Advancing Brain MRI Image Classification: Integrating VGG16 and ResNet50 with a Multi-Verse Optimization Method

Nazanin Tataei Sarshar [1], Soroush Sadeghi [2], Mohammadreza Kamsari [3], Mahrokh Avazpour [4], Saeid Jafarzadeh Ghoushchi [5] and Ramin Ranjbarzadeh [6,*]

[1] Department of Engineering, Islamic Azad University, Tehran North Branch, Tehran 1584743311, Iran; ab.tataee@gmail.com
[2] School of Electrical and Computer Engineering, University of Tehran, Tehran 1417935840, Iran; soroush.sadeghi@ut.ac.ir
[3] Faculty of Electrical Engineering, Malek-Ashtar University of Technology (MUT), Esfahan 83154/115, Iran; mohammadrezakamsari@gmail.com
[4] Radio and Optics Communication Laboratory, School of Electronic Engineering, Dublin City University, D09 V209 Dublin, Ireland; mahrokh.avazpour@dcu.ie
[5] Faculty of Industrial Engineering, Urmia University of Technology, Urmia 5756151818, Iran; s.jafarzadeh@uut.ac.ir
[6] School of Computing, Faculty of Engineering and Computing, Dublin City University, D09 V209 Dublin, Ireland
* Correspondence: ramin.ranjbarzadehkondrood2@mail.dcu.ie

Abstract: Background/Objectives: The accurate categorization of brain MRI images into tumor and non-tumor categories is essential for a prompt and effective diagnosis. This paper presents a novel methodology utilizing advanced Convolutional Neural Network (CNN) designs to tackle the complexity and unpredictability present in brain MRI data. **Methods:** The methodology commences with an extensive preparation phase that includes image resizing, grayscale conversion, Gaussian blurring, and the delineation of the brain region for preparing the MRI images for analysis. The Multi-verse Optimizer (MVO) is utilized to optimize data augmentation parameters and refine the configuration of trainable layers in VGG16 and ResNet50. The model's generalization capabilities are significantly improved by the MVO's ability to effectively balance computational cost and performance. **Results:** The amalgamation of VGG16 and ResNet50, further refined by the MVO, exhibits substantial enhancements in classification metrics. The MVO-optimized hybrid model demonstrates enhanced performance, exhibiting a well-calibrated balance between precision and recall, rendering it exceptionally trustworthy for medical diagnostic applications. **Conclusions:** The results highlight the effectiveness of MVO-optimized CNN models for classifying brain tumors in MRI data. Future investigations may examine the model's applicability to multiclass issues and its validation in practical clinical environments.

Keywords: brain tumor classification; MRI image analysis; optimization; deep learning; data augmentation; multi-verse optimizer

1. Introduction

Brain tumors, one of the most daunting medical diagnoses, present a formidable challenge in healthcare. These abnormal cell growths in the brain can have devastating effects on health and well-being [1–4]. Early and accurate recognition is paramount for effective treatment planning and improved prognosis. Traditionally, the diagnosis of brain tumors relies heavily on the analysis of MRI images by experts or radiologists. However, this manual process is not only labor intensive but also prone to human error. The advancement of automated classification methods, particularly through machine learning, offers a transformative solution, promising greater accuracy, efficiency, and consistency in diagnosing brain tumors [4,5]. Machine learning (ML), a subgroup of artificial intelligence

(AI), involves the development of techniques that enable computers to learn from data and make decisions without being explicitly programmed [6–8]. It encompasses numerous methods that allow systems to improve automatically through experience, making it a powerful tool for interpreting and analyzing vast amounts of data [9–12].

In contemporary applications, supervised models based on ML have become integral to a wide range of applications. They are prominently employed in natural language processing (NLP), where they help in understanding and interpreting human languages with applications ranging from translation services to sentiment analysis [13,14]. In the medical field, these techniques are revolutionizing medical image processing, enabling more precise diagnoses from imaging data and also enhancing medical signal processing for better monitoring and treatment of health conditions [15,16]. The widespread adoption of ML models in diverse fields underscores their versatility and transformative potential in technology and science [4,6,17,18].

In the field of automated brain tumor classification, two main ML methodologies have emerged as predominant methods: hand-crafted feature extraction methods and deep learning (DL) methods. Each technique offers distinct strategies and benefits for the analysis and classification of brain tumors from MRI images [19–22]. Hand-crafted feature extraction relies on predefined strategies to identify specific characteristics or 'features' in MRI images. These features, such as intensity, shape, and texture, are then utilized to classify the images into tumor or non-tumor categories. While these methods have been instrumental in early efforts at automated classification, they are limited by the need for expert knowledge in feature selection and can struggle with the high variability present in brain tumor images [4,23].

Leveraging neural networks, particularly CNNs, DL models automatically learn feature descriptions directly from the data, bypassing the need for manual feature extraction. These approaches have shown remarkable success in various image recognition tasks, including medical image analysis [24–26]. CNNs, with their ability to hierarchically extract and learn complicated hidden patterns in data, have become the cornerstone of modern automated image classification, including the differentiation of MRI images into tumor and non-tumor categories [25,27,28].

2. Literature Review

Vankdothu et al. [29] developed an innovative automated system for detecting and classifying medical images. This system encompasses several stages: the preprocessing of MRI images, segmentation, feature extraction, and classification. In the preprocessing stage, an adaptive filter is applied to reduce noise in the MRI images. For image segmentation, an enhanced version of the K-means clustering technique, known as IKMC, is utilized. The feature extraction procedure utilizes the gray-level co-occurrence matrix (GLCM) strategy to draw out critical and hidden patterns from the images. Subsequently, these features are input into a DL model for classification into various categories, such as non-tumors, meningiomas, gliomas, and pituitary tumors, utilizing recurrent CNNs (RCNN). This strategy demonstrated improved outcomes in classifying brain images from a specific dataset. The evaluation of this method was carried out using a dataset from Kaggle, comprising 394 test images and 2870 training images. The findings indicate that this method outperforms previous approaches in terms of performance. Additionally, the effectiveness of the RCNN model was benchmarked against contemporary classification techniques like U-Net, backpropagation, and RCNN, showcasing its superior capabilities in medical image classification.

Siddiqi et al. [30] suggested a cutting-edge feature extraction technique specifically designed for MRI images aimed at identifying and selecting prominent features associated with different brain diseases. This strategy stands out for its ability to discern key features from MRI scans, facilitating the differentiation between various disease classes. The approach employs a novel method that utilizes recursive values like the partial Z-value for class discrimination. The algorithm works by extracting a select set of features through

backward and forward recursion models. In the forward recursion model, the most interrelated features are identified based on the partial Z-test values. Conversely, the backward model aims to decrease the least interrelated features from the feature space. In both instances, the Z-test values are computed based on the predefined labels of the diseases, aiding in the precise identification of localized features, which is a major advantage of this method. Once the optimal features are extracted and selected, the model employs a Support Vector Machine (SVM) for training. This training allows the model to assign predictive labels to the MRI images accurately.

Ullah et al. [31] proposed a theory emphasizing the critical role of image quality, particularly when enhanced during the preprocessing phase, in improving the classification accuracy of statistical methods. To support this theory, they introduced an advanced image enhancement procedure comprising three distinct sub-stages. Initially, noise is mitigated using a median filter, followed by the enhancement of image contrast through histogram equalization. The process culminates with the conversion of images from grayscale to RGB. Once the images are enhanced, feature extraction is conducted on the improved MR brain images employing discrete wavelet transform. These extracted features are then further processed by implementing color moments, which include skewness, mean, and standard deviation, to effectively reduce the feature set. The final step involves training an advanced deep neural network (DNN) with these processed features. The purpose of this DNN is to correctly classify human brain MRI images, differentiating between normal and pathological conditions. This approach underscores the significance of image quality in preprocessing and its impact on the efficacy of DL models in medical imaging classification tasks.

Alrashedy et al. [32] introduced BrainGAN, a framework designed for both generating and classifying brain MRI images. This framework utilizes Generative Adversarial Network (GAN) architectures alongside DL models. A key aspect of this study was the development of an automated technique to ensure the quality of images generated by GANs. For this purpose, the framework used three different models: MobileNetV2, CNN, and ResNet152V2. These deep transfer models were trained employing images generated by two types of GANs: Deep Convolutional GAN (DCGAN) and Vanilla GAN. The effectiveness of these models was then assessed using a test set comprising actual brain MRI images. The experimental results indicated that the ResNet152V2 model exhibited superior performance compared to the CNN and MobileNetV2 models. This outcome underscores the potential of using advanced neural networks in conjunction with GAN-generated data for the accurate classification of medical images, especially in scenarios where real-world data are limited or difficult to access.

Many studies have mostly concentrated on binary classification problems, neglecting the intricacies of multiclass settings, such as differentiating between various tumor kinds. Our research has explicitly addressed these restrictions with a strong binary classification methodology, while proposing possibilities for expansion to multiclass scenarios. Furthermore, earlier models frequently encounter challenges associated with class imbalances and data scarcity, which impede performance and generalizability. In our study, we utilized advanced data augmentation techniques optimized using the Multi-verse Optimizer (MVO) to enhance training data diversity and reduce the risk of overfitting.

A significant drawback in previous research is the absence of effective optimization algorithms for deep learning systems, resulting in inferior performance and heightened computing expenses. We addressed this by incorporating the MVO to optimize data augmentation parameters and the configuration of trainable layers in VGG16 and ResNet50, thereby enhancing performance and generalization.

We employ a sophisticated strategy that combines the strengths of two renowned CNN architectures, VGG16 and ResNet50. These models have been selected for their proven efficacy in image classification tasks, providing a robust foundation for our classification strategy. The proposed approach begins with a critical preprocessing phase, pivotal for preparing MRI images for analysis by CNN models. This part incorporates several proce-

dures, including image resizing, conversion to grayscale, and the application of Gaussian blurring. These processes are instrumental in minimizing noise and standardizing the images, ensuring the focus is on features crucial for classification. Further, advanced image processing techniques are utilized to accurately isolate the brain region from each scan, a key step in precise tumor detection.

