A Hybrid Approach for Multiclass Brain Tumor Classification: Ensemble Convolutional Neural Networks and PSO Optimization

Mohamed Djemai[1]* , Hilal Naimi[1] , Mohammed Bourennane[2]

[1] Laboratory for Modeling Simulation and Optimization of Systems Complexes, University of Djelfa, Djelfa 17000, Algeria
[2] Laboratory of Telecommunications and Smart Systems, University of Djelfa, Djelfa 17000, Algeria

Corresponding Author Email: m.djemai@univ-djelfa.dz

ABSTRACT

Brain tumors are irregular cell growths occurring in the brain or central spinal canal, classified as benign or malignant, posing challenges in diagnosis and treatment. Convolutional Neural Networks (CNNs) have become powerful tools in medical imaging analysis, particularly for classifying and segmenting brain tumors from Computed Tomography (CT) scans or Magnetic Resonance Imaging (MRI). However, using a single pretrained CNN model may not fully capture the data's variability and complexity, potentially reducing classification accuracy due to missed features. In this paper, we propose an ensemble fusion method integrating three deep CNNs: VGG19, GoogLeNet, and ResNet50 to classify brain MR images into Glioma tumor, Meningioma tumor, Pituitary tumor, and Normal brain categories. By combining these models into an ensemble, we aim to incorporate all pertinent data features, enhancing classification accuracy and generalization to new data. We employed the Particle Swarm Optimization (PSO) algorithm to optimize the penalty parameter C for linear SVM and select optimal features from multiple CNNs, enabling our ensemble model to significantly enhance performance with a reduced feature subset. Our ensemble approach achieved outstanding results across all metrics: accuracy, precision, sensitivity, and F1-score was achieved 99.30%, accompanied by a high specificity of 99.70%. Moreover, our framework demonstrates competitive performance compared to prior studies.

1. INTRODUCTION

Brain tumors are abnormal masses of neoplastic tissue characterized by uncontrolled cell multiplication and growth, unchecked by the mechanisms regulating normal cell division. They can appear at any stage of life [1] and are among the significant diseases affecting the human central nervous system (CNS) [1]. Brain tumors are regarded as among the devastating illnesses that profoundly impact the human body [2]. The majority of brain tumors lack a clear cause [3]. They can be classified as benign or malignant. Benign tumors, which are noncancerous, do not spread and do not invade other parts of the body. In contrast, a malignant tumor is a cancerous tumor, characterized by rapid growth and the possibility of spreading to other areas of the body. Over 120 types of tumors in CNS have been reported by the World Health Organization (WHO) [4].

Three main types of brain tumors exist:

Pituitary tumors, located in the pituitary gland and responsible for hormone production related to growth and other glands. Meningioma tumors, commonly benign and characterized by slow growth, surround the meninges, exhibiting a higher occurrence among women than men. The incidence rates of pituitary and meningioma tumors in clinical practice are approximately 12% and 15-20% respectively [5].

Gliomas, originating from glial cells or the supportive tissue surrounding nerve cells, account for 45% of tumors [6].

Detecting tumors early is key to timely intervention, improving treatment results, and potentially saving lives.

Medical imaging is now a fundamental tool for diagnosis and intervention, offering visual insights into the functionality of organs and tissues. The increasing use of advanced imaging technologies such as Magnetic Resonance Imaging (MRI) and Computed Tomography (CT) has created a pressing need for automated processing of scanned data [7]. These technologies produce accurate images, allowing doctors to accurately detect tumors and initiate appropriate treatment plans. MRI is preferred over CT scans because it has beneficial qualities and does not impact the human body [8].

Diverse approaches have been applied to medical databases, covering MRI images of brain tumors and tumors in other regions of the human body [9].

Artificial intelligence (AI) has experienced significant progress with the emergence of deep learning techniques, especially Convolutional Neural Networks (CNNs).

In image processing, CNNs stand out as the most commonly utilized and effective algorithm in Deep Learning [10], revolutionizing how computers analyze and interpret visual data. CNNs have shown outstanding effectiveness in various computer vision applications, such as image classification,

facial recognition, and object detection. These neural networks use convolutional layers to automatically learn and extract features from images, allowing them to capture diverse structures and patterns. For increased flexibility and customization, ultimately leading to improved performance and outcomes in image analysis, classification, and segmentation tasks, various modified pretrained networks on large datasets like ImageNet have been utilized. These pretrained networks expedite the training process and frequently lead to better performance on subsequent tasks, making them a popular choice in computer vision applications [11]. Modifications to pretrained architectures often involve adding or modifying layers, tuning hyperparameters, or changing activation functions.

A single pretrained CNN model might have limitations in capturing the complete variability and complexity of the data. It may not be easily adaptable to different tasks or domains without extensive retraining or fine-tuning, processes that are often time-consuming and resource-intensive. These limitations may result in the model not encompassing all relevant features or patterns in the data, leading to decreased accuracy in classification and generalization performance on unseen data.

This article proposes an ensemble fusion method that concatenates three different CNN models, to classify brain MR images into four categories: Pituitary tumor, Glioma tumor, Meningioma tumor, and Normal brain. The three pre-trained models serve as deep feature extractors from images. Subsequently, we combine the features extracted from these neural networks to create a synthetic feature, and the dominant features are selected using the PSO algorithm. Finally, a linear kernel SVM classifier is employed for classification.

The major difficulty associated with using linear SVM classifiers is the need to select an optimal value for the penalty parameter C. This tuning parameter C balances the tradeoff between expanding the margin and minimizing errors in the machine learning problem. In this experiment, we chose to use PSO to optimize the parameter C and also to search for relevant features that maximize the score on the testing set.

Our goal is to achieve superior performance in multiclass brain tumor classification by enhancing accuracy and generalization capabilities.

The contributions of this study are briefly:

•Combination of CNN and linear SVM classifier for classifying brain MR images into four categories: Meningioma tumor, Pituitary tumor, Glioma tumor, and Normal brain;

•Fusion of three deep CNNs: VGG19, GoogLeNet, and ResNet50 achieves an even higher accuracy than individual models;

•Employing the PSO algorithm to enhance our ensemble model's performance with a reduced feature subset.

The rest of the document is organized as follows: Sect II details the methodology employed to design the ensemble deep CNN model. The results obtained from the proposed method and comparisons with existing studies are presented in Sect III. Sect IV concludes the paper.

2. RELATED WORK

Recently, ensemble learning has become a significant area of research, particularly in classification tasks. By combining multiple classifiers, ensemble methods strive to enhance performance by leveraging the diversity among individual models.

