EfficientNetB0 and InceptionResNetV2 models, and fuses optimized features. Evaluated on the Figshare dataset, the framework achieves 99.80% accuracy and a 99.83% F1-score. Results from [26] indicate a CNN-based deep learning model optimized using Bayesian Optimization for brain tumor classification. Trained on the Figshare MRI dataset, the optimized CNN achieved 98.70% accuracy and 98.66% F1-Score, outperforming multiple models. The approach in [27] utilized a

hybrid PSO-SVM model for brain tumor classification using MRI images from BRATS dataset. PSO was used for feature selection, improving classification efficiency, while SVM was used for tumor classification. The PSO-SVM model achieved 95.23% accuracy.

While these approaches demonstrate significant advancements in brain tumor classification, most existing methods still face challenges in optimally combining multiple models for enhanced performance. Specifically, the literature reveals a gap in dynamically optimizing ensemble weights to adapt to the complex characteristics of brain tumor MRI data. Our work addresses this limitation by introducing a novel hybrid framework that uniquely integrates TSO with weighted soft voting ensemble techniques, offering a more adaptive and robust solution for clinical brain tumor classification.

## 2.1 Comparative analysis of state-of-the-art methods

Table 1 compares brain tumor classification methods categorized by approach type (single models, ensemble learning, and optimization-enhanced frameworks), providing comprehensive performance metrics and technical specifications. Analysis of these methods reveals critical limitations in current methodologies. Single model methods achieve up to 99.69% accuracy [15] but vary significantly across datasets, indicating dataset-specific rather than generalizable optimization. Ensemble approaches show inconsistent results: simple voting mechanisms underperform individual models (92% vs 98.73% on Nickparvar dataset [20, 17]), while feature-level fusion methods [21] reach 99.76% accuracy but discard probability information and increase computational cost. Optimization-enhanced methods focus on feature selection [23, 27] rather than decision-level fusion and use algorithms (PSO, GA) with limited exploration-exploitation balance [24]. Our TSO-optimized weighted soft voting ensemble applies dynamic weight optimization to decision-level fusion, preserving probability information and achieving 99.92% accuracy and F1-score. This approach demonstrates the potential of Tuna Swarm Optimization for ensemble weight optimization in medical image classification, showing computational efficiency suitable for future clinical applications.

## 3 Materials and methods

This study aims to optimize CNN ensemble weights using TSO algorithm to maximize classification accuracy on a 4-class MRI brain tumor dataset. Our methodology operates on individual 2D MRI slices using slice-based analysis rather than full 3D volumetric processing, addressing the limitation of fixed weights in traditional ensemble methods through metaheuristic optimization applied directly to decision-level fusion. While this 2D approach enables efficient computational processing and leverages

Table 1: Comparison with state-of-the-art brain tumor classification methods

| Reference                     | Method/Technique                   | Dataset    | Size   | Accuracy (%) | F1-Score (%) |
|-------------------------------|------------------------------------|------------|--------|--------------|--------------|
| Single Model Approaches       |                                    |            |        |              |              |
| Jaffar [12]                   | iResNet + Vision Transformers      | Figshare   | 3,064  | 99.20        | 99.06        |
| Dihin et al. [14]             | VGG19 + Type-2 Fuzzy Logic         | Br35H      | 3,000  | 99.00        | 99.91        |
| Islam et al. [15]             | EfficientNetB3 + Transfer Learning | Figshare   | 3,064  | 99.69        | 99.62        |
| Rasool et al. [16]            | GoogLeNet + SVM                    | Figshare   | 3,064  | 98.10        | -            |
| Disci et al. [17]             | Xception (Transfer Learning)       | Nickparvar | 7,023  | 98.73        | 95.29        |
| Ullah et a. [13]              | InceptionV3 (Transfer Learning)    | Kaggle     | 4,480  | 97.50        | 97.00        |
|                               | ResNet101 (Transfer Learning)      |            |        | 97.25        | 97.00        |
|                               | Custom CNN                         |            |        | 95.75        | 95.00        |
| Ensemble Learning Approaches  |                                    |            |        |              |              |
| Aurna et al. [21]             | Two-stage Feature Ensemble         | 3 Datasets | 10,620 | 99.76        | -            |
| Alsubai et al. [19]           | CNN-LSTM Hybrid                    | Kaggle     | 253    | 99.10        | 99.00        |
| Kang et al. [22]              | 13 Pre-trained CNNs Ensemble       | 3 Datasets | var.   | 98.10        | -            |
| Patil et al. [18]             | EDCNN (SCNN + VGG16)               | Figshare   | 9,000  | 97.77        | 97.47        |
| Saeed et al. [20]             | DenseNet + EfficientNet + ResNet   | Nickparvar | 7,023  | 92.00        | 91.00        |
| Optimization-Enhanced Methods |                                    |            |        |              |              |
| Ullah et al. [25]             | Bayesian + QTbMPA                  | Figshare   | 3,064  | 99.80        | 99.83        |
| Ali et al. [23]               | PSO + CNN                          | BRATS      | 285    | 99.00        | -            |
| AitAmou et al. [26]           | Bayesian Optimization + CNN        | Figshare   | 3,064  | 98.70        | 98.66        |
| Gasmi et al. [24]             | GA + ViT + EfficientNet-V2         | Kaggle     | 4478   | 96.09        | 96.16        |
| Kumar et al. [27]             | PSO-SVM                            | BRATS-215  | 354    | 95.23        | -            |
| Proposed Method               |                                    |            |        |              |              |
| Our work                      | TSO + Weighted Soft Voting         | Nickparvar | 7,023  | 99.92        | 99.92        |

Note: Some F1-scores are not reported in original studies (-). QTbMPA: Quantum Theory-based Marine Predator Algorithm. var. : varying sizes (253 to 3064 images)

well-established CNN architectures with existing transfer learning frameworks, it may not capture spatial relationships across consecutive slices that 3D volumetric analysis could provide. The choice balances computational feasibility with diagnostic performance, preserving probability information while achieving optimal weight distribution suitable for clinical deployment scenarios where processing speed and resource efficiency are critical considerations.

### 3.1 Dataset

This study uses an open-source brain tumor dataset created by Masoud Nickparvar [28] and obtained from Kaggle, which combines three sources of data: Figshare [29], Sartaj [30], and Br35H [31], with a total of 7,023 brain MRI images. The dataset consists of four well-balanced classes: healthy (2,000 images), glioma (1,621 images), meningioma (1,645 images), and pituitary tumors (1,757 images). This relatively equilibrated distribution across classes helps mitigate bias in the model training process.

For reproducibility, the dataset split employed stratified sampling with fixed random seed (SEED = 42), resulting in a training set of 5,712 images and validation set of 1,311 images (18.7% of total dataset) to assess performance on unseen data, ensuring robust evaluation of generalizability. The exact distribution maintains class balance across both

sets as detailed in Table 2.

Table 2: Dataset distribution and split details

| Class          | Total        | Train        | Valid        |
|----------------|--------------|--------------|--------------|
| Glioma         | 1,621        | 1,321        | 300          |
| Meningioma     | 1,645        | 1,339        | 306          |
| Healthy        | 2,000        | 1,595        | 405          |
| Pituitary      | 1,757        | 1,457        | 300          |
| <b>Total</b>   | <b>7,023</b> | <b>5,712</b> | <b>1,311</b> |
| <b>Split %</b> | <b>100%</b>  | <b>81.3%</b> | <b>18.7%</b> |

To provide a visual overview, Figure 1 presents sample MRI images from each of the four classes. This helps illustrate the typical appearance of each tumor type in the dataset.