An innovative optimization technique is integrated into the model development, employing the MVO. A significant application of the MVO in this study is the optimization of data augmentation parameters. Data augmentation is an essential step in enhancing the robustness and generalization of DL models. By artificially enlarging the training dataset with modified versions of the original images, the risk of overfitting is diminished. This also prepares the models to manage diverse image variations. The MVO plays a critical role in identifying the most effective augmentation techniques and their parameters, optimizing the process for brain MRI images.

Furthermore, the MVO is instrumental in determining the optimal configuration of trainable layers in the VGG16 and ResNet50 architectures, a crucial factor in adapting the models for brain tumor classification. The selective unfreezing of layers for training achieves a balance between employing pre-learned features and adapting to the specific dataset. The MVO directs this process, identifying layers that when trained, significantly boost the model's efficacy.

The structure of this paper is accurately organized to offer a comprehensive understanding of the methodology and results. The introduction is followed by the materials and methods section, which details the preprocessing steps and the MVO optimization process. Following this, the data augmentation strategies and the process of building and training the VGG16 and ResNet50 models are thoroughly examined. The model evaluation section follows, discussing the metrics and techniques utilized to assess the performance of the models. A critical analysis of the findings is presented in the results and discussion section, where the effectiveness of this approach is compared with other methods, particularly emphasizing its role in the automated detection of brain tumors. The paper concludes with a summary of the key contributions and insights, and it outlines potential future research avenues in this crucial area of medical diagnostics.

3. Materials and Methods

This section methodically outlines the proposed method. Initially, the visual summary of the recommended model is illustrated in Figure 1.
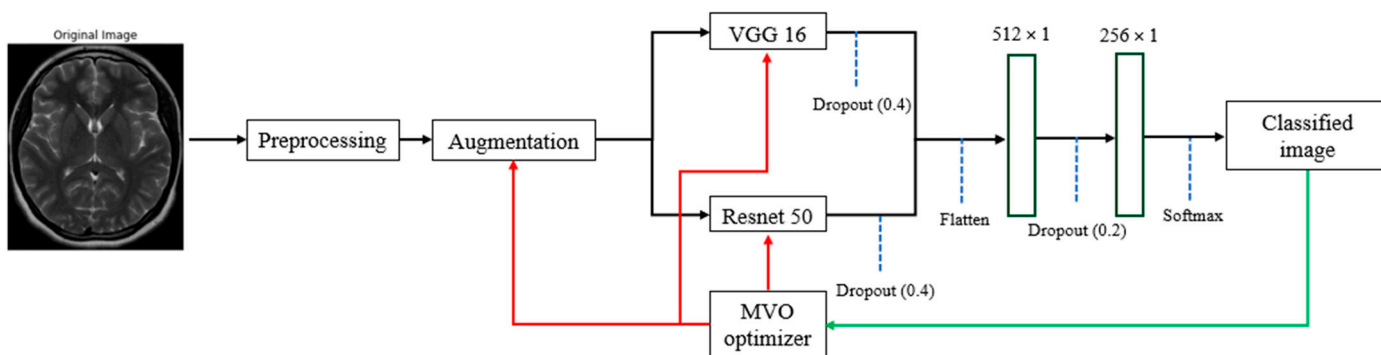

Figure 1. The graphical abstract of the suggested model.

3.1. Data Preprocessing

The preprocessing of MRI images is a fundamental step towards achieving accurate classification results when discerning between scans with and without tumors. Our procedure was designed to extract just the brain region from each MRI image, utilizing a combination of image processing techniques provided by the OpenCV library (version 4.8.1). These steps are described as follows:

- Image Resizing: Each image was resized to a standard dimension, certifying uniformity across the dataset. This was important for consistency when feeding the images into the neural network models.
- Grayscale Conversion: The resized images were converted to grayscale to moderate computational complexity. Color information was typically redundant for the task at hand, as the focus was on texture and shape within the images.
- Gaussian Blurring: To reduce high-frequency noise, a Gaussian blur was applied to the grayscale images. This smoothing technique aided in highlighting the more significant structures within the brain by softening edges and details.
- Otsu's Thresholding: We implemented Otsu's thresholding to separate the brain tissue from the background. This method automatically computed a threshold value for image binarization, which was employed to detect contours.
- Contour Detection and Selection: By applying contour detection, we identified the boundaries of all the objects in the binary image. We assumed the largest contour to be the brain's boundary, a reasonable assumption in a typical MRI scan.
- Extreme Points and Cropping: Once the largest contour was identified, we verified its extreme points. These points represented the furthest pixels in the horizontal and vertical directions within the contour, which we employed to define a cropping boundary.
- Image Cropping: The image was cropped using the extreme points as vertices, with an optional padding added to ensure no part of the brain was excluded.
- Displaying the Process: We visualized the preprocessing steps, displaying the original image, the contour of the largest detected region, the extreme points, and finally, the cropped brain region.

It is important to note that the performance of this preprocessing step depends on the characteristics and quality of the MRI scans provided. If the scans are not standardized or contain artifacts, additional preprocessing techniques might be required. Figure 2 demonstrates the results of applying the proposed preprocessing method. Our analysis indicates that although these procedures were computationally efficient, they did not substantially augment the overall runtime, hence maintaining the model's suitability for real-time or near-real-time applications.
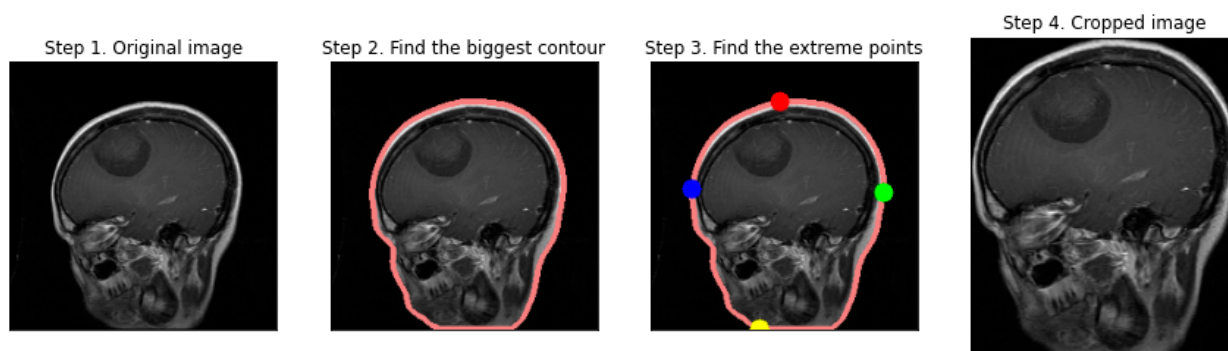

Figure 2. An example of applying the suggested preprocessing method. The original image is indicated to provide a baseline. The contour image illustrates the identified largest contour. The extreme point image marks the extreme points on the brain contour, demonstrating the cropping boundaries. The cropped image displays the isolated brain region, ready for input into the classification models.

3.2. Optimization Method

Optimization in ML is the manipulation of model parameters in order to minimize errors and enhance the precision of predictions. It is a vital process in model training that directly affects the performance of the model [7,33,34]. Metaheuristics are high-level problem-independent algorithmic frameworks that provide a set of guidelines or approaches to developing heuristic optimization algorithms. They are often nature inspired and designed to investigate the search space efficiently, which is crucial in avoiding local optima and finding a near-global optimum in complex problems. Unlike exact techniques, metaheuris-

tics do not guarantee an optimal solution, but they often discover good solutions with less computational effort, especially in large and complex search spaces [35–39].

In our study, we utilized the MVO technique, an effective metaheuristic algorithm inspired by the multi-verse theory in physics. The MVO algorithm models each solution as a universe and the features of the solutions as objects within these universes. The algorithm iterates through a process of exploration and exploitation, guided by the concepts of wormholes, black holes, and white holes, to share information among the universes and converge toward the best solution [40–42]. White holes act as channels for exporting matter from a universe, which in MVO terms means exporting solution features from a higher-ranked solution to others. Black holes serve the opposite function, absorbing matter, and thus importing solution features from other universes into the current solution. Wormholes provide a means for solutions to jump across the search space, encouraging exploration and aiding in avoiding local minima [40,43,44]. We used the MVO method for two main optimization tasks:

- Optimization of Data Augmentation Parameters: MVO was utilized to uncover the best data augmentation techniques and their respective parameter values. This step was crucial to enhance the dataset variability without deviating from the realistic transformations applicable to MRI images. By doing so, we aimed to maximize the performance of the classification models while avoiding overfitting.
- Layer-wise Trainability in CNNs: We also applied the MVO strategy to find the optimal configuration of trainable layers within VGG16 and ResNet50 architectures. The method helped in identifying which layers should be frozen (weights not updated during training) and which should be trainable (weights updated during training) to improve the models' accuracy and efficiency. This is particularly important, as deep neural networks can be computationally expensive to train, and freezing certain layers can significantly diminish the number of parameters that are essential to be updated, thus speeding up the training process and potentially improving generalization.

The MVO relies on various parameters that guide the optimization process. Adjusting these parameters can notably influence the algorithm's ability to find optimal solutions. Below is a detailed description of these parameters and the role they play in the MVO algorithm [42–44]:

- Universe Size (Population Size): This parameter defines the number of potential solutions (universes) that the algorithm will consider simultaneously. A larger universe size allows for greater exploration of the solution space. A universe size of 30 achieves a compromise between computing efficiency and diversity within the search space. More extensive populations enhance diversity but may impede the optimization process, whereas smaller populations might converge rapidly yet risk overlooking ideal solutions due to inadequate research. A universe size of 30 facilitates a sufficiently diversified population to investigate many solutions while minimizing computational expenses.
- Wormhole Existence Probability (WEP): This probability controls the creation of wormholes, which are mechanisms for sharing information between universes. A value of 0.6 signifies a considerable probability of the formation of these wormholes. This likelihood facilitates an effective combination of exploration (investigating new regions of the solution space) and exploitation (enhancing current solutions). A high WEP (approaching 1) may result in premature convergence due to excessive emphasis on exploitation, whereas a low WEP (approaching 0) restricts exploitation, hence obstructing convergence to an ideal solution. A WEP of 0.6 guarantees that the MVO algorithm can efficiently navigate the solution space while simultaneously enhancing viable solutions.
- Travelling Distance Rate (TDR): This rate determines how much a solution can be altered when a wormhole event occurs. A value of 0.4 signifies that alterations to the solutions are moderate. If the TDR is overly elevated, solutions may be excessively

disturbed, resulting in irregular exploration that could overlook excellent solutions. If TDR is excessively low, the algorithm may adopt an overly conservative approach and become trapped in local optima. A TDR of 0.4 offers a balanced methodology, enabling the algorithm to implement significant modifications to solutions while remaining close to attractive regions.