Goyal et al. [12] presented a new dataset of foot images with diabetic foot ulcers. They introduced a novel labeled dataset, which is a DFU dataset with ground truth labels, for the presence of ischemia and infection. This is the first publicly available dataset with these labels. They developed then an ensemble model by concatenating the features from three CNNs models (ResNet50, InceptionResNet-V2, and Inception-V3) and use SVM classifier to classify infection versus non-infection and ischaemia versus non-ischaemia. This work pioneered recognizing important diabetic foot conditions from images using machine learning. Haq et al. [13] proposed a deep CNN approach for breast cancer detection and classification from mammogram images. The DCNN architecture uses feature fusion from different blocks of three subnetworks, with the last block of each sub-network containing various classifier (sigmoid, SVM, and random forest). An ensemble of the three classifiers is then used in the last block of the DCNN, utilizing majority voting for the final prediction. Wu et al. [14] presented a method called ZipperNet for merging multiple well-trained deep CNN models into a single model for efficient multi-task inference. It generates a sequence of merged models via trade-off between speedup and accuracy drop. The ZipperNet method can achieve up to 3x speedup and memory reduction with less than 3% average accuracy drop across the merged tasks compared to the individual models. Bourennane et al. [15] performed a deep learning model for binary classification of brain tumors using fusion of pre-trained CNNs (EfficientNetB0 and VGG19). This novel model is employed to extract the most features, which are then classified into two categories—tumor or no tumor—using a cubic SVM classifier. The proposed method achieved excellent performance on the Br35H dataset with 99.78% accuracy, 99.78% precision, 99.78% recall, 99.78% specificity and 99.78% F1-score. Remzan et al. [6] combined three CNN models, namely ResNet50, VGG19, and EfficientNetV2B1 to create an ensemble model for extracting features from MR images utilizing a dataset comprising of 5712 MR images, which includes four classes (Pituitary tumor, Meningioma tumor, Glioma tumor, and Normal brain). This ensemble is then combined with the Multilayer Perceptron (MLP) classifier to classify the four categories. The resulting ensemble model achieves 96.67% accuracy, outperforming previous state-of-the-art methods on this dataset. Rostami et al. [16] developed an ensemble Deep CNN for classifying wound images into categories such as diabetic, surgical, and venous ulcers. The ensemble classifier combines two individual classifiers - a patch-wise classifier using a fine-tuned AlexNet on wound patches, and an image-wise classifier using AlexNet trained on whole images. The outputs from these classifiers are then combined in an MLP to achieve high-quality classification performance. Alhichri [17] suggested a novel method called RS-DeepSuperLearner for classifying remote sensing (RS) scenes.

This method fine-tunes and combines five CNN models: InceptionResNet-V2, Inception-V3, VGG16, EfficientNet-B3, and DenseNet-121. Subsequently, a novel deep CNN, termed SuperLearner, is trained on the predicted probability outputs and cross-validation accuracies of these five CNN models to optimally combine their outputs. Juan et al. [18] presented a new approach called SkinFLNet for multi-class skin cancer recognition. It uses a fusion strategy that combines predictions from two deep CNN models (Inception-V3 and

ResNet50) to improve classification accuracy. Furthermore, this approach employs lifelong learning to retrain the model using a merged dataset that includes both newly collected data and a portion of the original data. Pan et al. [19] proposed an ensemble learning method called Wheat Rust Based on Ensemble Learning (WR-EL) for identifying wheat rust diseases from images. It integrates five convolutional neural network (CNN) models - VGG16, ResNet101, ResNet152, DenseNet169, and DenseNet201 using bagging, snapshot ensembling, and the stochastic gradient descent with warm restarts (SGDR) algorithm. Additionally, they proposed the SGDR-S algorithm which is an improved version of the SGDR to improve the F1- scores of leaf rust wheat, stem rust wheat, and healthy wheat. Experiments show WR-EL achieves 92% accuracy, outperforming any single CNN model. Babar et al. [20] developed a feature fusion-based system for brain tumor classification using multiple CNN architectures (AlexNet, ResNet18, DenseNet201, VGG16). The dataset used comprises 3,064 MRI images of glioma, pituitary, and meningioma tumors. The best performing CNNs, AlexNet and DenseNet201, provide features that are fused into a single vector and then classified using SVM and KNN classifiers. This method achieved a maximum accuracy of 92.2% using the SVM classifier on the fused features, surpassing the 85-89% accuracy achieved by individual CNN features.

Salih and Abdulazeez [21] introduced a novel method for classifying brain tumors from MRI images into four distinct categories: Meningioma tumor, Pituitary tumor, Glioma tumor, and Normal brain. This approach combines features from ResNet18 and ResNet50 models, utilizing preprocessing with Gaussian filtering, feature extraction, fusion, and classification with a Softmax classifier. Trained on a dataset with 3264 MRI images, this method achieved 92.47% accuracy, 94.44% recall, 94.37% precision, and a 96.89% F1-score. The fusion model outperformed the individual models and demonstrated competitive performance compared to other recent methods. Patil and Kirange [22] presented an ensemble deep convolutional neural network (EDCNN) to classify brain tumors using MRI images. This method integrates a custom shallow CNN (SCNN) for high-level spatial feature extraction with a fine-tuned VGG16 model for deep feature extraction, followed by feature fusion. This method maintains the spatial information of the tumor and reduces information loss during training. The EDCNN model achieves 97.77% accuracy, outperforming recent methods on the same dataset.

Among the various optimization approaches, Particle Swarm Optimization (PSO) emerges as a promising technique for enhancing the performance of ensemble CNNs.

PSO has the ability to optimize hyperparameters, model weights, and the configuration of the ensemble itself, leading to more accurate and reliable results in various deep learning tasks.

Donuk et al. [23] suggested a method for facial emotion recognition from images employing CNNs, binary particle swarm optimization (BPSO), and SVM. First, a CNN-based network is trained on the Fer+ dataset. Next, the BPSO algorithm is used to select features from the feature vector within the fully connected layer of the trained CNN. The selected features are then classified by SVM. The system attained an accuracy of 85.74% on the FER+ test set, higher than using just CNN (84.28%) or CNN+SVM (84.81%). Rahman et al. [24] presented a novel method for multi-class classification of Acute Lymphoblastic Leukemia (ALL) using machine deep learning techniques. PSO and Cat Swarm

Optimization (CSO) were used to find the best features. Support Vector Classifier (SVC) classifiers were then utilized to perform the multi-class malignant classification. This method Achieved highly accurate multi-class classification of different ALL subtypes from blood cell images. Arianti et al. [25] used an ensemble of diverse CNNs to improve classification accuracy for detecting abnormalities like polyps and ulcers in endoscopy images. PSO is utilized to find the best weights for each model in the ensemble, giving more influence to stronger models. Experiments on the Kvasir dataset showed the proposed weighted ensemble with diversity improved classification accuracy compared to a single CNN or a standard averaging ensemble. Islam et al. [26] developed an automated approach to classify seven medicinal plants from Bangladesh using smartphone-captured plant images. This approach uses a cascaded neural network architecture combining a pre-trained ResNet50 CNN for feature extraction, PSO for feature selection, and a SVM classifier. The ResNet50-PSO-SVM network achieved 99.60% accuracy outperforming previous methods that relied on leaf images.

3. PROPOSED APPROACH

3.1 Dataset

The dataset employed. in this experiment is comprised of data from three different sources: figshare, Br35H, and the SARTAJ dataset, made publicly available by Masoud Nickparvar [27].

This merged dataset comprises 7023 MRI images of the human brain. For our study, we used 5712 images, classified into four categories: pituitary (1457 images), meningioma (1339 images), glioma (1321 images), and no tumor (1595 images). The no tumor images were sourced from the Br35H dataset. The image sizes vary within the dataset; therefore, the images were resized to 224×224 pixels to match the input size requirements of the three deep learning networks used.

The dataset was split into two groups: 80% allocated for training and 20% reserved for testing. Sample images collected from the database are displayed in Figure 1.
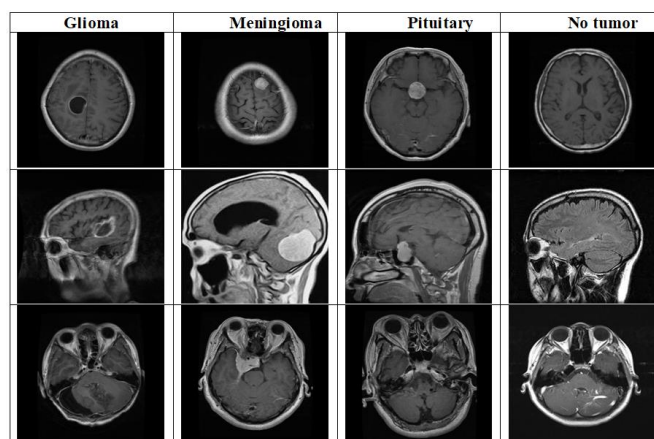
Figure 1. Sample brain MRI images from the dataset used in this study [27]

3.2 Deep neural network models

This study employs three CNN models: VGG19, ResNet50,

and GoogLeNet to sequentially extract feature vectors from brain MR image.