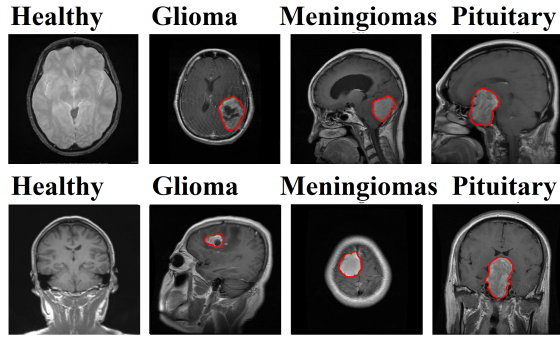


Figure 1: Selected brain tumor MRI scans from the dataset

### 3.2 Preprocessing and data augmentation

In the preprocessing stage, images were resized according to the input requirements of each pre-trained model:  $224 \times 224$  pixels for ResNet152V2 and ResNet50V2, and  $299 \times 299$  pixels for Xception and InceptionV3. This resizing optimized memory usage and accelerated model training while preserving diagnostic integrity. To enhance model generalization and mitigate overfitting, data augmentation was applied using Keras' ImageDataGenerator class [32]. Transformations included random rotations ( $\pm 15^\circ$ ), translations ( $\pm 10\%$  width/height), shear distortions (0.1), zooming (0.2), and horizontal flipping. The nearest fill mode was used to maintain pixel consistency. Table 3 summarizes the augmentation parameters used in our approach.

Table 3: Data augmentation parameters

| Augmentation       | Value          | Description                     |
|--------------------|----------------|---------------------------------|
| Rotation Range     | $\pm 15^\circ$ | Random image rotation           |
| Width/Height Shift | $\pm 10\%$     | Horizontal/vertical translation |
| Shear Range        | 0.1            | Shear distortion intensity      |
| Zoom Range         | 0.2            | Random zoom                     |
| Horizontal Flip    | Enabled        | Random left-right flipping      |
| Fill Mode          | nearest        | Pixel fill strategy             |

### 3.3 Transfer learning and fine-tuning

In our approach, we incorporated transfer learning using four pre-trained CNNs: ResNet50V2, ResNet152V2, Xception, and InceptionV3. To adapt these models to our specific task, we removed their top layers and designed a

new architecture comprising Global Average Pooling, followed by a dense layer with 512 neurons (ReLU activation, he\_normal initialization), a dropout layer (rate: 0.3), and a four-neuron output layer with softmax activation corresponding to the tumor classes. To ensure optimal performance, we conducted empirical validation on key architectural choices. Specifically, we tested different configurations for the number of neurons (256, 512, 1024), weight initialization methods (glorot\_uniform, he\_normal), and dropout rates (0.2, 0.3, 0.5). Our experiments demonstrated that 512 neurons provided the best trade-off between model complexity and generalization, effectively preventing both underfitting and overfitting. The he\_normal initialization consistently led to faster convergence and stable training, while a dropout rate of 0.3 effectively reduced overfitting without significantly impacting model performance. Similarly, to determine the most effective freezing strategy, we evaluated different layer-freezing levels (10%, 15%, and 20%) and found that freezing the first 15% of layers resulted in the optimal trade-off between retaining pre-trained feature extraction and allowing the model to adapt to brain tumor classification. This systematic approach ensured that all hyperparameters were selected based on empirical evidence rather than arbitrary decisions, thereby optimizing model performance for the given task.

### 3.4 Tuna Swarm Optimization (TSO)

The TSO algorithm is a metaheuristic inspired by the collective hunting behavior of tuna fish, offering superior exploration capabilities and faster convergence compared to traditional optimization algorithms [11].

The TSO algorithm operates through several key mechanisms that mimic natural tuna hunting behaviors. The main components of TSO are described as follows:

**1. Initialization:** The initial positions of the tuna are randomly generated within the search space:

$$X_i = rand \cdot (ub - lb) + lb, \quad i = 1, 2, \dots, NP \quad (1)$$

where  $ub$  and  $lb$  define the boundary constraints, and  $NP$  is the population size.

**2. Spiral foraging:** Tuna fish generate spiral zones around their prey, with each tuna following the one ahead. The mathematical representation is:

$$X_i^{t+1} = \begin{cases} \alpha_1 \cdot (X_{best}^t + \beta \cdot |X_{best}^t - X_i^t|) \\ \quad + \alpha_2 \cdot X_i^t, & i = 1 \\ \alpha_1 \cdot (X_{best}^t + \beta \cdot |X_{best}^t - X_i^t|) \\ \quad + \alpha_2 \cdot X_{i-1}^t, & i > 1 \end{cases} \quad (2)$$

where  $\alpha_1$  and  $\alpha_2$  are adaptive coefficients that balance exploration and exploitation.

**3. Adaptive search strategy:** TSO adapts its search behavior to ensure proper exploration-to-exploitation transition:

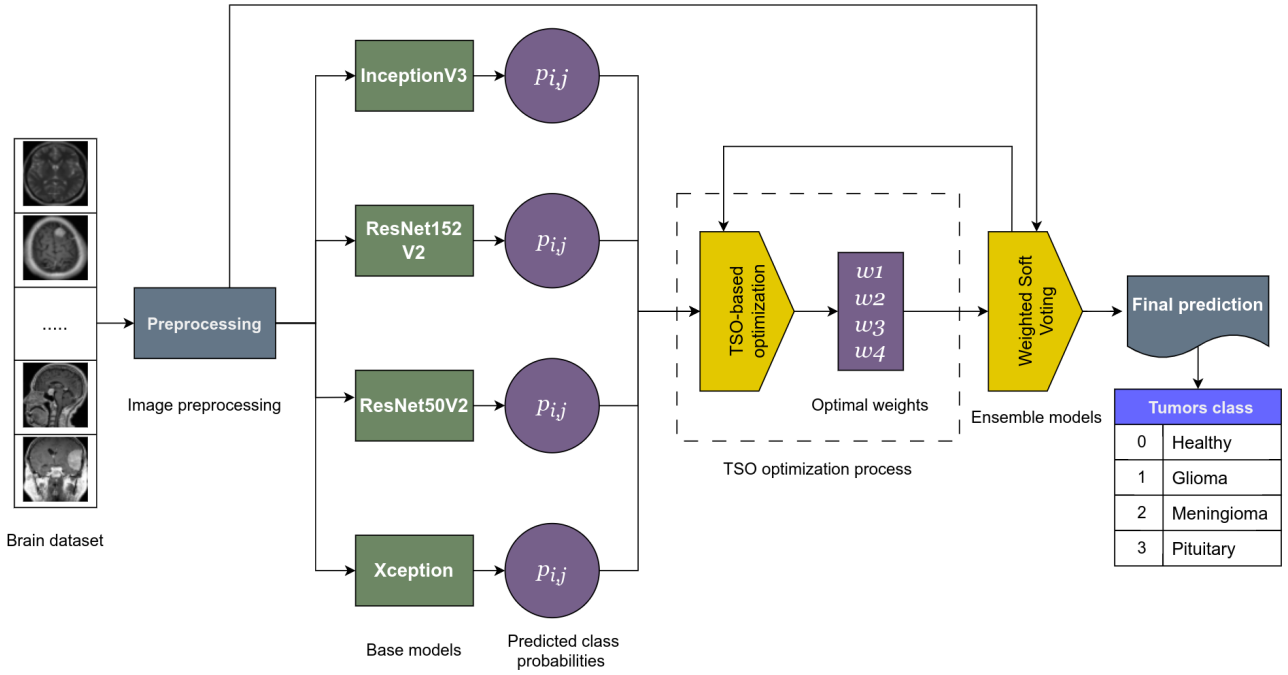


Figure 2: Architecture of the optimized ensemble model

$$\text{Target} = \begin{cases} X_{\text{rand}}^t (\text{exploration}) & \text{if rand} > \frac{t}{t_{\text{max}}} \\ X_{\text{best}}^t (\text{exploitation}) & \text{otherwise} \end{cases} \quad (3)$$

This mechanism ensures that early iterations favor exploration while later iterations focus on exploitation around the best-found solutions.