- Maximum/Minimum WEP: These parameters set the lower and upper bounds for the wormhole existence probability, introducing dynamic variability in the algorithm. The interval [0.2, 1.0] guarantees that during the initial phases of optimization, the method prioritizes exploration (with a reduced WEP at 0.2), thus avoiding premature convergence. As the algorithm advances, WEP approaches 1.0, highlighting exploitation in the latter phases to optimize the results. The incremental rise in WEP enables the early investigation of several options while refining the search in the concluding iterations for enhanced precision.

- Maximum/Minimum TDR: Similar to WEP, these parameters set the boundaries for the traveling distance rate, controlling the extent of solution alterations through the optimization process. Commencing with a minimum TDR of 0.4 guarantees that initial exploration remains closely aligned with the existing optimal solutions. As optimization progresses, the TDR may rise to 1.0, facilitating more substantial alterations in subsequent phases. This adaptive modification allows the method to maintain flexibility throughout the initial phases of optimization while intensifying solution refinement as it nears convergence.

These parameters are not static; they can be adapted during the optimization process to dynamically adjust the search behavior of the method. For our study, Table 1 outlines the specific parameters of the MVO algorithm that were employed to optimize data augmentation techniques and determine trainable layers in the VGG16 and ResNet50 models:

Table 1. Specific parameters of the MVO algorithm were employed for optimization purposes.

Hyperparameters	Description	Best Value
Universe Size	Number of solutions in the population	30
Wormhole Existence Probability (WEP)	Probability of wormholes' appearance	0.6
Travelling Distance Rate (TDR)	How far a wormhole can alter a solution	0.4
Max/Min WEP	Bounds the occurrence of wormholes	Max: 1.0, Min: 0.2
Max/Min TDR	Limits the modification extent by the wormholes	Max: 1.0, Min: 0.4

3.3. Data Augmentation

Data augmentation is an essential strategy in the field of DL, especially when dealing with image data. It refers to the process of generating new training samples from the original ones by applying a sequence of random changes that result in realistic variances [45–47]. This technique is crucial for several reasons [45,46,48]:

- Mitigates Overfitting: Augmentation spreads the variety of the training samples, which helps prevent the model from memorizing specific images and overfitting.
- Improves Generalization: By simulating various scenarios, data augmentation allows the model to generalize better to new, unseen data.
- Compensates for Imbalanced Datasets: In cases where some classes are underrepresented, augmentation can help to balance the samples without the need to collect more data.
- Enhances Model Robustness: Augmented data aid the model in learning more robust features that are invariant to certain transformations, which is significant for real-world applications.

In our study, the 'ImageDataGenerator' from 'Keras' is used to implement data augmentation, with a range of parameters set to apply random transformations to the images. Table 2 summarizes the data augmentation approaches employed in our study. Figure 3 demonstrates the results of applying the augmentation methods to an image.

Table 2. A description of all data augmentation methods employed in our study. * Note: The 'Channel Shift' and 'Fill Mode' methods are not selected by the optimization approach for data augmentation.

| Index	Augmentation Method	Description	Used in Study	Suggested Values by MVO Optimizer
1	Rotation range	Random rotation within a specified range of degrees.	Yes	10
2	Width Shifting range	Random horizontal shift.	Yes	0.1
3	Height Shifting range	Random vertical shift.	Yes	0.1
4	Shear range	Random shearing.	Yes	0.1
5	Brightness range	Random brightness adjustments.	Yes	[0.5...1.5]
6	Horizontal Flip	Random horizontal flipping.	Yes	True
7	Vertical Flip	Random vertical flipping.	Yes	True
8	Zoom range	Random zooming into the images.	Yes	0.1
9	Channel Shift *	Random channel shifting.	No	-
10	Fill Mode *	Method for filling points outside the boundaries of the input.	No	-
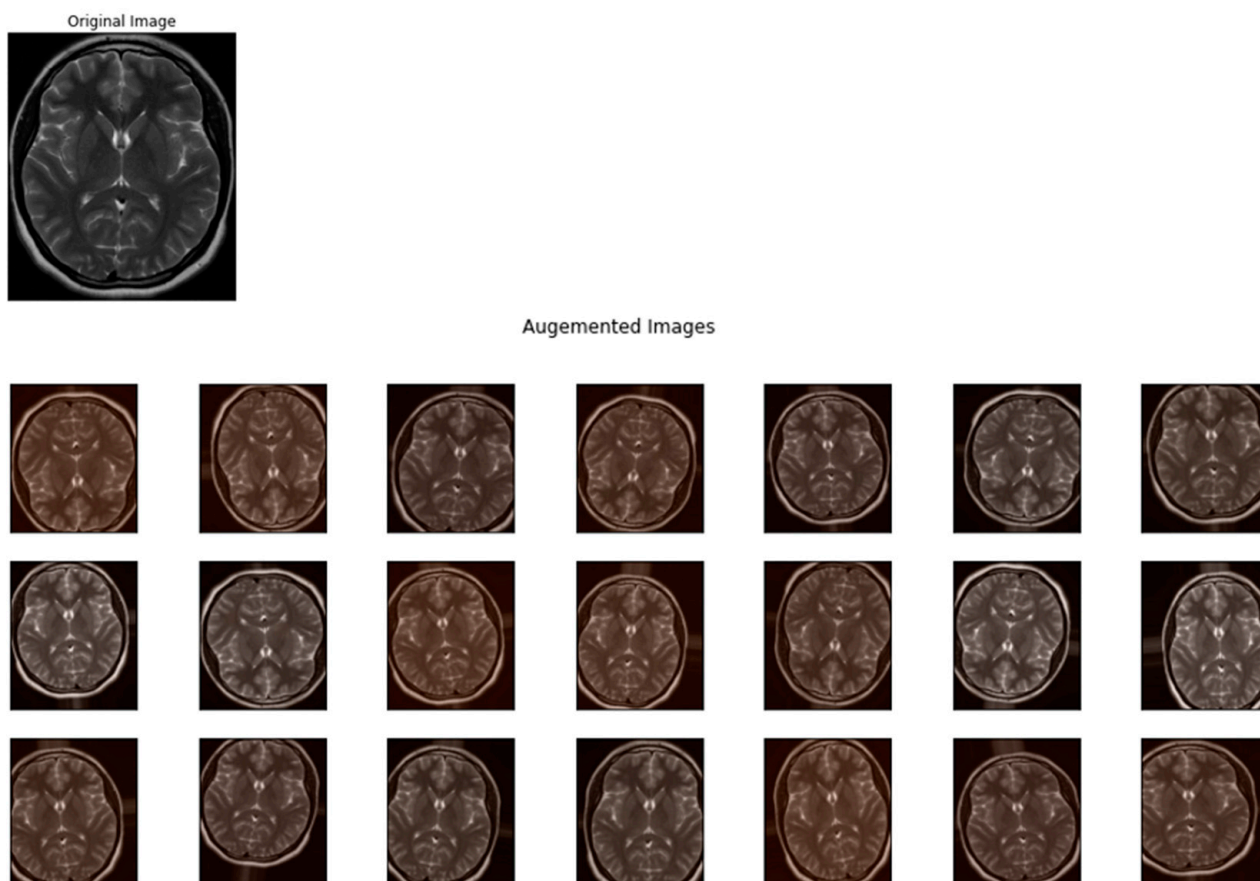
Figure 3. The results of applying the augmentation methods to an image.

3.4. Model Building with VGG16 and ResNet50

In our study, we focused on building and fine-tuning two prominent CNN architectures: ResNet50 and VGG16. These models were chosen for their proven track records in image classification tasks. VGG16 stands out for its depth and utilization of uniform 3×3 convolutional layers. Its architecture is a demonstration of the idea that depth is crucial for attaining high levels of accuracy in complex image classification tasks. A pre-trained VGG16 model, originally trained on the ImageNet dataset, provides a strong feature extraction base due to its exposure to a widespread variety of images. However, this model is often criticized for its high computational cost and extensive memory requirement, mainly due to its deep architecture and fully connected layers [49–53].

The adaptation of VGG16 for binary classification involves fine-tuning the dense layers to suit the specific task. The main advantage here is the ability to capture fine-grained details, which can be necessary for discovering subtle features in MRI images that are indicative of tumors. ResNet50, with its residual blocks and skip connections, addresses the challenge of training deep networks without falling prey to the vanishing gradient problem. The skip connections facilitate the training of the model by letting gradients flow through the architecture more effectively. A pre-trained ResNet50 model brings robustness and a quicker training convergence to the table, owing to these residual blocks. The architecture is more efficient than VGG16, both in terms of computational resources and required training time, making it a strong candidate for DL tasks. However, one could argue that ResNet50 may sometimes lead to feature redundancy due to its very deep architecture [48,54–57].

In the binary classification of MRI images, ResNet50's architecture is beneficial for its ability to learn from residuals, potentially refining the model's ability to distinguish relevant patterns associated with tumors. Combining ResNet50 and VGG16 can be particularly advantageous for the task at hand. While VGG16 is adept at capturing texture and detailed features, ResNet50 excels in leveraging deeper contextual information and solving the degradation problem in deep networks. The fusion of these models could lead to a comprehensive feature extraction mechanism, where VGG16 contributes detailed local features and ResNet50 contributes broader contextual understanding.

This combined methodology may offer a more robust representation of the MRI images, capturing both the minute details essential for distinguishing small or early-stage tumors and the broader patterns necessary for identifying more significant anomalies. The ensemble of these two architectures has the potential to capitalize on the strengths of both while mitigating their individual weaknesses, leading to an improved classification performance.