3.2.1 VGG19

VGG19 is a pre-trained CNN model initially trained on 1000 classes from the ImageNet dataset [28]. It accepts 224×224-pixel images as input and comprises of 19 weight layers: 16 convolutional layers followed by 3 fully connected layers, totaling approximately 144 million parameters.

The convolutional layers use small 3×3 convolutional filters, starting at 64 in the first layer and doubling in number after each max-pooling layer, up to a maximum of 512 filters. Its architecture has demonstrated exceptional performance across a various image classification task [29]. Despite its simplicity, VGG19 tends to generalize well when provided with sufficient training data [30].
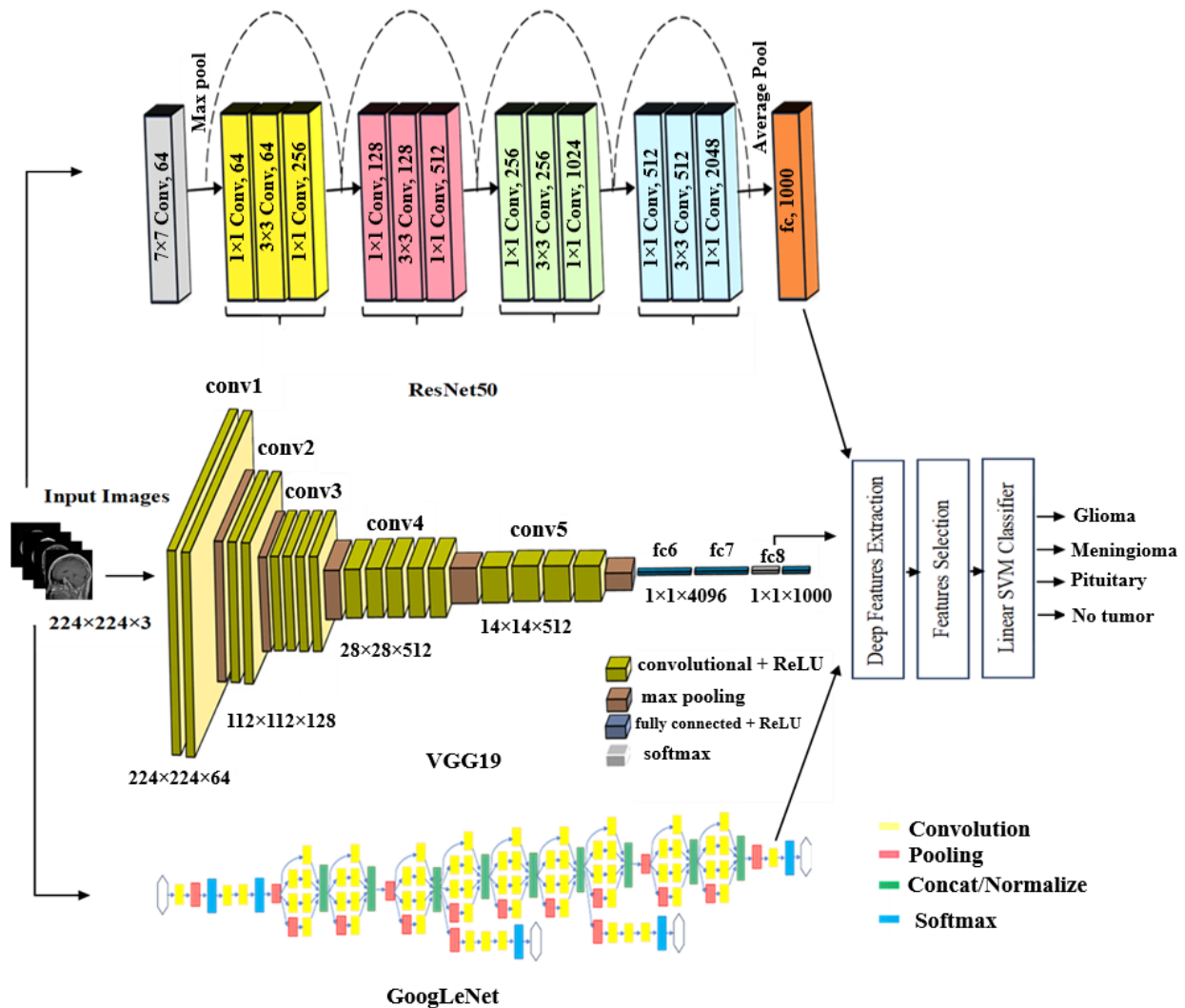
3.2.2 ResNet50

ResNet50 is a CNN that is 50 layers deep. It was trained on 1.28 million training images across 1000 classes and has an image input size of 224 by 224. The architecture of ResNet50 consists of four key parts: convolution layers for feature extraction, convolution blocks comprising multiple convolution layers with normalization and activation functions for high-level feature extraction, residual blocks that provide shortcut connections to mitigate the vanishing gradient problem, and fully connected layers that make predictions based on the extracted features. The residual layers present in ResNet50 are crucial for transferring large gradient values to their prior adjacent layers [31]. ResNet50 has indeed demonstrated remarkable efficiency in solving the vanishing gradient problem compared to previous methods. This problem occurs due to the iterative multiplication of derivative values in the first layers, which causes these values to decrease and subsequently reduces network accuracy in deep learning [32, 33].

3.2.3 GoogLeNet

GoogLeNet is a deep learning model capable of classifying patterns among approximately 1000 images. It comprises of 22 layers and incorporates 9 inception modules. These modules enable the network to capture features at different scales and resolutions, improving its capability to recognize diverse patterns in images. The 1×1 convolutions at the module's bottom reduce the number of inputs, leading to a dramatic decrease in computational cost [34]. Furthermore, GoogLeNet utilizes the global average pooling layer rather than a fully connected layer, which decreases the number of parameters in the network. In general, it stands out as an efficient and precise deep learning architecture, significantly influencing the development of subsequent models in the field [35].

Figure 2. Summary of the suggested framework for categorizing brain tumors using an ensemble of VGG19, ResNet50, and Inception V1 with an SVM classifier

Integrating VGG19, GoogLeNet, and ResNet50 leverages their unique feature extraction strengths: VGG19 captures fine-grained textures and tumor edges, essential for identifying subtle tumor patterns; GoogLeNet's inception modules enhance feature diversity by efficiently extracting both local and global information; and ResNet50's skip connections facilitate deeper feature learning, preventing vanishing gradients and improving hierarchical feature extraction. This ensemble approach provides a more robust feature representation, mitigating individual model weaknesses and enhancing classification performance.

The proposed approach involves fusing the outputs of the fully connected layers of VGG19, ResNet50, and GoogLeNet into a synthetic feature set, as illustrated in Figure 2. This results in a set comprising 3000 features, with 1000 features contributed by each model. The dominant features selected using the PSO algorithm are then fed into a linear SVM classifier for classifying 4 classes.

3.3 PSO

In this subsection, we describe the functional principles of the feature selection model included in this study, specifically PSO.

PSO is a robust and efficient algorithm that mimics the behavior of birds searching for food [36]. It has proven successful in addressing search and optimization problems across a range of domains.