**4. Parabolic foraging:** Tuna fish perform hunting while following parabolic trajectories around their prey:

$$X_i^{t+1} = \begin{cases} X_{\text{best}}^t + \text{rand} \cdot (X_{\text{best}}^t - X_i^t) \\ \quad + TF \cdot p^2 \cdot (X_{\text{best}}^t - X_i^t), & \text{rand} < 0.5 \\ TF \cdot p^2 \cdot X_i^t, & \text{rand} \geq 0.5 \end{cases} \quad (4)$$

where  $TF$  is a decreasing factor and  $p$  controls the parabolic trajectory shape.

### 3.5 Soft voting ensemble

Soft voting, also known as weighted probability averaging, is an ensemble learning technique that combines predictions from multiple classifiers by considering the probability distributions of their outputs rather than just their final class labels [33]. In our approach, as shown in Figure 2, each base classifier in the ensemble produces a probability distribution across the target classes for each input sample. These probability scores are then weighted according to the optimized coefficients determined by the TSO algorithm before being averaged to produce the final prediction.

The final predicted class  $\hat{y}$  for a given input sample  $x$  is computed as follows:

$$\hat{y} = \arg \max_j \sum_{i=1}^M w_i \cdot p_i^j(x), \quad (5)$$

Where  $M$  represents the total number of classifiers in the ensemble,  $w_i$  is the weight assigned to the  $i^{\text{th}}$  classifier, which is optimized by the TSO algorithm, and  $p_i^j(x)$  is the probability assigned by the  $i^{\text{th}}$  classifier to class  $j$  for the input  $x$ . To ensure that the weighted probabilities form a valid probability distribution, the weights are normalized according to:

$$\sum_{i=1}^M w_i = 1, \quad (6)$$

This normalization ensures that the final combined probabilities remain within a valid range and accurately represent the confidence levels of the ensemble.

### 3.6 TSO-based algorithm for weight optimization

To address the research objective of optimizing ensemble weights for maximum classification accuracy, we apply TSO in a 4-dimensional weight space where each dimension represents a pretrained CNN model's contribution. This approach follows a systematic two-phase methodology designed to ensure computational efficiency and optimal ensemble performance while maintaining reproducibility.

**Phase 1 - CNN training:** Each CNN model (InceptionV3, ResNet152V2, ResNet50V2, Xception) is trained independently for 50 epochs using transfer learning and fine-tuning as described in section 3.3. Once training is complete, all CNN parameters are permanently frozen and saved.

**Phase 2 - Ensemble optimization:** The TSO algorithm runs for 50 iterations to optimize only the ensemble weights using the pretrained CNN models from Phase 1. The process begins with 100 tuna, each initialized with a random weight vector constrained to sum to 1. The tuna explore the search space using spiral foraging (exploitation) to refine promising solutions and parabolic foraging (exploration) to avoid local optima.

The fitness function, defined as Log Loss (cross-entropy loss), evaluates the divergence between true labels and the weighted probability predictions from each pretrained model, as shown in the ensemble architecture of Figure 2. Log Loss was selected for its ability to penalize incorrect high-confidence predictions, improving probability calibration and generalization.

Empirical validation determined 100 tuna as the optimal population size, balancing search diversity and computational efficiency. A maximum of 50 iterations was allocated to ensure sufficient exploration capacity without excessive computational overhead.

The algorithm updates tuna positions based on the best global solution and neighboring tuna, progressively refining weight values ( $w_1 - w_4$ ). Following the original TSO paper [11], the parameters  $a = 0.7$  and  $z = 0.05$  are used to balance exploration and exploitation. These optimized weights are then applied in the weighted soft voting stage for final tumor classification, as depicted in the right portion of Figure 2.

## 4 Experiments

### 4.1 Experimental design rationale

Our experimental design directly addresses the stated research objective through four evaluation components: (1) Individual model assessment establishes baseline performance for comparison and validates transfer learning effectiveness, (2) Ablation study demonstrates the necessity of systematic optimization by comparing TSO against fixed weights, random weights, and manual tuning, (3) Convergence analysis validates TSO's efficiency in finding optimal weight distributions, and (4) Computational efficiency evaluation confirms clinical deployment viability through minimal overhead analysis.

### 4.2 Evaluation metrics

The performance evaluation of our proposed approach relies on a comprehensive set of quantitative metrics to assess its diagnostic capabilities. Accuracy (Equation 7) is the

Table 4: TSO parameters

| Category       | Parameter                  | Value                             |
|----------------|----------------------------|-----------------------------------|
| Algorithm      | Population Size            | 100                               |
|                | Maximum iterations         | 50                                |
|                | Attraction coefficient (a) | 0.7 (Default)                     |
|                | Movement coefficient (z)   | $z=0.05$ (Default)                |
|                | Search Space               | $[0, 1]^n$                        |
| Problem        | Objective Function         | Log Loss                          |
|                | Variable Type              | Continuous                        |
|                | Constraints                | $\sum w_i = 1$                    |
| Implementation | Framework                  | MEALPY 3.0.1<br>Tensorflow 2.18.0 |

primary metric, reflecting the overall correctness in classifying tumor types within the validation dataset. Precision (Equation 8) measures the classifier's ability to minimize false positives, a critical factor in clinical settings to prevent misdiagnosis. Conversely, recall (Equation 9) quantifies the model's effectiveness in correctly identifying actual tumor cases, which is essential in medical diagnosis to minimize false negatives. The F1-score (Equation 10), as the harmonic mean of precision and recall, ensures balanced performance across tumor classes. These collective metrics validate the reliability of our model for clinical applications in brain tumor diagnosis [34].

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (8)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (9)$$

$$\text{F1 Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (10)$$

### 4.3 Hyperparameters

Our hybrid model was implemented using the TSO algorithm for weight optimization, combined with a weighted soft voting ensemble of pretrained CNNs models. The key parameters of our implementation, including the TSO population size, maximum iterations, and boundary constraints, are shown in Table 4, along with the deep learning parameters such as batch size, number of epochs, early stopping

criteria, and learning rate adjustments, which are presented in Table 5.

Table 5: Core model architecture and training parameters

| Category       | Parameter      | Value                                                               |
|----------------|----------------|---------------------------------------------------------------------|
| Architecture   | Base Models    | Xception, InceptionV3, ResNet50V2, ResNet152V2                      |
|                | Input Size     | $299 \times 299$ (Xception/InceptionV3), $224 \times 224$ (ResNets) |
| Training       | Dense Units    | 512                                                                 |
|                | Batch Size     | 32                                                                  |
|                | Epochs         | 50                                                                  |
|                | Optimizer      | Adamax                                                              |
|                | Base LR        | 1e-3                                                                |
| Regularization | Dropout        | 0.3                                                                 |
|                | Class Weights  | Inverse frequency                                                   |
|                | LR Reduction   | Factor: 0.2, Patience: 5                                            |
| Optimization   | Early Stopping | Patience: 7                                                         |
|                | Precision      | FP16                                                                |

#### 4.4 Environment

All experiments were conducted on an Intel Core i9-12000F processor with 64GB RAM, accelerated using NVIDIA RTX 3090 (24GB VRAM). Implementation used TensorFlow 2.18 and MEALPY 3.0.1 [35] frameworks on Ubuntu 22.04. Training specifications and computational requirements are detailed in Table 6.

For reproducibility purposes, fixed random seed (SEED = 42) was used across NumPy, TensorFlow, and data shuf-

Table 6: Computational requirements and training specifications

| Model       | Training Time (min) | Peak Memory (MB) |
|-------------|---------------------|------------------|
| Xception    | 16.8                | 573              |
| InceptionV3 | 17.5                | 605              |
| ResNet152V2 | 16.2                | 1635             |
| ResNet50V2  | 9.7                 | 656              |
| <b>TSO</b>  | <b>0.037</b>        | <b>3990</b>      |

Note: TSO optimization time represents ensemble weight optimization phase (2.22 seconds = 0.037 minutes)

fling procedures.