3.5. Model Training

The model was created with Python, employing tools such as TensorFlow (version 2.10.0), Keras (version 2.10.0), and OpenCV (version 4.8.1) for data processing and model training. The experiments were performed on an NVIDIA GeForce RTX 3080 GPU, facilitating effective parallel computation, complemented by a high-performance 12th Gen Intel(R) Core(TM) i7-12700H and 32 GB of RAM. The model training phase was characterized by a dynamic and adaptive method. Each training iteration was contingent upon the new sets of hyperparameters recommended by the optimization process for data augmentation and layer unfreezing. This strategy ensured that the training process was continually refined and tailored based on the model's performance feedback.

The iterative nature of the training was closely tied to the MVO, which provided updated values for data augmentation parameters and discovered the layers to be unfrozen for each training cycle. This method permitted the model to evolve in conjunction with the optimizer's findings, effectively creating a feedback loop between training performance and parameter adjustment. The optimizer's role was essential in defining the non-trainable and trainable parameters of the model. With a total of 102,659,777 parameters, the optimizer fine-tuned the model by selectively unfreezing layers. This strategic balance played a substantial role in the model's ability to adapt to the specific features of MRI images. Data augmentation was not a static preprocessing phase but a variable aspect of training. As the

optimizer advocated new augmentation values, the training samples were transformed accordingly, which permitted the model to learn from a more diverse set of features. This continual alteration of the training samples helped mitigate overfitting and improved the model's generalizability.

During each training iteration, the structure was assessed using a validation set to monitor its performance against data that were not part of the training process. This regular validation served as a checkpoint to verify the efficacy of the current set of hyperparameters and informed subsequent adjustments from the optimizer. Throughout the training phase, the model underwent evolution, not only in its biases and weights but also in its structure as layers were unfrozen and data augmentation techniques were varied. This evolution aimed to refine the model's capability to identify and classify the presence of tumors in MRI scans with increasing precision. The outcome of this iterative and adaptive training process was a highly tuned model, uniquely customized to the task at hand. The model's training was directly influenced by the optimizer, confirming that with each training cycle, the model was progressively equipped to handle the complexities of tumor detection in MRI images.

3.6. Dataset

The employed dataset consisted of 3264 brain MRI scans [58]. These scans were derived from individuals seeking medical assessment and evaluation of their brain health. Each MRI scan held the potential to reveal critical information about the presence or absence of brain tumors. MRI, being non-invasive, ensures minimal discomfort and risk to the patients while delivering invaluable insights into the brain's intricate anatomy. We utilized 10% of the samples for validation, 10% of the samples for testing, and the rest for training purposes. Some images from the dataset are indicated in Figure 4. We utilized 80% of the original dataset (training samples) for data augmentation. The utilization of eight distinct augmentation techniques substantially enhanced the training dataset, hence augmenting model resilience and generalization.

The dataset is carefully labeled, attributing two distinct classes to each MRI scan [58]:

1. No Tumor:
    *   This class encompasses brain MRI scans attained from individuals who do not demonstrate any detectable brain tumors. These scans serve as a crucial baseline for comparative analysis against scans that do depict tumors.
    *   Individuals within this class are typically those seeking routine brain examinations or experiencing neurological symptoms unrelated to tumor presence.

2. Tumor:
    *   The "Tumor" class comprises brain MRI scans from individuals who have been clinically diagnosed with brain tumors. Brain tumors are marked in various locations, sizes, and types, making their accurate finding and grouping a formidable challenge in the field of medical imaging.
    *   The tumors within this class may include metastatic tumors, gliomas, meningiomas, and other forms of abnormal growth within the brain.

3.7. Model Evaluation

The model assessment part was tailored to analyze the performance of the trained models, which were configured to classify MRI images. The models were compiled with 'binary crossentropy loss' to suit the binary classification task. The Adam optimizer with a learning rate of 0.001 was chosen for its effectiveness in handling the noise associated with the inherently stochastic nature of training deep neural networks. The assessment metrics selected were precision, accuracy, and recall, providing a rounded perspective on the models' classification abilities, as described in Equations (1)–(3) [28,35,59,60]:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{1}$$

$$\text{Accuracy} = \frac{TP + TN}{TN + FN + TP + FP} \tag{2}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3}$$

These metrics are especially crucial in the context of binary classification, such as identifying the presence or absence of tumors in MRI images.
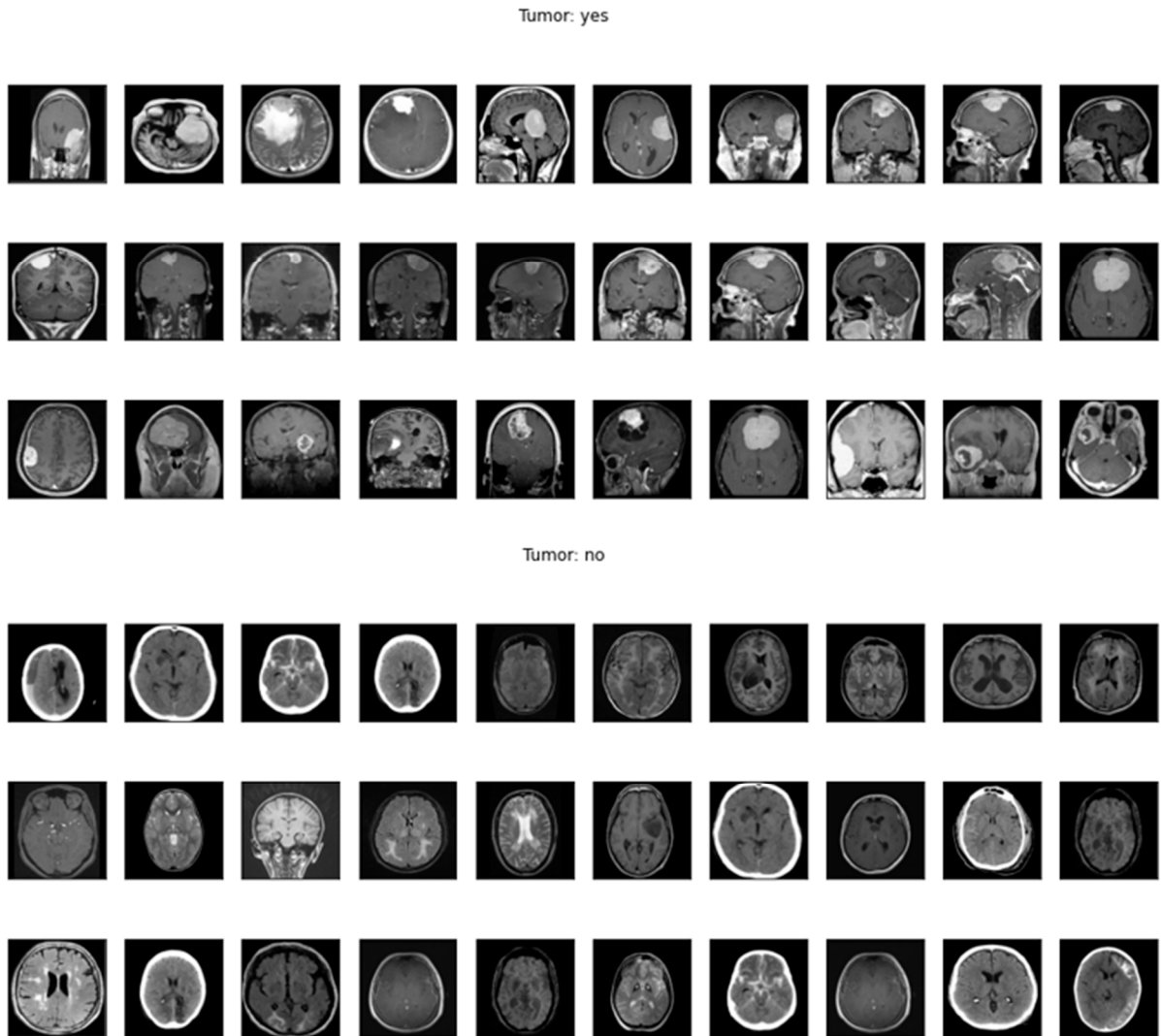

Figure 4. Some images from the dataset samples.

1. Precision (Equation (1)):
   - Description: Precision assesses the accuracy of the positive predictions made by the structure. In simpler terms, it answers the following question: of all the instances where the model predicted 'positive', how many were actually positive?
   - Components:
     - True Positives (TP): These are the instances where the model precisely predicts the positive class.
     - False Positives (FP): These are the instances where the model wrongly predicts the positive class.

- Usage: Precision is particularly essential in situations where the cost of a false positive is high. For example, in medical diagnostics, falsely recognizing a healthy patient as having a tumor can lead to unnecessary stress and treatment.

2. Accuracy (Equation (2)):
   - Description: Accuracy measures the proportion of true outcomes (both true negatives and positives) among the total number of cases examined. It essentially quantifies how often the model is correct.
   - Components:
     - True Negatives (TN): These are the instances where the model properly predicts the negative class.
     - False Negatives (FN): These are the instances where the model mistakenly predicts the negative class.
   - Usage: Accuracy is a valuable measure when the classes in the dataset are balanced. However, its usefulness is reduced when dealing with imbalanced datasets, as it can be misleadingly high in cases where the model mostly predicts the majority class correctly.

3. Recall (Sensitivity) (Equation (3)):
   - Description: Recall measures the proportion of actual positives that were accurately recognized by the model. It answers the following question: of all the instances that were actually positive, how many did the model identify?
   - Usage: Recall is critical in contexts where missing a positive instance is significantly worse than falsely detecting a negative instance as positive. For example, in medical diagnostics, a false negative (failing to detect a tumor) can be more dangerous than a false positive.

A 50-epoch training process was implemented with an early stopping set to monitor the validation loss and patience of seven epochs. This strategy aimed to avoid overfitting by halting the training if the validation loss did not improve for seven consecutive epochs. During training, the model learned from the training generator with a specified number of steps per epoch, while the validation generator provided data for evaluating performance after each epoch. Post-training, the history object was queried to extract the training and validation metrics across epochs. These metrics included accuracy and loss for both training and validation sets, providing insight into the learning curve and model convergence. These plots are critical for identifying trends such as overfitting or underfitting and for confirming the efficiency of early stopping.