In PSO algorithm, each solution is represented as a particle, with the swarm comprising all these particles. Each particle's movement is influenced by three factors: its current velocity, its best position found thus far ($pbest$), and the swarm's best position thus far ($gbest$). These factors guide the particles in exploring the search space effectively and facilitate convergence toward the optimal solution. Initially, particles are randomly positioned and assigned random velocities. Their fitness is evaluated, and pbest and gbest are updated if the new fitness values are better than previously recorded values. Each particle's velocity and position are updated iteratively based on the following equations:

$$v_{t+1} = \omega v_t + c_1 r_1 (pbest_t - x_t) + c_2 r_2 (gbest_t - x_t) \tag{1}$$

$$x_{t+1} = x_t + v_{t+1} \tag{2}$$

where, t refers to number of iterations. x_t indicates the particle's position at time t, v_t denotes the current velocity at time t, $pbest_t$ is the personal best position, $gbest_t$ is the global best position in the swarm thus far, ω denotes the inertia weight, r_1 and r_2 are uniformly distributed random numbers between 0 and 1, and c_1 and c_2 are cognitive and social coefficients. This process is repeated until a predefined stopping criterion, such as a satisfactory fitness level or a maximum number of iterations, is met.

PSO is easy to implement and requires fewer parameter adjustments [37]. It provides stable results in parameter optimization compared to other methods [38]. PSO is a crucial element in ML for SVM parameter adjustment, optimizing weights in back-propagation neural networks, and achieving better results compared to traditional backpropagation methods. PSO effectively balances exploration and exploitation, making it suitable for high-dimensional optimization tasks such as feature selection and hyperparameter tuning [39]. This balance enables PSO to efficiently navigate large search spaces. Its efficiency in navigating search spaces via particle position updates reduces computational costs compared to exhaustive methods such as grid search. PSO also converges more quickly than genetic algorithms by avoiding complex genetic operations such as crossover and mutation. Unlike probabilistic models such as Bayesian optimization, which often face challenges in high-dimensional spaces due to increased statistical and computational complexity, PSO does not rely on gradient information. These features make PSO a valuable tool for various optimization tasks, particularly in complex search spaces.

The role of PSO is to determine the values the penalty parameter C and fs: a real number whose binary representation is used as a mask to select a subset of features which maximize the value of the objective function.

In this experiment, the PSO algorithm parameters are set as follows:

Initial population: In the PSO algorithm, a randomly initialized population of potential solutions was created, with the initial population size set to 30.

Fitness function: The 10-fold-Cross-Validation method will be utilized to evaluate the performance of the SVM model and the wrapper approach will be used to select the most important feature subset that maximize the fitness function value. A vector composed of 3000 bits, which is the binary representation of fs, is used as a mask. This mask includes all bits with a value of 1, representing the features used in the training phase, and ignores all features masked by bits set to 0. Feature subset selection and the penalty parameter C should be optimized simultaneously to find the best solution. In this study, overall classification accuracy is utilized as the fitness function.

Maximum generations: The total number of generations was set to 50.

Termination criterion: Can be either a lack of improvement in the population over a specified number of generations or reaching an upper limit on the number of generations.

In this experiment, the tested ranges of values for parameters C and fs were set to [0.001, 10] and [0.1, 100], respectively.

3.4 SVM based classification

SVM is a supervised learning algorithm used for solving classification and regression problems. It was enhanced by Cortes and Vapnik in 1995 for binary classification. The algorithm was later developed and generalized for multiclass and nonlinear datasets [40].

This classifier seeks to identify the optimal separating hyperplane that achieves the maximum margin. The margin is defined as the distance between the hyperplane and the closest data points, termed support vectors. The hyperplane is defined by the equation:

$$w^T x_i + b = 0 \tag{3}$$

The weight vector w and bias term b are parameters that define the decision boundary. These parameters are calculated as follows:

$$\begin{cases} y_i(w^T x_i + b) \geq 1 - \xi_i \; if \; y_i = +1 \\ y_i(w^T x_i + b) \leq 1 - \xi_i \; if \; y_i = -1 \end{cases} \tag{4}$$

The last two constraints can be combined into:

$$y_i(w^T x_i + b) \geq 1 - \xi_i \qquad (5)$$

where, ξ_i is the slack variable. The optimal hyperplane can then be found as follows:

$$\Phi(w, \xi) = \frac{1}{2} w^T w + C \sum_{i=1}^{N} \xi_i \qquad (6)$$

The parameter $C > 0$ regulates the balance between maximizing the margin and minimizing classification errors.

It significantly impacts the efficiency and performance of the SVM classifier [41].

Using the method of Lagrangian multiplier, the solution the optimal classification hyperplane can be formulated as follows:

$$Q(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j K(x_i, x_j) \qquad (7)$$

where, $\alpha_i = (\alpha_1, \alpha_2, \dots, \alpha_N)$ is the vector of Lagrange multipliers, most multipliers satisfy the condition $\alpha_i = 0$ with only the sample for which $\alpha_i \neq 0$ being considered a support vector. $K(x_i, x_j)$ is the kernel function in SVMs that maps data, which is not linearly separable, into a higher-dimensional feature space where linear separation may become possible. Currently, the most widely used SVM kernel functions are linear, polynomial, radial basis function (RBF), and sigmoid, among others [42]. In this study, the linear SVM classifier is integrated into the final layer of the fully connected CNN to enhance efficiently fit the date length to turn the kernel [33]. Figure 3 provides a detailed depiction of the process employed by the proposed model.
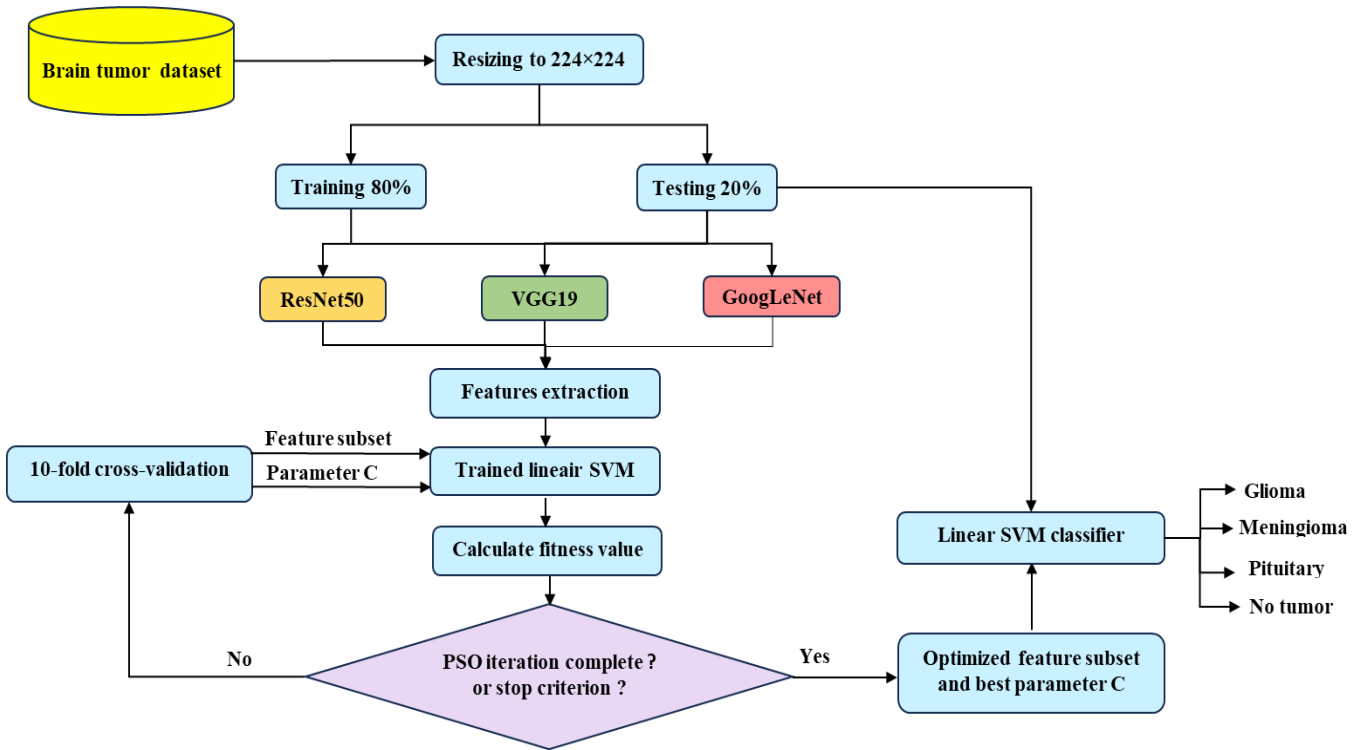
Figure 3. Flowchart of the proposed ensemble approach for brain image classification