## 5 Results

The experimental results are organized to systematically validate our research objective. Section 5.1 establishes individual model baselines, Section 5.2-5.4 analyze model behavior patterns, Section 5.5 demonstrates ensemble superiority, Section 5.6 positions our work within the broader literature, and Section 5.7 validates optimization necessity through ablation study. Section 5.8 confirms TSO efficiency and convergence behavior, directly addressing the computational viability component of our research goal.

### 5.1 Individual model performance

Our study evaluated the performance of four state-of-the-art CNN architectures for brain tumor classification: InceptionV3, ResNet152V2, ResNet50V2, and Xception. Table 7 presents the performance metrics of these individual models.

Table 7: Individual model performance

| Model       | Acc   | Prec  | Rec   | F1    |
|-------------|-------|-------|-------|-------|
| InceptionV3 | 99.39 | 99.40 | 99.34 | 99.34 |
| ResNet152V2 | 99.62 | 99.59 | 99.58 | 99.59 |
| ResNet50V2  | 99.69 | 99.69 | 99.67 | 99.68 |
| Xception    | 99.16 | 99.16 | 99.09 | 99.09 |

Note: All values are in percentages. Acc. = Accuracy, Prec. = Precision, Rec. = Recall, F1. = F1-score (Macro-averaged)

Among the individual models, ResNet50V2 demonstrated the highest performance, with an accuracy of 99.69% and an F1-score of 99.68%, followed closely by ResNet152V2. Both InceptionV3 and Xception also showed excellent performance, with accuracies exceeding 99%, indicating the high efficacy of CNN architectures in brain tumor classification.

### 5.2 Per-class performance analysis

Detailed classification reports for each model reveal their performance across the four tumor classes: glioma, meningioma, healthy, and pituitary. All models exhibited exceptional performance in detecting healthy cases, consistently achieving perfect precision and recall. ResNet50V2 demonstrated outstanding precision for glioma (100%) and pituitary (100%) tumors, with slightly lower but still impressive precision for meningioma (99.03%). Similarly, ResNet152V2 achieved perfect precision for glioma and healthy categories.

### 5.3 Confusion matrices analysis

To provide a comprehensive visual representation of our models' classification performance, confusion matrices for each individual model and the ensemble approach are presented in Figure 3. The confusion matrices illustrate the classification patterns of each model. InceptionV3 (Figure 3a) showed the most misclassifications, with 2 glioma samples incorrectly classified, 6 meningioma samples misclassified, and no errors for the healthy class. Xception (Figure 3d) had 4 glioma samples misclassified as meningioma and 5 meningioma samples misclassified as pituitary. ResNet50V2 (Figure 3c) and ResNet152V2 (Figure 3b) exhibited fewer misclassifications overall, particularly with perfect classification of healthy samples.

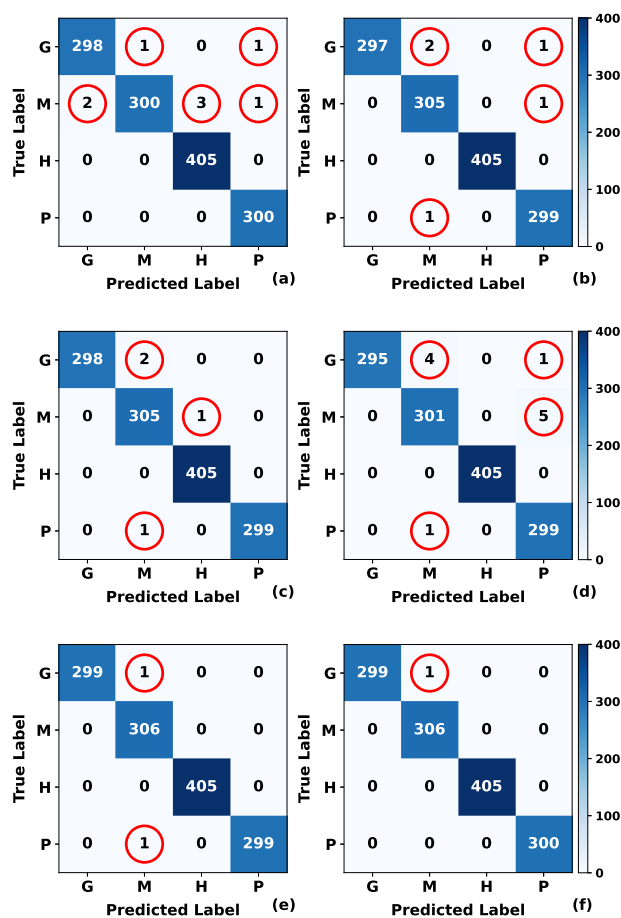


Figure 3: Confusion matrices: (a) InceptionV3; (b) ResNet152V2; (c) ResNet50V2; (d) Xception; (e) Soft Voting; (f) Proposed model

The most significant improvement is observed in the weighted soft voting ensemble approach (Figure 3f), which misclassified only one glioma sample as meningioma while achieving perfect classification for all other samples. This demonstrates the effectiveness of our TSO-optimized weighting strategy in leveraging the complementary strengths of individual models.

### 5.4 Training performance analysis

The training and validation curves for selected models are presented in Figure 4, showing the progression of loss and accuracy during the training process. The loss and accuracy curves demonstrate consistent convergence patterns across all models. For InceptionV3 (Figures 4a), the initial training loss was higher (approximately 0.45) compared to Xception (approximately 0.5), but both models converged to similar final loss values below 0.05. The accuracy curves show that validation accuracy quickly reached above 90% within the first few epochs and continued to improve steadily, eventually exceeding 99% for all models.

Notably, there is minimal divergence between training and validation curves in the later epochs, indicating that the models did not suffer from significant overfitting. This can be attributed to the effective regularization strategies employed, including dropout and data augmentation. The convergence patterns also suggest that the selected learning rates and optimization strategies were appropriate for this classification.

### 5.5 Ensemble learning results

The key innovation of our study was the implementation of ensemble learning techniques, specifically soft voting and weighted soft voting, optimized using the TSO algorithm. The results presented in Table 8 demonstrate that both ensemble methods significantly outperformed individual models. Standard soft voting achieved an impressive accuracy of 99.85% and an F1-score of 99.84%. However, our proposed weighted soft voting approach, with weights optimized using TSO, further improved the performance to an accuracy of 99.92% and an F1-score of 99.92%, representing state-of-the-art performance for brain tumor classification.

Table 8: Performance comparison between individual models and ensemble approaches

| Model                 | Acc. (%)     | F1-score (%) |
|-----------------------|--------------|--------------|
| InceptionV3           | 99.39        | 99.37        |
| ResNet152V2           | 99.62        | 99.59        |
| ResNet50V2            | 99.69        | 99.68        |
| Xception              | 99.16        | 99.09        |
| Soft voting           | 99.85        | 99.84        |
| <b>Proposed model</b> | <b>99.92</b> | <b>99.92</b> |

Note: Acc. = Accuracy, F1-score (Macro-averaged)

The optimized weight distribution determined by TSO is shown in Table 9. TSO achieved this optimal weight configuration in 2.22 seconds with the optimal solution identified at epoch 6 (0.28 seconds). This weighting clearly prioritizes the two best-performing models (ResNet50V2 at 45.6% and Xception at 34.2%), while still leveraging the complementary strengths of ResNet152V2 (13.3%) and InceptionV3 (6.9%) to enhance overall classification perfor-