The model's predictive power was further scrutinized by making predictions on the validation and test sets and constructing confusion matrices for each. The confusion matrices provided detailed insight into the true negative, true positive, false positive, and false negative rates, offering a granular view of the model's classification capabilities. Cases of misclassification were identified, and the corresponding MRI images were displayed, offering an opportunity for qualitative error analysis. This analysis could reveal characteristics or patterns in the images that were challenging for the model, informing potential improvements.

Finally, the receiver operating characteristic (ROC) curve was plotted, and the area under the curve (AUC) was determined for the test set. The ROC curve is a powerful tool for evaluating the true positive rate against the false positive rate at various threshold settings, while the AUC provides a single metric summarizing the overall performance of the model in discriminating between classes. The comprehensive evaluation provided by these models allows for a detailed understanding of the models' performance and is instrumental in guiding future iterations of the model training and optimization process.

4. Results and Discussion

In a comprehensive evaluation of DL models for classifying MRI images, the study investigated four distinct configurations: ResNet50, VGG16, a combination of VGG16

and ResNet50, and a further combination augmented with the MVO. The models were evaluated on both validation and test datasets to gauge their recall, precision, and accuracy, providing insights into their respective performances and suitability for tumor detection tasks. The outcomes of applying four models to the dataset are demonstrated in Table 3 and Figures 5–12.

Based on accuracy ratings of 89% and 88%, respectively, the VGG16 model performs rather well in both the validation and test stages. Although it has good generalization, its precision (91% validation, 92% test) is marginally lower than that of other models, suggesting a moderate amount of false positives. It is efficient at identifying true positives, though, demonstrated by its high recall, which reached 96% in validation and 94% in testing. This feature of VGG16 is useful when obtaining a high number of positive cases—even at the expense of some false positives—is necessary.

Conversely, ResNet50 achieves 92% in both the test and validation phases, yielding higher accuracy. With 96% in testing and 95% in validation, there has been a notable increase in precision. ResNet50 is a good option in situations where the cost of false positives is high, as this signals that it is more effective in minimizing false positives. Although the recall for ResNet50 remains consistent at 96% throughout validation, it slightly decreases to 94% during testing, indicating a little decline in its capacity to generalize. Overall, ResNet50 presents itself as a trustworthy model for tasks requiring accuracy by striking a good mix between precision and recall.
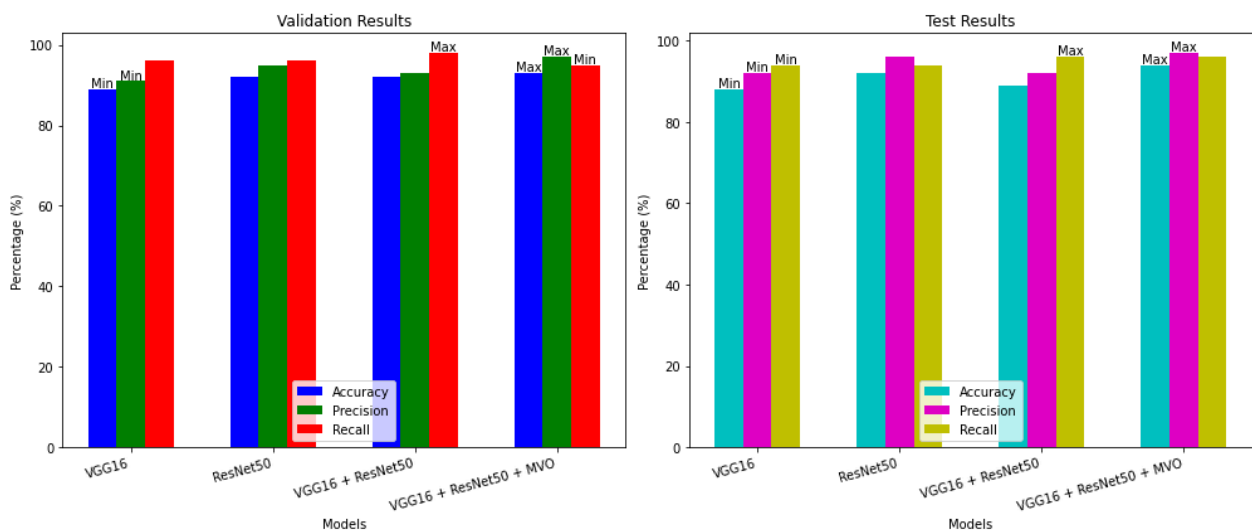

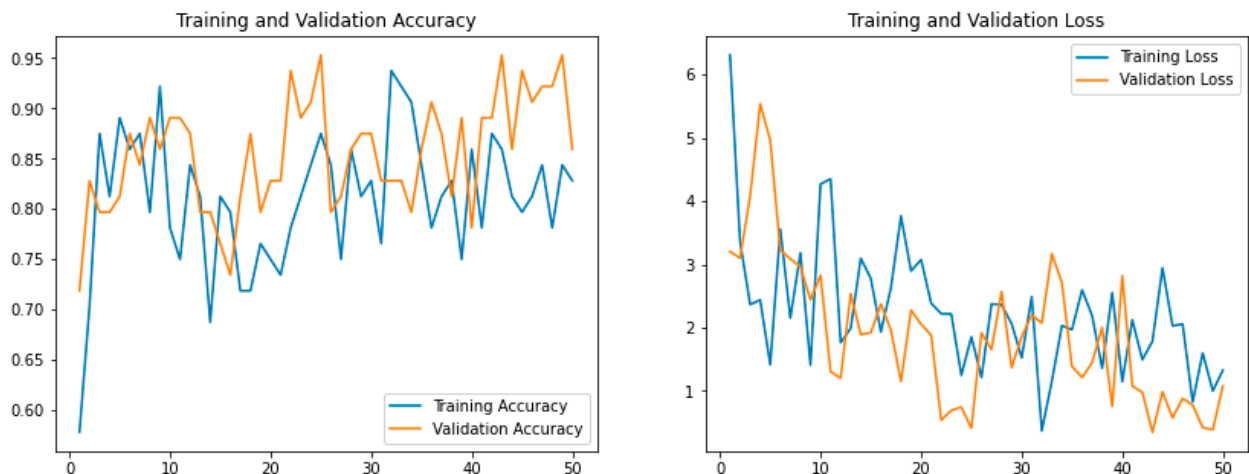**Figure 5.** Comparative performance analysis of DL models on MRI image classification.
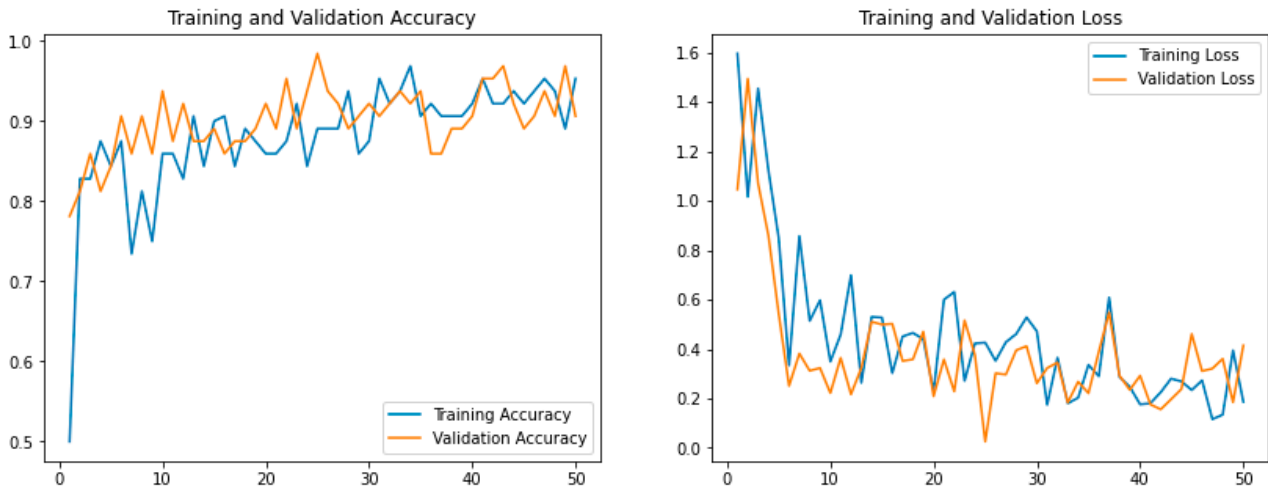

Figure 6. The performance of the VGG16 model.

Figure 7. The performance of the ResNet50 model.