3.5 Performance metrics

In this research, we evaluated the performance of each model using accuracy, precision, recall, and F1-score metrics. These classification performance measures are derived from the four values of the confusion matrix: True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). Additionally, we plotted the AUC (Area Under the Curve) for the ROC (Receiver Operating Characteristics) curve. The formulas for these evaluation metrics are given in Eqs. (8) to (12).

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \qquad (8)$$

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

$$\text{Specificity} = \frac{TN}{TN + FP} \qquad (10)$$

$$\text{Precision} = \frac{TP}{TP + FP} \qquad (11)$$

$$\text{F1} - \text{Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \qquad (12)$$

4. EXPERIMENTAL RESULTS AND DISCUSSION

This study aimed to present a method for classifying brain MR images using machine learning technique. Various models, including our ensemble approach, were trained and evaluated. This section examines whether the proposed method has improved classification accuracy. The 10-fold

cross-validation method was selected to evaluate the performance of the SVM model in this experiment.

4.1 Performance analysis

Performance metrics such as Accuracy, Precision, Specificity, Sensitivity, and F1-score for three deep CNNs: VGG19, GoogLeNet, and ResNet50 are displayed in Table 1. It also evaluates their combinations for classifying brain MR images into four categories: Meningioma tumor, Pituitary tumor, Glioma tumor, and Normal brain.

Table 1. Performance metrics of proposed model and individual models with their ensembles for brain MRI classification