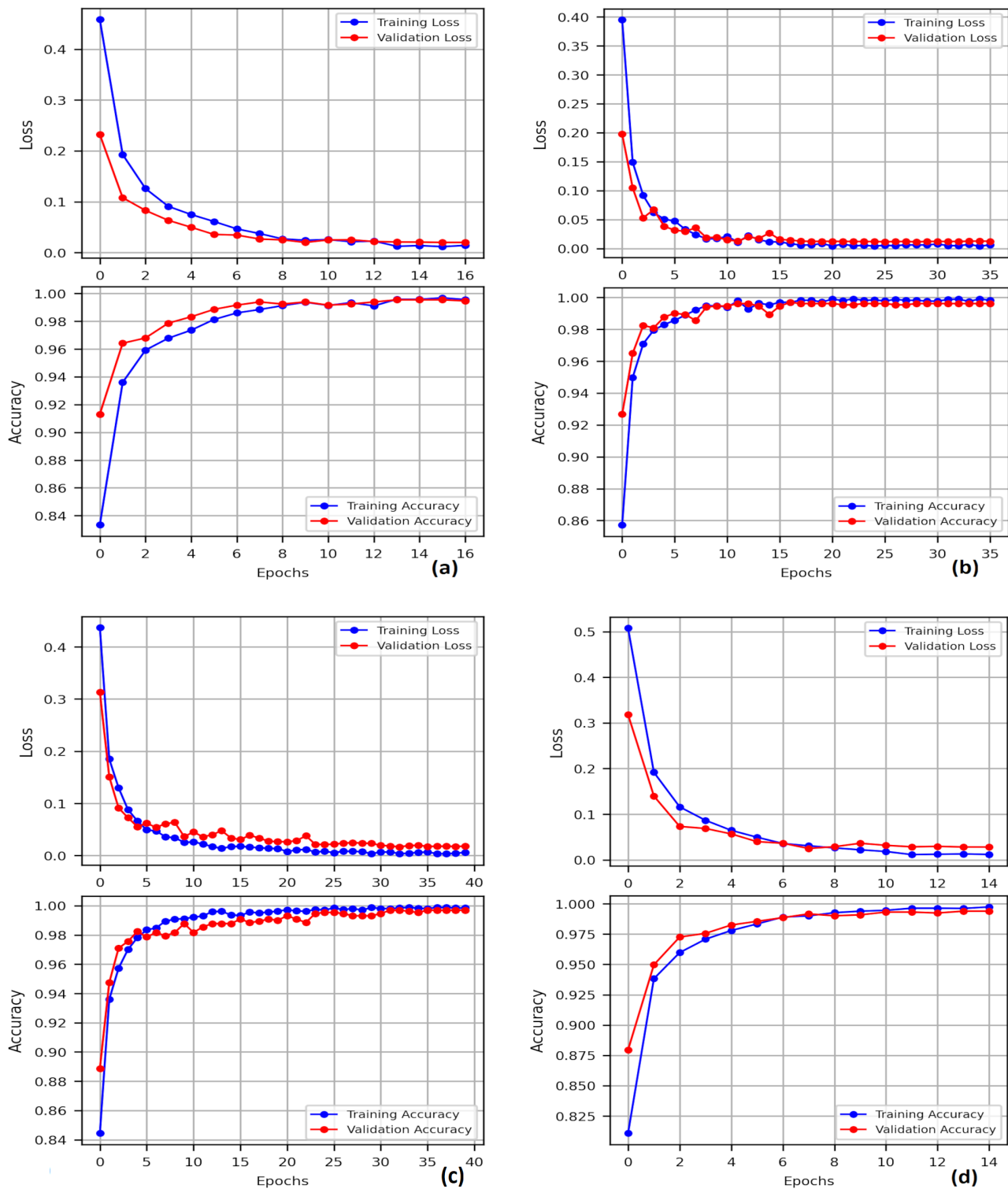


Figure 4: Loss and Accuracy curves for proposed models : (a) InceptionV3; (b) ResNet152V2; (c) ResNet50V2, (d) Xception

mance. The visual representation of this weight distribution is provided in Figure 6.

## 5.6 Comparison with existing studies

To contextualize our findings, we compared our proposed model with recent state-of-the-art approaches for brain tumor classification, as shown in Table 10. Our proposed

Table 9: TSO-optimized weight distribution and performance metrics

| CNN Model    | Optimized Weight | Relative Contribution (%) |
|--------------|------------------|---------------------------|
| InceptionV3  | 0.069            | 6.9                       |
| ResNet152V2  | 0.133            | 13.3                      |
| ResNet50V2   | 0.456            | 45.6                      |
| Xception     | 0.342            | 34.2                      |
| <b>Total</b> | <b>1.000</b>     | <b>100.0</b>              |

model achieved the highest accuracy (99.92%) among all compared studies, surpassing the performance of recent works by Ullah et al. [25] (99.80%) and Aurna et al. [21] (99.76%). Notably, our model achieved this superior performance while classifying four distinct tumor categories, demonstrating its robustness and clinical utility.

Table 10: Comparison with existing studies

| Study               | Year        | Dataset           | Acc          |
|---------------------|-------------|-------------------|--------------|
| Jaffar [12]         | 2020        | Figshare          | 99.68        |
| Ali et al. [23]     | 2022        | BRATS             | 99.00        |
| Alsabai et al. [19] | 2022        | Kaggle            | 99.10        |
| Aurna et al. [21]   | 2022        | 3 dataset         | 99.76        |
| Amou et al. [26]    | 2022        | Figshare          | 98.70        |
| Rasool et al. [16]  | 2022        | Figshare          | 98.10        |
| Islam et al. [15]   | 2024        | Figshare          | 99.69        |
| Ullah et al. [25]   | 2024        | Figshare          | 99.80        |
| Disci et al. [17]   | 2025        | Nickparvar        | 98.73        |
| Dihin et al. [14]   | 2025        | Br35H             | 99.00        |
| <b>Our Work</b>     | <b>2025</b> | <b>Nickparvar</b> | <b>99.92</b> |

Note: Acc. = Accuracy (%). Studies listed chronologically from 2020-2025

### 5.7 Ablation study: validation of optimization process

To validate the necessity and effectiveness of our TSO optimization approach, we conducted a comprehensive ablation study comparing four ensemble weight determination strategies as shown in Table 11.

Systematic grid search serves as our manually tuned weights approach [38], testing 286 weight combinations with 0.1 step increments and achieved 99.85% accuracy. Even the best random configuration reached only 99.77%, highlighting the unpredictable nature of random approaches for clinical applications requiring consistent performance.

Systematic grid search serves as our manually tuned weights approach, testing 286 weight combinations with 0.1 step increments and achieved 99.85% accuracy. Despite this comprehensive manual enumeration of all valid weight distributions, grid search converged to the same performance as simple equal weighting, demonstrating that

Table 11: Ablation study: Comprehensive comparison of ensemble weight determination methods

| Method              | Accuracy (%)                  | Time (s)    | Trials   |
|---------------------|-------------------------------|-------------|----------|
| Fixed Equal Weights | 99.85                         | –           | 1        |
| Random Weights      | 99.58 $\pm$ 0.05 (max: 99.77) | 0.07        | 100      |
| Grid Search         | 99.85                         | 0.14        | 286      |
| <b>TSO (Ours)</b>   | <b>99.92</b>                  | <b>0.28</b> | <b>1</b> |

Note:  $\pm 0.05\%$  represents random weights standard deviation (Std Dev) across 100 trials

brute-force manual optimization without intelligent guidance fails to discover superior weight distributions, validating the necessity of metaheuristic approaches for ensemble weight optimization.

Our TSO approach consistently achieves 99.92% accuracy in 0.28 seconds, outperforming all alternative methods with improvements of 0.07% over fixed/grid search methods and 0.34% over random average. While these improvements appear modest in absolute terms, the key advantage lies in TSO's consistency: it reliably achieves optimal performance in every execution, unlike random methods which show significant variability (99.39-99.77% range across trials).