Table 3. Quantitative outcomes of applying four models to the dataset.

| Data	Model	Accuracy (%)	Precision (%)	Recall (%)	Run Results
Validation	VGG16	89	91	96	12/12 [==============================] - 2s 172ms/step Validation Accuracy = 0.89 Validation Precision = 0.91 Validation Recall = 0.96 12/12 [==============================] - 1s 74ms/step Test Accuracy = 0.88 Test Precision = 0.92 Test Recall = 0.94 359/359 [==============================] - 2s 5ms/step
Test		88	92	94	
Validation	ResNet50	92	95	96	12/12 [==============================] - 1s 49ms/step Validation Accuracy = 0.92 Validation Precision = 0.95 Validation Recall = 0.96 12/12 [==============================] - 1s 49ms/step Test Accuracy = 0.92 Test Precision = 0.96 Test Recall = 0.94 359/359 [==============================] - 4s 11ms/step
Test		92	96	94	
Validation	VGG16 + ResNet50	92	93	98	12/12 [==============================] - 2s 118ms/step Validation Accuracy = 0.92 Validation Precision = 0.93 Validation Recall = 0.98 12/12 [==============================] - 1s 118ms/step Test Accuracy = 0.89 Test Precision = 0.92 Test Recall = 0.96 359/359 [==============================] - 5s 13ms/step
Test		89	92	96	
Validation	VGG16 + ResNet50 + MVO	93	97	95	12/12 [==============================] - 2s 130ms/step Validation Accuracy = 0.93 Validation Precision = 0.97 Validation Recall = 0.95 12/12 [==============================] - 1s 129ms/step Test Accuracy = 0.94 Test Precision = 0.97 Test Recall = 0.96 359/359 [==============================] - 5s 15ms/step
Test		94	97	96	
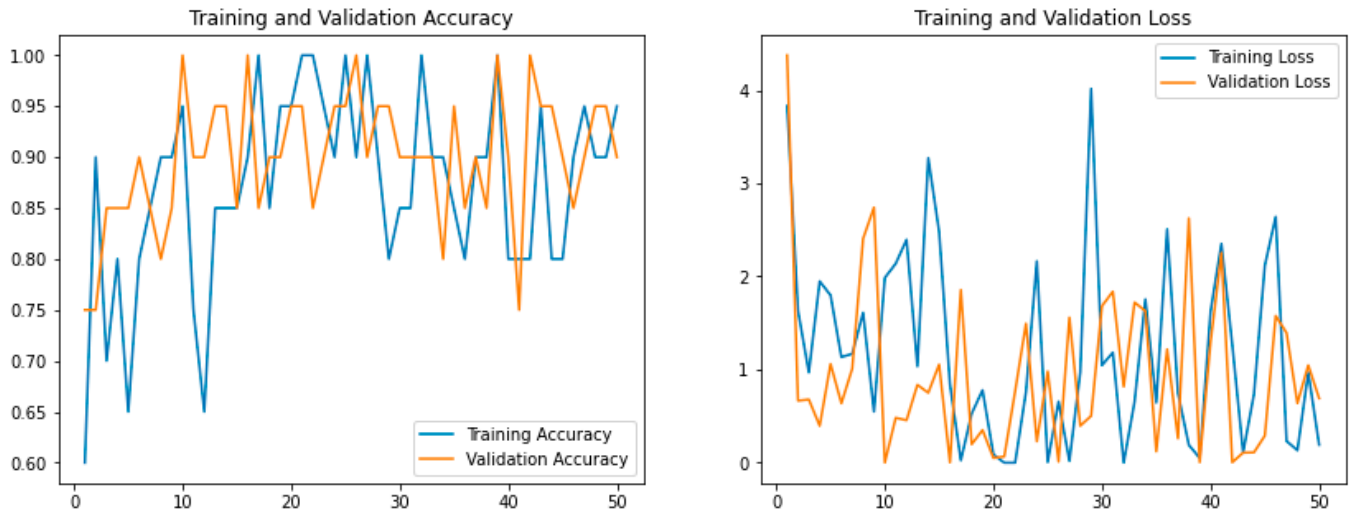
Figure 8. The performance of our model without applying the MVO. All layers in the ResNet50 and VGG16 were frozen.

There are some interesting performance differences when VGG16 and ResNet50 are combined. The test accuracy falls to 89%, indicating possible difficulties with generalization, while validation accuracy remains at 92%, mirroring ResNet50's performance. However, recall is where this hybrid model succeeds, reaching 98% in the validation phase. With medical applications like tumor identification, where missing a positive case could have major repercussions, this increase in recall suggests that the combination model is quite good at reducing false negatives.
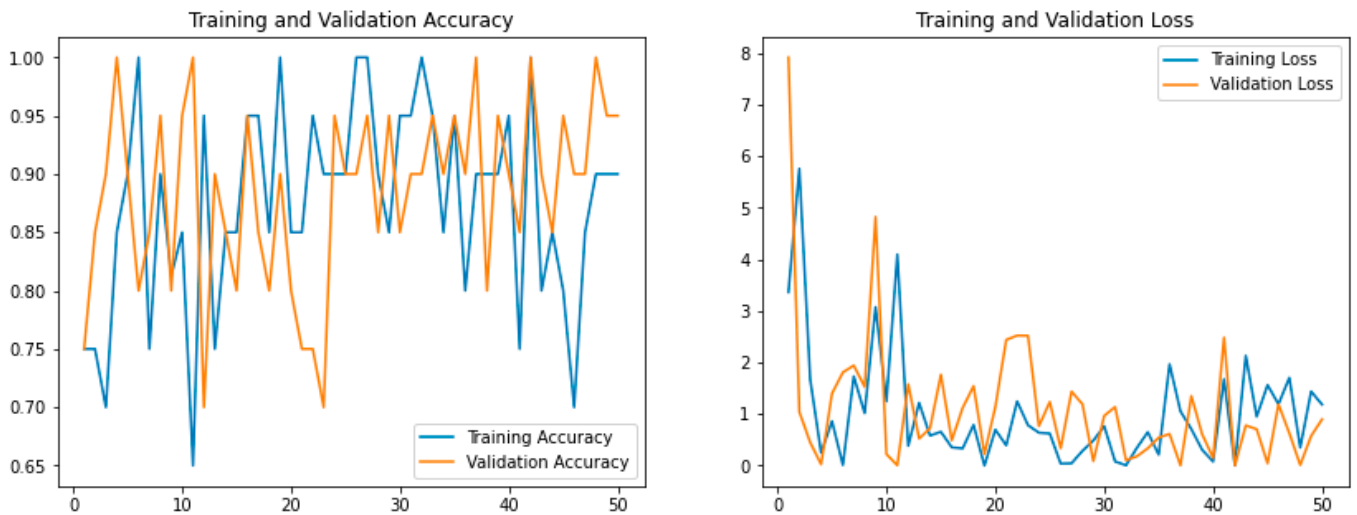

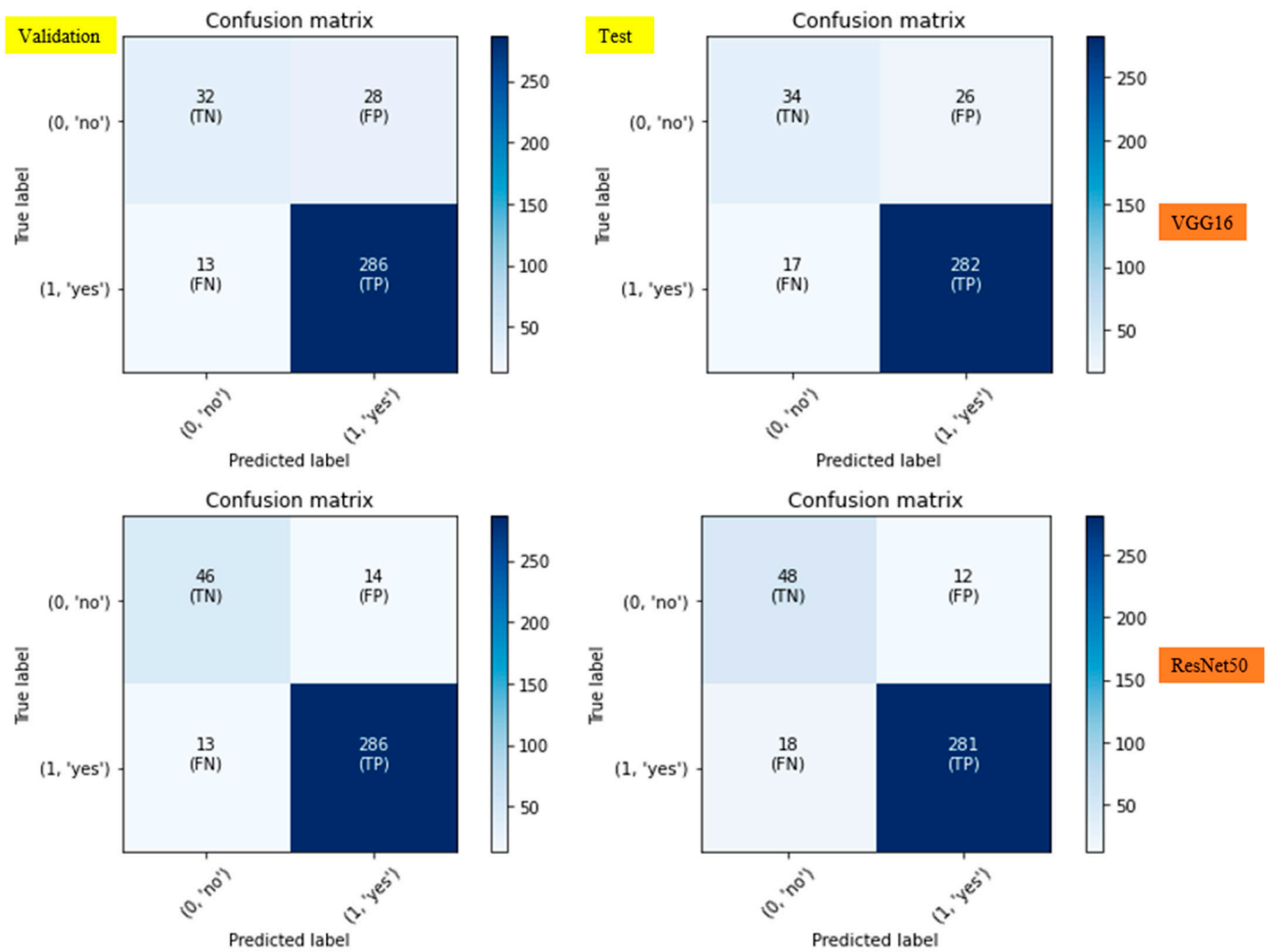**Figure 9.** The performance of our model using the MVO.

Figure 10. The confusion matrix of the ResNet50 and VGG16 models.