Model	Feature Subset	Accuracy	Precision	Specificity	Sensitivity	F1-score
GoogLeNet	1000	98.07	98.07	99.36	98.07	98.07
VGG19	1000	97.55	97.55	99.18	97.55	97.55
ResNet50	1000	98.25	98.25	99.42	98.25	98.25
GoogLeNet+ VGG19+ ResNet50	3000	98.77	98.77	99.60	98.77	98.77
PSO+GoogLeNet+ VGG19+ ResNet50	1740	99.30	99.30	99.77	99.30	99.30
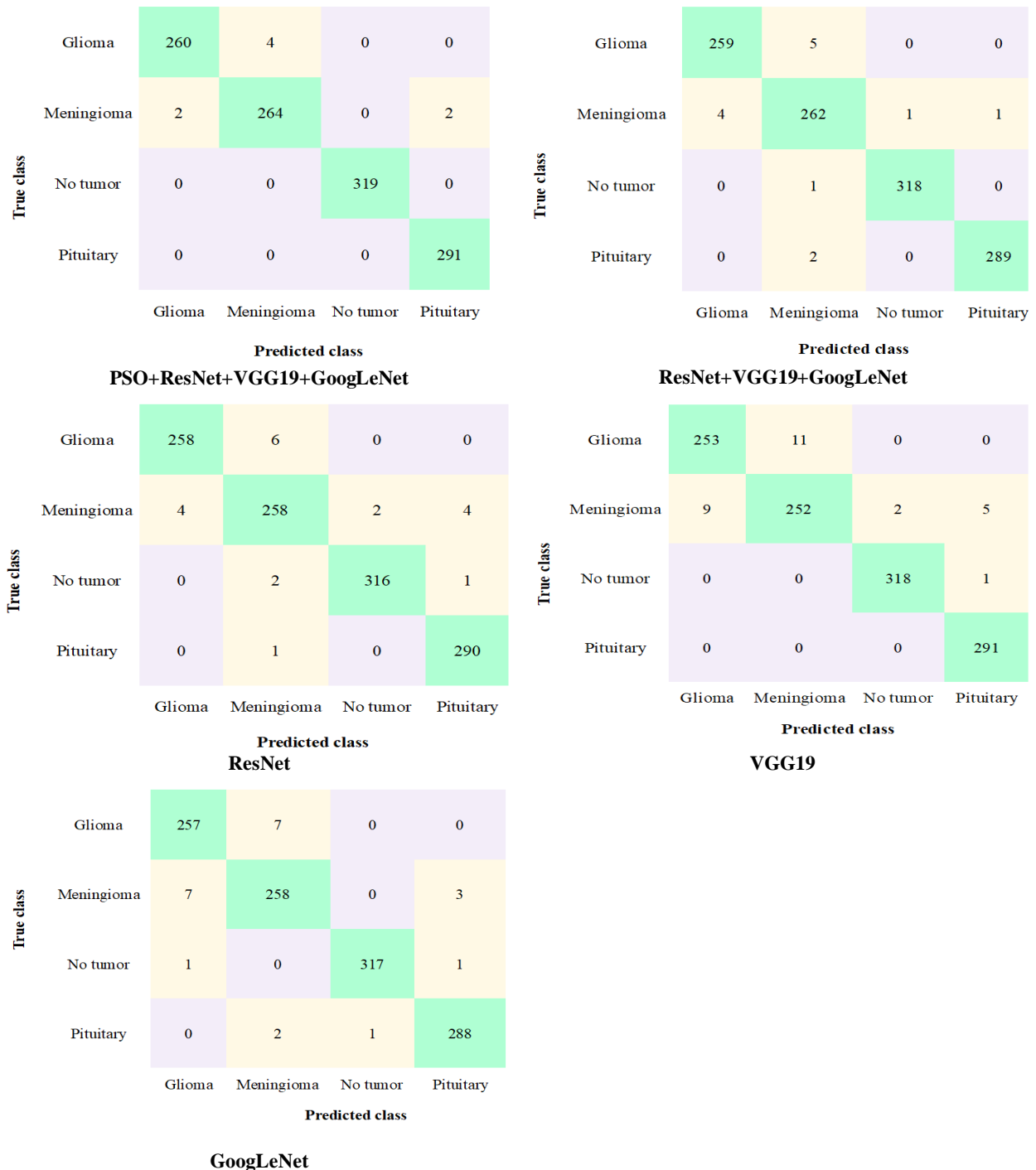
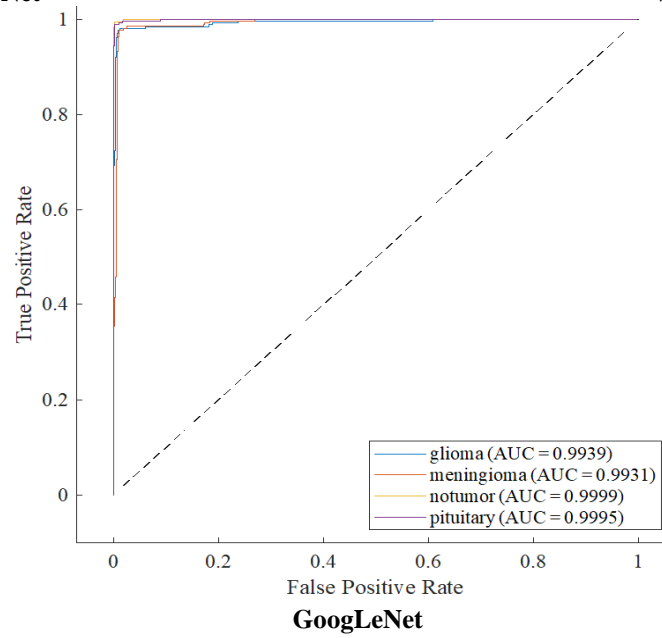**Figure 4.** Confusion matrices comparing the proposed model with other models
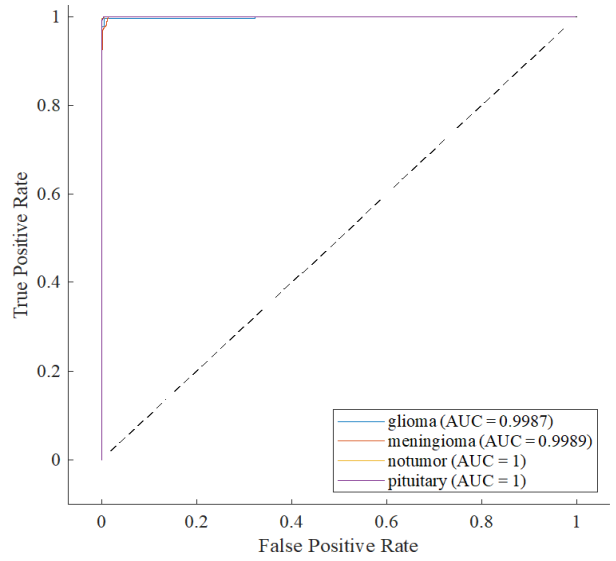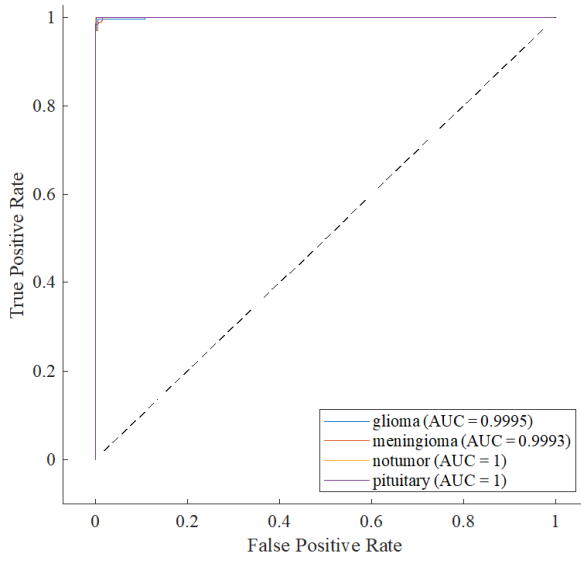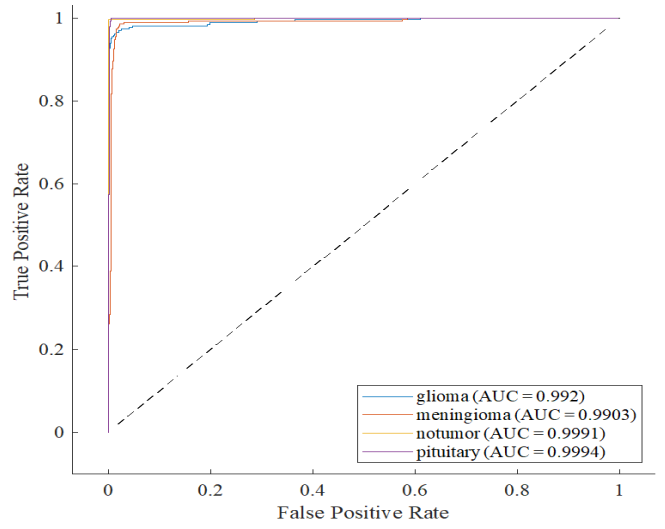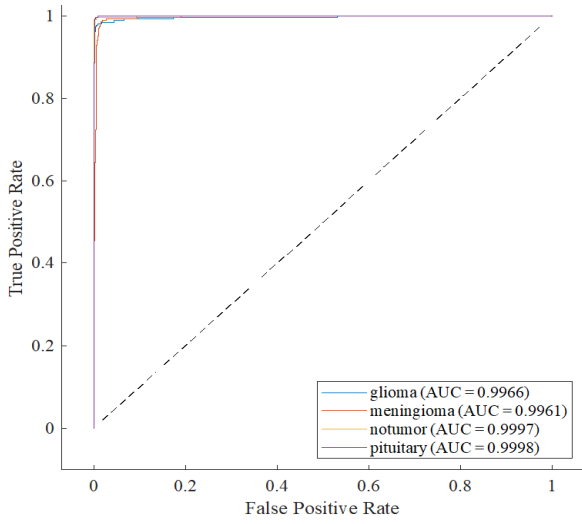
Figure 5. ROC curve comparison for the proposed model and alternative models

Table 2. Comparison of accuracy for various brain MRI classification methods from recent studies

Reference	Method	Brain MRI Dataset	Accuracy (%)
Babar et al. [20]	AlexNet+DenseNet201	3064 images	92.20
Salih and Abdulazeez [21]	ResNet18+ResNet50	3264 images	92.47
Patil and Kirange [22]	SCNN+VGG16	3064 images	97.77
Remzan et al. [6]	ResNet50 + VGG19 + EfficientNetV2B1+ MLP	5712 images	96.67
Our proposed approach	PSO+GoogLeNet+ VGG19+ ResNet50	5712 images	99.30

GoogLeNet, VGG19, and ResNet50 each demonstrate strong individual performance. Among these models, ResNet50 achieves maximum scores across all metrics, including accuracy (98.25%), precision (98.25%), specificity (99.42%), sensitivity (98.25%), and F1-score (98.25%). The inclusion of skip connections in ResNet50 effectively addresses the vanishing gradient problem, allowing for the training of deeper networks. This capability enables the model to capture both fine details and global structures in MRI images, leading to superior classification performance across various tumor types. Notably, all individual models exhibit very high specificity, indicating their robust ability to accurately identify negative cases, with ResNet50 leading at 99.42%. Combining these models with a subset of 3000 features significantly improves overall performance. The combined model achieves uniformly higher scores across all metrics compared to the individual models, with accuracy, precision, sensitivity, and F1-score all at 98.77%, along with a particularly high specificity of 99.60%.

Each of the models VGG19, GoogLeNet, and ResNet50 employs distinct strategies for feature extraction across different scales. By combining these models into an ensemble, it becomes feasible to incorporate all relevant features or patterns present in the data. This approach was able to enhance the accuracy of classification performance and bolster the models' capability to apply learned patterns to new, unseen data.

Fine-tuning the penalty parameter C in SVM establishes a well-balanced decision boundary, thereby reducing misclassification errors. PSO, we efficiently selected the optimal C value from a vast parameter space. This approach, combined with a reduced feature subset of 1,740 features, significantly enhanced our model's performance.

The method achieves maximum scores across all metrics, with accuracy, sensitivity, precision, and F1-score all at 99.30%, and particularly high specificity at 99.70%.

Figure 4 shows the confusion matrices. Additionally, we plotted the ROC curves for the individual models, GoogLeNet, VGG19, and ResNet50, as well as for the ensemble models, GoogLeNet + VGG19 + ResNet50 and PSO + GoogLeNet + VGG19 + ResNet50, which are shown in Figure 5.