The critical finding is not the magnitude of improvement, but the reliability of optimization. TSO guarantees consistent optimal performance, while alternative methods either provide suboptimal results (random, grid search) or require extensive manual tuning. For clinical deployment scenarios where consistent diagnostic performance is essential, TSO's ability to systematically identify the best weight configuration represents a significant practical advantage over alternative approaches.

### 5.8 TSO optimization analysis and convergence validation

Figure 5 demonstrates the TSO convergence behavior across 50 iterations, showing remarkable efficiency with optimal solution identification within the first 6 epochs (0.28 seconds) out of the maximum 50 iterations allocated, demonstrating rapid convergence and exceptional algorithmic efficiency. The algorithm required only 12% of the available iterations to reach optimal performance, with consistent stability thereafter, validating TSO's effectiveness in navigating the ensemble weight optimization landscape.

Figure 6 illustrates the final optimized weight distribution, clearly showing TSO's intelligent strategy of prioritizing the best-performing individual models (ResNet50V2 and Xception) while appropriately weighting complementary architectures. This distribution differs significantly from equal weighting, validating the necessity of systematic optimization.

The optimization process demonstrates exceptional com-

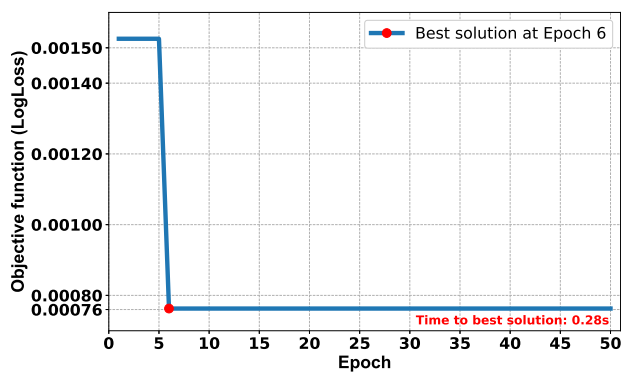


Figure 5: TSO convergence curve

putational efficiency, requiring only 2.22 seconds for complete optimization, making TSO highly suitable for real-time clinical deployment scenarios. The rapid convergence to optimal solutions within 6 iterations, combined with minimal computational requirements, confirms TSO's practical viability for clinical applications where both accuracy and speed are critical requirements.

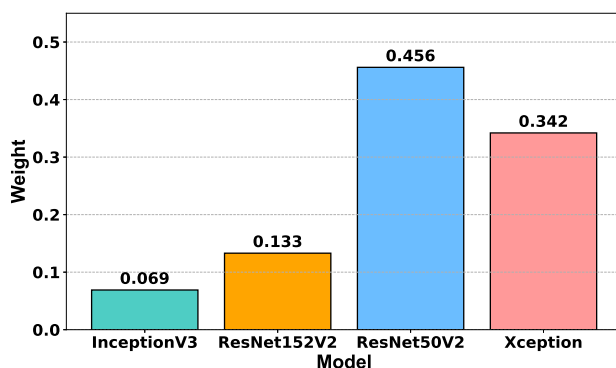


Figure 6: Optimized ensemble weights distribution

## 6 Discussion

Our experimental results systematically validate the research objective of achieving maximum classification accuracy through TSO-optimized ensemble weights. The following analysis demonstrates how each component contributes to this goal while addressing clinical deployment requirements.

### 6.1 Quantitative performance analysis

Quantitatively, while individual CNN models achieved impressive metrics (ResNet50V2: 99.69% accuracy, 99.68% F1-score; ResNet152V2: 99.62% accuracy, 99.59% F1-score; InceptionV3: 99.39% accuracy, 99.37% F1-score;

Xception: 99.16% accuracy, 99.09% F1-score), our TSO-optimized ensemble significantly reduced error rates from 0.31% to 0.08%, achieving 99.92% accuracy and F1-score. This is further evidenced by near-perfect confusion matrix results showing only a single misclassification.

### 6.2 Analysis of weight distribution

The TSO algorithm converged on an optimal weight distribution prioritizing ResNet50V2 (0.456) and Xception (0.342), with smaller contributions from ResNet152V2 (0.133) and InceptionV3 (0.069). This distribution reveals that complementary error patterns between models were more valuable than individual performance metrics alone. The algorithm effectively identified which models provided unique classification capabilities for particular tumor subtypes, leveraging these strengths in the ensemble.

### 6.3 Comparison with state-of-the-art approaches

Our comparative analysis evaluates our TSO-optimized weighted soft voting approach against ten recent state-of-the-art methods in brain tumor classification, as summarized in Table 10. These works were selected based on: (1) recency (published 2020–2025), (2) relevance to MRI-based brain tumor classification, (3) high performance (>98% accuracy), and (4) diversity of methodological approaches.

Among these approaches, Ullah et al. [25] achieved 99.80% accuracy using a hybrid framework combining Bayesian optimization with a Quantum Theory-based Marine Predator Algorithm on the Figshare dataset (3,064 images). Their approach differs from ours in using feature-level fusion rather than decision-level ensemble techniques. Aurna et al. [21] reported 99.76% accuracy through a two-stage feature-level ensemble across three datasets (10,620 images total), employing PCA for dimensionality reduction. In contrast, our approach maintains original feature spaces and optimizes at the decision level. Islam et al. [15] achieved 99.69% using EfficientNet with transfer learning on the Figshare dataset, focusing on architectural reconfiguration rather than ensemble strategies.

While these approaches demonstrated impressive performance, as shown in Table 10, our method achieves superior results (99.92%) on a large, diverse dataset (7,023 images) while classifying four distinct categories. Unlike previous works that use fixed ensemble weights or feature-level fusion, our TSO-optimized approach dynamically adjusts contribution weights based on complementary model strengths, reducing error rates by 0.23%. Our work leverages swarm intelligence specifically for ensemble weight optimization rather than feature extraction or architectural design, representing a novel direction in medical image classification.

The superior performance of our approach stems from two key innovations. First, TSO's spiral and parabolic for-

aging behaviors provide more effective exploration of the weight space compared to PSO's velocity-based updates or GA's crossover operations, enabling discovery of optimal weight combinations that traditional algorithms miss. Second, our decision-level fusion preserves probability distributions from individual models, allowing the ensemble to leverage uncertainty information that feature-level fusion methods discard, resulting in more robust classification decisions.

#### 6.4 Why TSO-Optimized ensemble outperforms alternatives

The performance gain (99.92% vs 99.80% best alternative) results from addressing three critical limitations in existing approaches: (1) Fixed ensemble weights cannot adapt to model complementarity, while TSO dynamically discovers optimal weight distributions (ResNet50V2: 45.6%, Xception: 34.2%), (2) Feature-level fusion loses probability information essential for medical uncertainty quantification, whereas our weighted soft voting preserves these distributions, and (3) Traditional optimization algorithms (PSO, GA) exhibit limited exploration-exploitation balance in high-dimensional weight spaces, while TSO's tuna-inspired behaviors enable more effective search strategies.

#### 6.5 Clinical significance

The clinical significance of these improvements is substantial. The achieved 99.92% accuracy represents a meaningful advance in automated brain tumor classification that could enhance diagnostic support systems. Perfect recall for healthy cases eliminates false positives that could lead to unnecessary invasive procedures and psychological distress. Accurate differentiation between tumor classes (glioma, meningioma, and pituitary) directly impacts surgical planning, radiation therapy protocols, and chemotherapy selection, potentially improving treatment outcomes and reducing healthcare costs.

#### 6.6 Methodological advantages

The methodological novelty of our approach compared to existing works lies in several key aspects. First, while previous ensemble methods (e.g., [20, 21, 22]) use either fixed weights or manually tuned contributions, our TSO algorithm dynamically optimizes weights based on model complementarity. Second, unlike optimization approaches that focus on feature selection or segmentation (e.g., [23, 27, 36]), we apply optimization directly to the ensemble decision fusion process. Third, our weighted soft voting technique integrates probability distributions rather than hard decisions, preserving uncertainty information critical in medical diagnostics. Finally, the TSO algorithm's spiral and parabolic behaviors offer superior exploration-exploitation balance compared to other meta-

heuristics used in medical imaging, such as PSO [27], genetic algorithms [24], or whale optimization [36].