The precision of the VGG16 + ResNet50 combination drops to 93% in validation despite this recall improvement, suggesting a minor rise in erroneous positives when compared to ResNet50 alone. In certain situations—like in medical diagnostics—where lowering false positives is more crucial than minimizing false negatives, this trade-off might be justified. Despite the possibility of more false positives, the strong recall of this combination model makes it especially helpful in situations when gathering all pertinent cases is essential. Adding the MVO to the VGG16 and ResNet50 combination model yields the best overall performance. Improvements are made to test accuracy to 94% and validation accuracy to 93%. The model's precision is further improved, attaining 97% in the validation and test stages, indicating its excellent efficacy in reducing false positives. For applications where minimizing misclassifications is crucial, this makes the MVO-optimized model incredibly dependable.
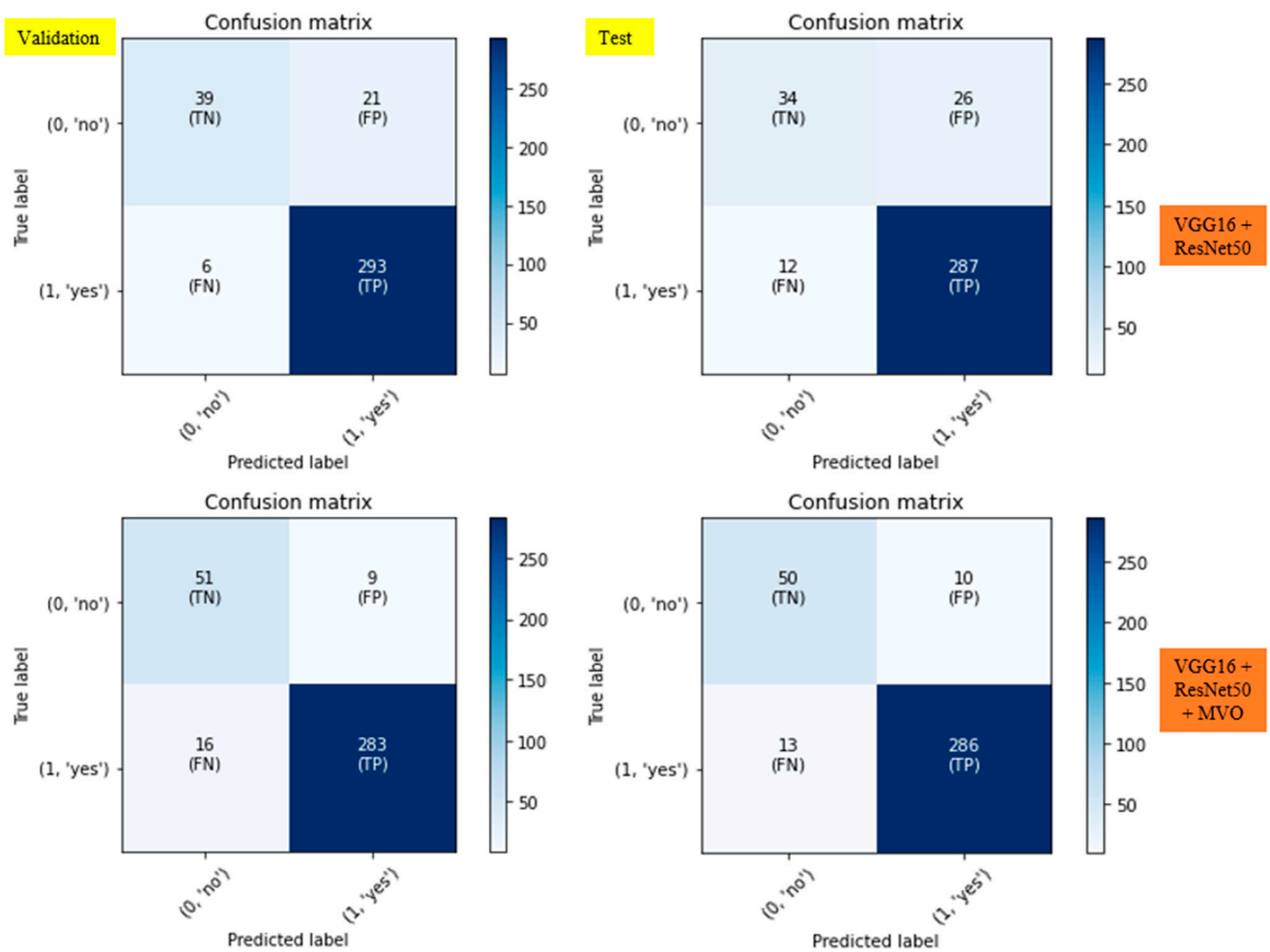
Figure 11. The confusion matrix of our model.

The MVO-optimized model continues to show strong recall values, 95% in validation and 96% in testing, demonstrating its persistent ability to detect true positives. For tasks requiring low error rates and high accuracy, the VGG16 + ResNet50 + MVO model is the most efficient option due to its precision and recall balance and high accuracy. This model offers reliability across many datasets and has good generalization, as evidenced by the consistency between test and validation measures. The trade-offs between precision and recall differ throughout the models. VGG16 performs well in recall at the expense of precision, which makes it appropriate in situations where finding every positive case—even if there are some false positives—is crucial. ResNet50 offers a more balanced method with excellent precision and accuracy, which makes it perfect for jobs where reducing false positives is crucial. The best recall is provided by the combined VGG16 + ResNet50 model, particularly in validation; however, its precision is marginally worse, suggesting that it might generate more false positives.
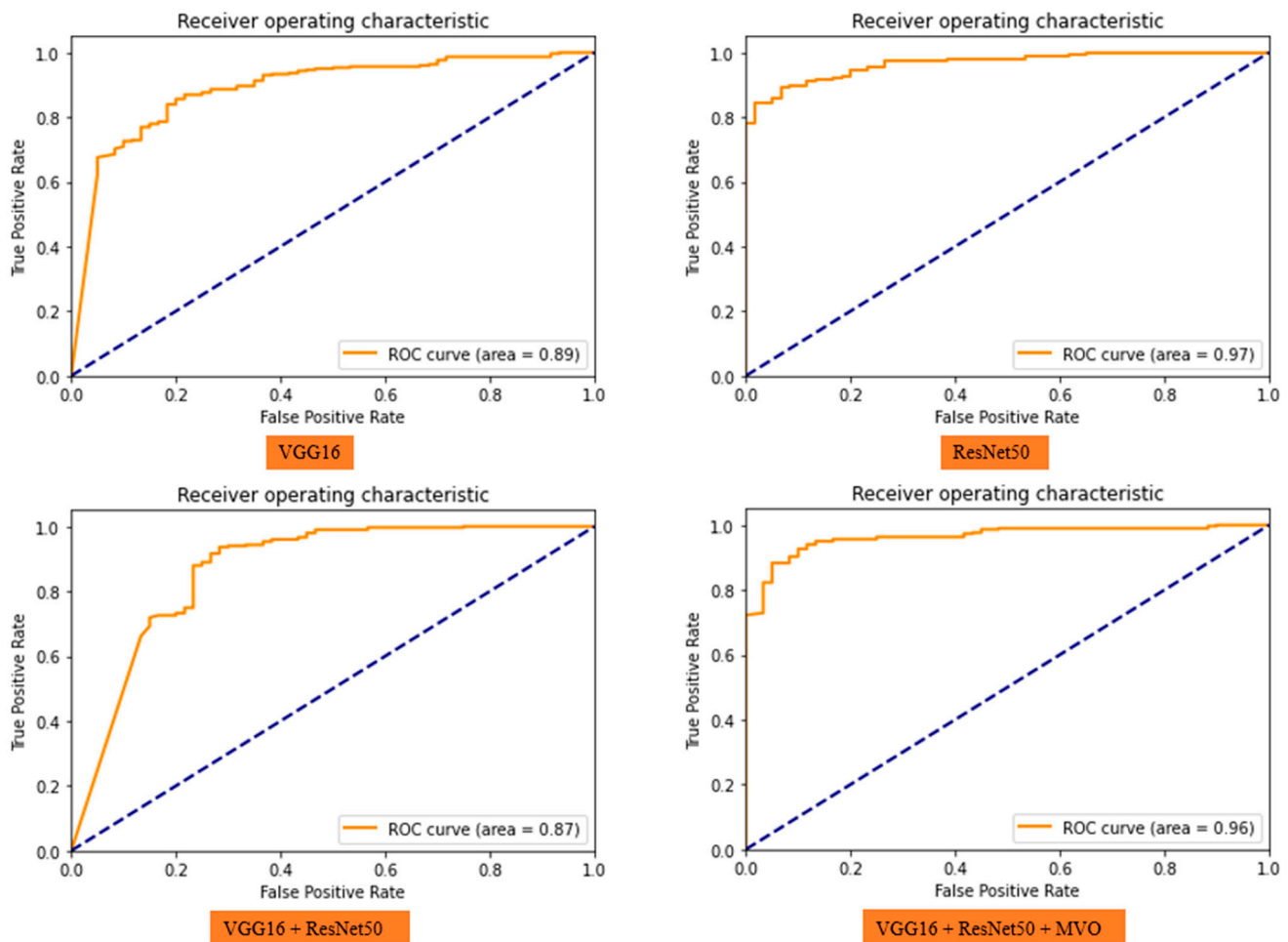
Figure 12. The ROC curves of different models for evaluating the true positive rate against the false positive rate at various threshold settings.

The combined model's MVO-optimized version provides the optimal balance in terms of precision and recall. Because it reduces false positives as well as false negatives, it is perfect for scenarios in which both kinds of errors might have serious repercussions. Since both overclassification and under-classification can result in unfavorable consequences, striking this balance is especially crucial when classifying medical images. When assessing the performance of a model, generalization is crucial. VGG16 exhibits a minor decline in test accuracy (88% vs. 89%) compared to validation, suggesting a minor problem with generalization. On the other hand, ResNet50 shows good generalization abilities, as it retains the same accuracy in both stages (92%). For deployment scenarios where the model must exhibit consistent performance across multiple datasets, ResNet50 is a dependable choice because of this.

Test accuracy is a little off for the combined VGG16 + ResNet50 model; it dropped from 92% in validation to 89%. This shows that although the model performs very well in recall, it may have overfitted the validation set. On the other hand, the model that has been optimized for maximum variance (MVO) exhibits better generalization. Its test accuracy of 94% is almost identical to its validation accuracy of 93%. This suggests that the MVO works very well to improve the model's generalization skills, enabling it to function well without overfitting on a variety of datasets. Every model has different advantages and disadvantages. Even though VGG16 occasionally produces false positives, it is the best option for high-recall jobs when finding every potential positive case is crucial. ResNet50 performs more evenly and excels in both precision and accuracy, which makes it appropriate for applications where minimizing false positives is essential. Recall is

increased by combining VGG16 and ResNet50, but test accuracy is somewhat decreased, suggesting possible overfitting.