The confusion matrices indicate that PSO+GoogLeNet+ VGG19+ ResNet50 has a high overall performance compared to other models, with very accurate classification for each class and minimal misclassifications. This suggests a robust model with strong generalization capabilities.

The results illustrate the performance of various models and their combinations using the AUC metric, which measures their effectiveness in differentiating between classes.

The results highlight the performance of individual models and the significant improvements achieved through ensemble methods and optimization techniques.

While all individual models perform exceptionally well, with AUCs above 0.99, ResNet50 stands out slightly above GoogLeNet and VGG19 with an AUC of 0.9956. Combining these models into an ensemble (GoogLeNet + VGG19 +

ResNet50) led to improved performance, with an AUC of 0.9987. The optimized ensemble, which included PSO alongside the three models, achieved the best overall performance with an AUC of 0.9995, indicating that the PSO-optimized ensemble is extremely proficient in distinguishing between classes.

4.2 Comparison with previous studies

We evaluated our proposed method against prior studies that examined identical brain tumor types but utilized different CNN models, as shown in Table 2.

Our proposed approach achieves the highest accuracy of 99.30%, significantly higher than other methods. It is also noted that both our study and Ramzan et al.'s study [6] utilized more than two models with a larger dataset of 5712 images, which may contribute to the higher performance observed. In contrast, the other methods employed fewer models. Babar et al. [20] used AlexNet + DenseNet201 with 3064 images, Salih and Abdulazeez [21] used ResNet18 + ResNet50 with 3264 images, and Patil and Kirange [22] used SCNN + VGG16 with 3064 images. These smaller datasets and fewer models may explain their relatively lower accuracy.

5. CONCLUSIONS

In this paper, we introduced an ensemble fusion method that integrates three deep CNNs: VGG19, GoogLeNet, and ResNet for classifying brain MR images into four categories: Meningioma tumor, Pituitary tumor, Glioma tumor, and Normal brain. The three pre-trained models serve as deep feature extractors from images provide features that are fused into a single vector and then classified using linear SVM classifiers. We employed the PSO algorithm to optimize the parameter C for linear SVM and select the best features from multiple CNNs, effectively enhancing classification accuracy on the testing set. The combined model consistently outperformed individual models across all evaluation metrics., with accuracy, precision, sensitivity, and F1-score all reaching 98.77%, and an exceptional specificity of 99.60%. Our results indicate that the fusion of VGG19, GoogLeNet, and ResNet50 harnesses their combined advantages, leading to improved feature extraction, enhancing accuracy, and bolstering generalization capabilities. This approach culminates in a more robust and effective model. Furthermore, experimental findings underscored that PSO improved the SVM algorithm's performance by finding the optimal parameter C from a large set, enabling our ensemble approach to significantly enhance performance with a reduced feature subset of 1740 features. The method excelled across all evaluated metrics, with accuracy, precision, sensitivity, and F1-score all achieving 99.30%, alongside an exceptionally high specificity of 99.70%.

Although our model effectively handles noisy images by leveraging the complementary strengths of VGG19,

GoogLeNet, and ResNet50, further analysis is required to assess its performance on extremely degraded images. Future research should focus on developing and validating real-time diagnostic tools or integrating the proposed methods into clinical workflows, which could facilitate the transition from research to practical application, thereby improving patient diagnosis and treatment outcomes. Additionally, investigating the applicability of the ensemble CNN and PSO framework to other types of cancers or medical conditions would help assess its versatility and effectiveness in various diagnostic scenarios. However, it's important to note that ensemble approaches increase computational complexity, necessitating high-performance GPUs for real-time inference. Addressing these computational challenges is essential for the effective implementation of such advanced diagnostic systems.

REFERENCES

[1] Vijithananda, S.M., Jayatilake, M.L., Hewavithana, B., Gonçalves, T., et al. (2022). Feature extraction from MRI ADC images for brain tumor classification using machine learning techniques. Biomedical Engineering Online, 21(1): 52. https://doi.org/10.1186/s12938-022-01022-6

[2] Rao, B.C., Raju, K., Babu, G.R., Pittala, C.S. (2023). An improved GABOR wavelet transform and rough k-means clustering algorithm for MRI BRAIN tumor image segmentation. Multimedia tools and applications, 82(18): 28143-28164. https://doi.org/10.1007/s11042-023-14485-z

[3] Eali, S.N.J., Bhattacharyya, D., Nakka, T.R., Hong, S.P. (2022). A novel approach in bio-medical image segmentation for analyzing brain cancer images with U-NET semantic segmentation and TPLD models using SVM. Traitement Du Signal, 39(2): 419-430. https://doi.org/10.18280/ts.390203

[4] Reyes, D., Sánchez, J. (2024). Performance of convolutional neural networks for the classification of brain tumors using magnetic resonance imaging. Heliyon, 10(3): e25468. https://doi.org/10.1016/j.heliyon.2024.e25468

[5] Bacak, A., Şenel, M., Günay, O. (2023). Convolutional neural network (CNN) prediction on meningioma, glioma with Tensorflow. International Journal of Computational and Experimental Science and Engineering, 9(2): 197-204. https://doi.org/10.22399/ijcesen.1306025

[6] Remzan, N., Hachimi, Y.E., Tahiry, K., Farchi, A. (2024). Ensemble learning based-features extraction for brain MR images classification with machine learning classifiers. Multimedia Tools and Applications, 83(19): 57661-57684. https://doi.org/10.1007/s11042-023-17213-9

[7] Müller, D., Kramer, F. (2021). MISCNN: A framework for medical image segmentation with convolutional neural networks and deep learning. BMC Medical Imaging, 21: 12. https://doi.org/10.1186/s12880-020-00543-7

[8] Kumar, D.M., Satyanarayana, D., Prasad, M.G. (2021). An improved Gabor wavelet transform and rough K-means clustering algorithm for MRI brain tumor image segmentation. Multimedia Tools and Applications, 80(5): 6939-6957. https://doi.org/10.1007/s11042-020-09635-6

[9] Badža, M.M., Barjaktarović, M.Č. (2020). Classification of brain tumors from MRI images using a convolutional neural network. Applied Sciences, 10(6): 1999. https://doi.org/10.3390/app10061999

[10] Allugunti, V.R. (2022). A machine learning model for skin disease classification using convolution neural network. International Journal of Computing, Programming and Database Management, 3(1): 141-147. https://doi.org/10.33545/27076636.2022.v3.i1b.53

[11] Saleh, A.Y., Chin, C.K., Penshie, V., Al-Absi, H.R.H. (2021). Lung cancer medical images classification using hybrid CNN-SVM. International Journal of Advances in Intelligent Informatics, 7(2): 151-162. https://doi.org/10.26555/ijain.v7i2.317

[12] Goyal, M., Reeves, N.D., Rajbhandari, S., Ahmad, N., Wang, C., Yap, M.H. (2020). Recognition of ischaemia and infection in diabetic foot ulcers: Dataset and techniques. Computers in Biology and Medicine, 117: 103616. https://doi.org/10.1016/j.compbiomed.2020.103616

[13] Haq, I.U., Ali, H., Wang, H.Y., Lei, C., Ali, H. (2022). Feature fusion and Ensemble learning-based CNN model for mammographic image classification. Journal of King Saud University-Computer and Information Sciences, 34(6): 3310-3318. https://doi.org/10.1016/j.jksuci.2022.03.023

[14] Wu, C.E., Lee, J.H., Wan, T.S., Chan, Y.M., Chen, C.S. (2020). Merging well-trained deep CNN models for efficient inference. In 2020 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), Auckland, New Zealand, pp. 1594-1600.