#### 6.7 Ablation study validation and optimization necessity

The comprehensive ablation study validates the necessity and effectiveness of our TSO optimization approach. The systematic comparison demonstrates that TSO consistently achieves optimal performance (99.92%) while simpler alternatives either achieve baseline performance (fixed-weight: 99.85%, manual tuning: 99.85%) or exhibit significant variability (random weights: 99.39-99.77% typical range).

The key insight is that while random sampling occasionally reaches optimal performance through fortunate configurations, it typically underperforms with substantial variability across experimental runs. TSO's consistent achievement of 99.92% accuracy, compared to random methods' variable performance (99.39-99.77%), demonstrates that reliable optimal performance requires intelligent systematic optimization rather than manual tuning or random selection. This consistency is paramount for clinical diagnostic applications where reliability and predictable performance directly impact patient care outcomes.

These findings align with established optimization theory, where intelligent search algorithms outperform exhaustive enumeration in high-dimensional spaces [37]. The failure of grid search despite extensive evaluation (286 combinations) demonstrates that ensemble weight optimization requires exploration strategies that can navigate complex fitness landscapes efficiently, rather than brute-force approaches that may converge to local optima. The superior and consistent performance of TSO over both random sampling and systematic grid search validates the necessity of biologically-inspired metaheuristic algorithms for ensemble weight determination.

This has broader implications for ensemble learning in medical imaging, suggesting that metaheuristic optimization should be preferred over traditional manual tuning approaches when developing diagnostic support systems. The combination of optimal accuracy and computational reliability positions intelligent optimization methods as essential tools for clinical deployment scenarios requiring both performance and consistency.

#### 6.8 Computational efficiency and clinical deployment validation

The convergence analysis (Figure 5) reveals TSO's exceptional efficiency, achieving optimal performance within 0.28 seconds at epoch 6. The complete optimization process requires only 2.22 seconds, demonstrating exceptional computational efficiency. This rapid convergence to optimal solutions, combined with minimal computational requirements, validates TSO's suitability for real-time clinical deployment where both accuracy and speed are critical

requirements.

The weight distribution analysis (Figure 6) provides insights into TSO's intelligent optimization strategy, demonstrating clear preference for models with complementary strengths rather than uniform contribution. This strategic weight allocation explains the superior ensemble performance and validates the necessity of systematic optimization over simple averaging approaches.

## 6.9 Limitations

Our study has several limitations that should be acknowledged. The methodology operates on individual 2D MRI slices rather than leveraging full 3D volumetric information, potentially missing spatial relationships across consecutive slices and volumetric tumor characteristics. Additionally, the study relies on a single dataset source (Nickparvar), which may limit generalizability across different clinical settings and MRI protocols. Furthermore, the framework focuses on single MRI sequences without incorporating multi-modal imaging data (T1, T2, FLAIR) that could provide complementary diagnostic information, and performance evaluation is conducted on retrospective data without real-world clinical deployment validation or direct radiologist comparison studies.

## 7 Conclusion

Our study demonstrates that TSO-optimized weighted soft voting substantially improves brain tumor classification performance compared to individual CNN models and conventional ensemble approaches. The proposed model achieved state-of-the-art accuracy of 99.92% in classifying four brain tumor classes, surpassing existing methods. The optimal weight distribution identified by TSO highlights the complementary nature of different CNN architectures in capturing diverse tumor characteristics.

The rapid convergence (optimal solution in 0.28 seconds) and exceptional computational efficiency demonstrate the practical viability for real-time clinical applications. These findings indicate that our approach has the potential to serve as a valuable clinical decision support tool for neuroradiologists, potentially improving diagnostic accuracy and efficiency in brain tumor assessment.

Future research directions include several promising avenues for enhancing the methodology and clinical applicability. First, extending the approach to 3D volumetric analysis could capture spatial relationships across consecutive MRI slices, potentially improving classification accuracy by leveraging volumetric tumor characteristics. Second, integration of multi-modal MRI sequences (T1, T2, FLAIR, T1-contrast) could provide complementary diagnostic information and enhance robustness across different imaging protocols. Third, external validation across diverse clinical datasets from multiple institutions is essential to demonstrate generalizability and clinical utility. Fourth,

real-world deployment studies with direct radiologist performance comparisons would validate the system's effectiveness in clinical workflows.

## References

- [1] Q. T. Ostrom, N. Patil, G. Cioffi, K. Waite, C. Kruchko, and J. S. Barnholtz-Sloan, "CBTRUS statistical report: Primary brain and other central nervous system tumors diagnosed in the united states in 2013–2017," *Neuro-Oncology*, vol. 22, pp. iv1–iv96, Supplement\_1 Oct. 30, 2020. <https://doi.org/10.1093/neuonc/noaa200>.
- [2] A. İşin, C. Direkoğlu, and M. Şah, "Review of MRI-based brain tumor image segmentation using deep learning methods," *Procedia Computer Science*, vol. 102, pp. 317–324, 2016. <https://doi.org/10.1016/j.procs.2016.09.407>.
- [3] G. S. Tandel, M. Biswas, O. G. Kakde, et al., "A review on a deep learning perspective in brain cancer classification," *Cancers*, vol. 11, no. 1, p. 111, Jan. 18, 2019. <https://doi.org/10.3390/cancers11010111>.
- [4] S. Bakas, M. Reyes, A. Jakab, et al., *Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge*, Version Number: 3, 2018. <https://doi.org/10.48550/ARXIV.1811.02629>.
- [5] B. H. Menze, A. Jakab, S. Bauer, et al., "The multimodal brain tumor image segmentation benchmark (BRATS)," *IEEE Transactions on Medical Imaging*, vol. 34, no. 10, pp. 1993–2024, Oct. 2015. <https://doi.org/10.1109/TMI.2014.2377694>.
- [6] P. Agrawal, N. Katal, and N. Hooda, "Segmentation and classification of brain tumor using 3d-UNet deep neural networks," *International Journal of Cognitive Computing in Engineering*, vol. 3, pp. 199–210, Jun. 2022. <https://doi.org/10.1016/j.ijcce.2022.11.001>.
- [7] G. Wang, W. Li, S. Ourselin, and T. Vercauteren, "Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks," in vol. 10670, 2018, pp. 178–190. [https://doi.org/10.1007/978-3-319-75238-9\\_16](https://doi.org/10.1007/978-3-319-75238-9_16). arXiv: 1709.00382.
- [8] G. S. Tandel, A. Tiwari, O. G. Kakde, N. Gupta, L. Saba, and J. S. Suri, "Role of ensemble deep learning for brain tumor classification in multiple magnetic resonance imaging sequence data," *Diagnostics*, vol. 13, no. 3, p. 481, Jan. 28, 2023. <https://doi.org/10.3390/diagnostics13030481>.