[15] Bourennane, M., Naimi, H., Mohamed, E. (2024). Deep feature extraction with Cubic-SVM for classification of brain tumor. Studies in Engineering and Exact Sciences, 5(1): 19-35. https://doi.org/10.54021/seesv5n1-002

[16] Rostami, B., Anisuzzaman, D.M., Wang, C., Gopalakrishnan, S., Niezgoda, J., Yu, Z. (2021). Multiclass wound image classification using an ensemble deep CNN-based classifier. Computers in Biology and Medicine, 134: 104536. https://doi.org/10.1016/j.compbiomed.2021.104536

[17] Alhichri, H. (2023). RS-DeepSuperLearner: Fusion of CNN ensemble for remote sensing scene classification. Annals of GIS, 29(1): 121-142. https://doi.org/10.1080/19475683.2023.2165544

[18] Juan, C.K., Su, Y.H., Wu, C.Y., Yang, C.S., Hsu, C.H., Hung, C.L., Chen, Y.J. (2023). Deep convolutional neural network with fusion strategy for skin cancer recognition: model development and validation. Scientific Reports, 13(1): 17087. https://doi.org/10.1038/s41598-023-42693-y

[19] Pan, Q., Gao, M., Wu, P., Yan, J., AbdelRahman, M.A. (2022). Image classification of wheat rust based on ensemble learning. Sensors, 22(16): 6047. https://doi.org/10.3390/s22166047

[20] Babar, M.A., Zeeshan, F., Waheed, S., Amin, R. (2023). A feature fusion based system for brain tumor classification. International Journal of Emerging Engineering and Technology, 2(1): 24-29. https://doi.org/10.57041/ijeet.v2i1.893

[21] Salih, M.S., Abdulazeez, A.M. (2024). A fusion-based deep approach for enhanced brain tumor classification.

Journal of Soft Computing and Data Mining, 5(1): 183-193. https://doi.org/10.30880/jscdm.2024.05.01.015

[22] Patil, S., Kirange, D. (2023). Ensemble of deep learning models for brain tumor detection. Procedia Computer Science, 218: 2468-2479. https://doi.org/10.1016/j.procs.2023.01.222

[23] Donuk, K., Arı, A., Özdemir, M.F., Hanbay, D. (2023). Deep feature selection for facial emotion recognition based on BPSO and SVM. Politeknik Dergisi, 26(1): 131-142. https://doi.org/10.2339/politeknik.992720

[24] Rahman, W., Faruque, M.G.G., Roksana, K., Sadi, A.S., Rahman, M.M., Azad, M.M. (2023). Multiclass blood cancer classification using deep CNN with optimized features. Array, 18: 100292. https://doi.org/10.1016/j.array.2023.100292

[25] Arianti, D., Abdullah, A., Sahran, S. (2024). Weighted PSO ensemble using diversity of CNN classifiers and color space for endoscopy image classification. International Journal of Advanced Computer Science & Applications, 15(3): 1137-1144. https://doi.org/10.14569/IJACSA.2024.01503113

[26] Islam, M.T., Rahman, W., Hossain, M.S., Roksana, K., et al. (2024). Medicinal plant classification using particle swarm optimized cascaded network. IEEE Access, 12: 42465-42478. https://doi.org/10.1109/ACCESS.2024.3378262

[27] Brain Tumor MRI Dataset. https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset?select=Training.

[28] Kandhro, I.A., Manickam, S., Fatima, K., Uddin, M., Malik, U., Naz, A., Dandoush, A. (2024). Performance evaluation of E-VGG19 model: Enhancing real-time skin cancer detection and classification. Heliyon, 10(10): e31488. https://doi.org/10.1016/j.heliyon.2024.e31488

[29] Girish, D. N., Priyanka, M. (2023). Tire imprint identification and classification using VGG19. In International Conference on Computer & Communication Technologies, Warangal, India, pp. 73-94. https://doi.org/10.1007/978-981-99-9704-6_7

[30] Kolla, B.S., Reddy, B.R., Sahithi, S.V., Madala, L.P. (2023). Comparative analysis of VGG19, ResNet50, and GoogLeNet inception models for BCI. Preprint. https://doi.org/10.21203/rs.3.rs-3511460/v1

[31] Behar, N., Shrivastava, M. (2022). ResNet50-based effective model for breast cancer classification using histopathology images. Computer Modeling in Engineering & Sciences, 130(2): 823-839. https://doi.org/10.32604/cmes.2022.017030.

[32] Abdallah, S.E., Elmessery, W.M., Shams, M.Y., Al-Sattary, N.S.A., Abohany, A.A., Thabet, M. (2023). Deep learning model based on ResNet-50 for beef quality classification. Information Sciences Letters, 12(1): 289-297. https://doi.org/10.18576/isl/120124

[33] Duan, Z., Wang, F., Wang, B., Luo, G., Jiang, Z. (2024). An adapted ResNet-50 architecture for predicting flow fields of an underwater vehicle. IEEE Access, 12: 66398-66407. https://doi.org/10.1109/ACCESS.2024.3399077

[34] Dai, M., Sun, W., Wang, L., Dorjoy, M.M.H., et al. (2023). Pepper leaf disease recognition based on enhanced lightweight convolutional neural networks. Frontiers in Plant Science, 14: 1230886. https://doi.org/10.3389/fpls.2023.1230886

[35] Gummaraju, A., Shenoy, A.K., Pai, S.N. (2023). Performance comparison of machine learning models for handwritten Devanagari numerals classification. IEEE Access, 11: 133363-133371. https://doi.org/10.1109/ACCESS.2023.3336912

[36] Cuong-Le, T., Nghia-Nguyen, T., Khatir, S., Trong-Nguyen, P., Mirjalili, S., Nguyen, K.D. (2022). An efficient approach for damage identification based on improved machine learning using PSO-SVM. Engineering with Computers, 38(4): 3069-3084. https://doi.org/10.1007/s00366-021-01299-6

[37] Al Bataineh, A., Manacek, S. (2022). MLP-PSO hybrid algorithm for heart disease prediction. Journal of Personalized Medicine, 12(8): 1208. https://doi.org/10.3390/jpm12081208

[38] Rahayu, E.S., Ma'arif, A., Cakan, A. (2022). Particle swarm optimization (PSO) tuning of PID control on DC motor. International Journal of Robotics and Control Systems, 2(2): 435-447. https://doi.org/10.31763/ijrcs.v2i2.476

[39] Elsedimy, E.I., AboHashish, S.M., Algarni, F. (2024). New cardiovascular disease prediction approach using support vector machine and quantum-behaved particle swarm optimization. Multimedia Tools and Applications, 83(8): 23901-23928. https://doi.org/10.1007/s11042-023-16194-z

[40] Ozaltin, O., Yeniay, O. (2023). A novel proposed CNN–SVM architecture for ECG scalograms classification. Soft Computing, 27(8): 4639-4658. https://doi.org/10.1007/s00500-022-07729-x

[41] Djemai, M., Guerti, M. (2022). A genetic algorithm-based support vector machine model for detection of hearing thresholds. Australian Journal of Electrical and Electronics Engineering, 19(2): 194-201. https://doi.org/10.1080/1448837X.2021.2023080

[42] Khairandish, M.O., Sharma, M., Jain, V., Chatterjee, J.M., Jhanjhi, N.Z. (2022). A hybrid CNN-SVM threshold segmentation approach for tumor detection and classification of MRI brain images. IRBM, 43(4): 290-299. https://doi.org/10.1016/j.irbm.2021.06.003