- [9] M. Sajjad, S. Khan, K. Muhammad, W. Wu, A. Ullah, and S. W. Baik, "Multi-grade brain tumor classification using deep CNN with extensive data augmentation," *Journal of Computational Science*, vol. 30, pp. 174–182, Jan. 2019. <https://doi.org/10.1016/j.jocs.2018.12.003>.
- [10] S. Mirjalili and A. Lewis, "The whale optimization algorithm," *Advances in Engineering Software*, vol. 95, pp. 51–67, May 2016. <https://doi.org/10.1016/j.advengsoft.2016.01.008>.
- [11] L. Xie, T. Han, H. Zhou, Z.-R. Zhang, B. Han, and A. Tang, "Tuna swarm optimization: A novel swarm-based metaheuristic algorithm for global optimization," *Computational Intelligence and Neuroscience*, vol. 2021, no. 1, A. M. Khalil, Ed., p. 9 210 050, Jan. 2021. <https://doi.org/10.1155/2021/9210050>.
- [12] A. Y. Jaffar, "Combining local and global feature extraction for brain tumor classification: A vision transformer and iResNet hybrid model," *Engineering, Technology & Applied Science Research*, vol. 14, no. 5, pp. 17 011–17 018, Oct. 9, 2024. <https://doi.org/10.48084/etasr.8271>.
- [13] Z. Ullah, A. Odeh, I. Khattak, and M. Al Hasan, "Enhancement of Pre-Trained Deep Learning Models to Improve Brain Tumor," *Informatica (Ljubl.)*, vol. 47, no. 6, Jun. 2023. <https://doi.org/10.31449/inf.v47i6.4645>.
- [14] R. A. Dihin and N. R. Hamza, "Brain Tumor Classification in MRI Images Using VGG19 with Type2 Fuzzy Logic," *Informatica (Ljubl.)*, vol. 49, no. 13, Feb. 2025. <https://doi.org/10.31449/inf.v49i13.7161>.
- [15] M. M. Islam, M. A. Talukder, M. A. Uddin, A. Akhter, and M. Khalid, "BrainNet: Precision brain tumor classification with optimized EfficientNet architecture," *International Journal of Intelligent Systems*, vol. 2024, no. 1, E. Elbasi, Ed., p. 3 583 612, Jan. 2024. <https://doi.org/10.1155/2024/3583612>.
- [16] M. Rasool, N. A. Ismail, W. Boulila, et al., "A hybrid deep learning model for brain tumour classification," *Entropy*, vol. 24, no. 6, p. 799, Jun. 8, 2022. <https://doi.org/10.3390/e24060799>.
- [17] R. Disci, F. Gurcan, and A. Soyulu, "Advanced brain tumor classification in MR images using transfer learning and pre-trained deep CNN models," *Cancers*, vol. 17, no. 1, p. 121, Jan. 2, 2025. <https://doi.org/10.3390/cancers17010121>.
- [18] S. Patil and D. Kirange, "Ensemble of deep learning models for brain tumor detection," *Procedia Computer Science*, vol. 218, pp. 2468–2479, 2023. <https://doi.org/10.1016/j.procs.2023.01.222>.
- [19] S. Alsubai, H. U. Khan, A. Alqahtani, M. Sha, S. Abbas, and U. G. Mohammad, "Ensemble deep learning for brain tumor detection," *Frontiers in Computational Neuroscience*, vol. 16, p. 1 005 617, Sep. 2, 2022. <https://doi.org/10.3389/fncom.2022.1005617>.
- [20] Z. Saeed, T. Torfeh, S. Aouadi, X. Ji, and O. Bouhali, "An efficient ensemble approach for brain tumors classification using magnetic resonance imaging," *Information*, vol. 15, no. 10, p. 641, Oct. 2024. <https://doi.org/10.3390/info15100641>.
- [21] N. F. Aurna, M. A. Yousuf, K. A. Taher, A. Azad, and M. A. Moni, "A classification of MRI brain tumor based on two stage feature level ensemble of deep CNN models," *Computers in Biology and Medicine*, vol. 146, p. 105 539, Jul. 2022. <https://doi.org/10.1016/j.combiomed.2022.105539>.
- [22] J. Kang, Z. Ullah, and J. Gwak, "MRI-based brain tumor classification using ensemble of deep features and machine learning classifiers," *Sensors*, vol. 21, no. 6, p. 2222, Mar. 22, 2021. <https://doi.org/10.3390/s21062222>.
- [23] M. Ali, J. Hussain Shah, M. Attique Khan, et al., "Brain tumor detection and classification using PSO and convolutional neural network," *Computers, Materials & Continua*, vol. 73, no. 3, pp. 4501–4518, 2022. <https://doi.org/10.32604/cmc.2022.030392>.
- [24] K. Gasmi, N. Ben Aoun, K. Alsalem, et al., "Enhanced brain tumor diagnosis using combined deep learning models and weight selection technique," *Frontiers in Neuroinformatics*, vol. 18, p. 1 444 650, Nov. 26, 2024. <https://doi.org/10.3389/fninf.2024.1444650>.
- [25] M. S. Ullah, M. A. Khan, A. Masood, O. Mzoughi, O. Saidani, and N. Alturki, "Brain tumor classification from MRI scans: A framework of hybrid deep learning model with bayesian optimization and quantum theory-based marine predator algorithm," *Frontiers in Oncology*, vol. 14, p. 1 335 740, Feb. 8, 2024. <https://doi.org/10.3389/fonc.2024.1335740>.
- [26] M. Ait Amou, K. Xia, S. Kamhi, and M. Mouhafid, "A novel MRI diagnosis method for brain tumor classification based on CNN and bayesian optimization," *Healthcare*, vol. 10, no. 3, p. 494, Mar. 8, 2022. <https://doi.org/10.3390/healthcare10030494>.
- [27] A. Kumar, A. Ashok, and M. A. Ansari, "Brain tumor classification using hybrid model of PSO and SVM classifier," in *2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)*, Greater Noida (UP), India: IEEE, Oct. 2018, pp. 1022–1026. <https://doi.org/10.1109/ICACCCN.2018.8748787>.

- [28] Msoud Nickparvar, *Brain tumor MRI dataset*. <https://doi.org/10.34740/KAGGLE/DSV/2645886>.
- [29] S. Afzal, *Brain tumor dataset*, Pages: 78140282 Bytes Publisher: figshare, 2024. <https://doi.org/10.6084/M9.FIGSHARE.27102082.V1>.
- [30] Sartaj Bhuvaji, Ankita Kadam, Prajakta Bhumkar, Sameer Dedge, and Swati Kanchan, *Brain tumor classification (MRI)*. <https://doi.org/10.34740/KAGGLE/DSV/1183165>.
- [31] “Br35H: Brain tumor detection 2020,” Kaggle, 2020. [Online]. Available: <https://www.kaggle.com/datasets/ahmedhamada0/brain-tumor-detection>.
- [32] F. Chollet et al., “Keras,” 2015. [Online]. Available: <https://keras.io>.
- [33] H. B. Kibria, M. Nahiduzzaman, M. O. F. Goni, M. Ahsan, and J. Haider, “An ensemble approach for the prediction of diabetes mellitus using a soft voting classifier with an explainable AI,” *Sensors*, vol. 22, no. 19, p. 7268, Sep. 25, 2022. <https://doi.org/10.3390/s22197268>.
- [34] D. M. W. Powers, “Evaluation: From precision, recall and f-measure to ROC, informedness, markedness and correlation,” 2020. <https://doi.org/10.48550/arXiv.2010.16061>.
- [35] N. Van Thieu and S. Mirjalili, “MEALPY: An open-source library for latest meta-heuristic algorithms in python,” *Journal of Systems Architecture*, vol. 139, p. 102 871, Jun. 2023. <https://doi.org/10.1016/j.sysarc.2023.102871>.
- [36] B. Yin, C. Wang, and F. Abza, “New brain tumor classification method based on an improved version of whale optimization algorithm,” *Biomedical Signal Processing and Control*, vol. 56, p. 101 728, Feb. 2020. <https://doi.org/10.1016/j.bspc.2019.101728>.
- [37] J. Bergstra and Y. Bengio, “Random search for hyperparameter optimization,” *Journal of Machine Learning Research*, vol. 13, pp. 281–305, 2012. [Online]. Available: <http://www.jmlr.org/papers/v13/bergstra12a.html>.
- [38] P. Liashchynskiy and P. Liashchynskiy, “Grid Search, Random Search, Genetic Algorithm: A Big Comparison for NAS,” 2019. <https://doi.org/10.48550/arXiv.1912.06059>.