Despite the fact that the MVO-optimized model introduces a slightly higher computational cost as a result of the additional optimization layer and hyper-parameter fine-tuning, this increase is justified by the model's improved performance and generalization. Table 3 illustrates that the MVO-optimized model attains enhanced accuracy (94% on the test set) and precision (97% for both validation and test sets) in comparison to ResNet50, which exhibits 92% accuracy and 96% precision. These enhancements are essential in medical diagnostics, where increased accuracy and precision can profoundly influence outcomes. Consequently, the trade-offs necessitate a balance between processing resources and the demand for enhanced classification performance, rendering our approach especially advantageous in contexts where accuracy is critical.

With good accuracy, precision, and recall across validation and test sets, the VGG16 + ResNet50 model performs best overall when the MVO is added. Because it offers a strong balance between precision and recall while retaining high generalization capabilities, the VGG16 + ResNet50 + MVO combination is the most efficient and dependable method for classifying brain MRI images. The MVO optimizes hyperparameters, including learning rates, dropout rates, and the choice of trainable layers. The MVO improves model performance and efficiency by the iterative refinement of these parameters.

As described in Figures 6–9, the ResNet50 model demonstrates considerable enhancement in accuracy and a decrease in loss across the 50 epochs. Initially, Epoch 1 commences with an accuracy of 50% and a significant loss of 1.59. By Epoch 6, accuracy rises to 87.50%, and loss diminishes to 0.33, showing the model's swift adaptation in the early training phase. The model has a stable accuracy trend during the intermediate epochs, averaging approximately 85–90%, while the loss persistently decreases, attaining a minimum of 0.22 by Epoch 20. Notwithstanding this robust performance, ResNet50 exhibits variations in loss, especially during the later epochs. Epoch 36 exhibits an elevated loss of 0.60, although accuracy stays somewhat consistent. The discrepancy in loss may indicate that although the model is learning efficiently, it encounters difficulties in consistently minimizing loss, suggesting the potential for further optimization or fine-tuning.

The VGG16 model has a more stable progression for accuracy and loss. Commencing at Epoch 1, VGG16 attains an accuracy of 57.81%, accompanied by a substantial loss of 6.3101, indicating initial challenges in categorization. Nonetheless, the model rapidly enhances, attaining 81.25% accuracy by Epoch 5, while its loss markedly declines to 1.4149, indicating more effective learning. At Epoch 10, VGG16 attains an accuracy of 89.06% and a loss of 1.3050, further illustrating the model's consistent convergence. While VGG16's loss diminishes throughout training, it does not attain loss levels as low as those of ResNet50. By Epoch 20, VGG16's loss is 2.05, while ResNet50's loss at the same epoch is 0.22. This indicates that although VGG16 continuously enhances accuracy, it faces greater challenges in minimizing its loss as effectively as ResNet50.

The combined model of ResNet50 and VGG16 yields superior overall accuracy relative to the standalone models. By Epoch 2, the hybrid model attains an accuracy of 90% and a comparatively low loss of 1.63, markedly surpassing the performance of both independent models at the moment. Throughout the subsequent epochs, the hybrid model sustains consistent accuracy, achieving 95% by Epoch 10 with a diminished loss of 0.0011, signifying robust convergence and negligible classification error. In subsequent epochs, the model maintains accuracy but exhibits minor fluctuations in loss; for instance, in Epoch 15, accuracy remains at 85%, while loss rises to 1.76. The hybrid technique exhibits superior overall performance in accuracy and loss reduction compared to the individual use of ResNet50 or VGG16.

The incorporation of the MVO into the hybrid architecture of ResNet50 and VGG16 significantly improves accuracy and loss performance. Beginning with Epoch 4, the model attains flawless accuracy (100%) with a remarkably low loss of 0.0248, representing a substantial enhancement compared to the previous models. The MVO-optimized model sustains near-perfect accuracy across the subsequent epochs, routinely achieving 95–100% accuracy in validation datasets. The MVO demonstrates significantly superior convergence regarding loss compared to the unoptimized hybrid model. Epoch 10 attains 95% accuracy with a loss of merely 0.22, indicating robust stability. Despite certain swings in loss during later epochs, notably Epoch 20, where the loss rises to 1.13, the overarching trend indicates that the MVO continually reduces loss more efficiently than the preceding models.

Figure 12 indicates that ResNet50 demonstrates marginally superior performance in several criteria, notably in the ROC curve. The MVO-optimized VGG16 + ResNet50 model offers a more balanced methodology by enhancing overall precision and recall, which is essential for applications that require the reduction of both false positives and false negatives. This equilibrium highlights the robustness of the suggested model for dependable tumor classification despite ResNet50's superior performance in particular domains. In the comparative analysis of the suggested structure integrated with several optimization algorithms (the Shuffled Frog Leaping Algorithm (SFLA), Multi-verse Optimizer (MVO), Red Deer Algorithm (RDA), and Whale Optimization Algorithm (WOA)) distinct patterns emerge regarding their influence on the model's precision, recall, and accuracy. The obtained results by applying different optimization approaches are described in Table 4 and Figure 13.

This approach, which begins with the SFLA, achieves a strong balance across all metrics, with test accuracy and validation accuracy at 91% and 91%, respectively. Recall and precision are both quite high, with recall reaching 96%. This makes it a good option for applications where reducing false negatives is essential. Although there are some precision trade-offs compared to some other approaches, the robustness of this approach is demonstrated by the consistency between validation and test measurements. However, an alternative strength is shown by the WOA optimization. The WOA is the least accurate approach overall, with test accuracy at 88% and validation accuracy at 85%, but it excels in precision. With a precision of 97% in validation and 98% in testing, it is the most effective way to lower false positives. The reduced recall scores (85% validation, 87% test), however, imply that the WOA forgoes its capacity to accurately identify positive cases in favor of conservatism, which may result in an increased number of false negatives in the classification task.
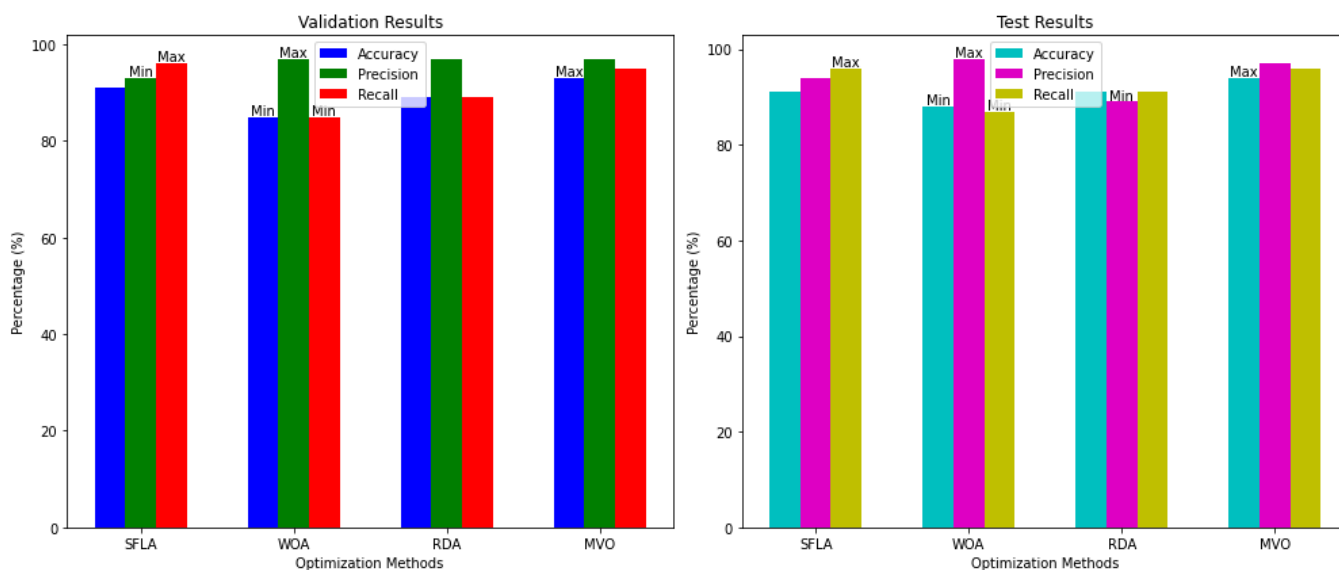

Figure 13. A performance comparison of the suggested model with different optimization techniques.

Table 4. The results by applying different optimization approaches.

Data	Model	Accuracy (%)	Precision (%)	Recall (%)	
Validation	Our model + SFLA [61]	91	93	96	12/12 [===========================] - 7s 231ms/step Validation Accuracy = 0.91 Validation Precision = 0.93 Validation Recall = 0.96 12/12 [===========================] - 1s 131ms/step Test Accuracy = 0.91 Test Precision = 0.94 Test Recall = 0.96
Test		91	94	96	359/359 [===========================] - 5s 14ms/step
Validation	Our model + WOA [62]	85	97	85	12/12 [===========================] - 2s 131ms/step Validation Accuracy = 0.85 Validation Precision = 0.97 Validation Recall = 0.85 12/12 [===========================] - 1s 130ms/step Test Accuracy = 0.88 Test Precision = 0.98 Test Recall = 0.87
Test		88	98	87	359/359 [===========================] - 5s 13ms/step
Validation	Our model + RDA [63]	89	97	89	12/12 [===========================] - 2s 127ms/step Validation Accuracy = 0.89 Validation Precision = 0.97 Validation Recall = 0.89 12/12 [===========================] - 1s 129ms/step Test Accuracy = 0.91 Test Precision = 0.99 Test Recall = 0.91
Test		91	89	91	359/359 [===========================] - 5s 13ms/step
Validation	Our model + MVO [40]	93	97	95	12/12 [===========================] - 2s 130ms/step Validation Accuracy = 0.93 Validation Precision = 0.97 Validation Recall = 0.95 12/12 [===========================] - 1s 129ms/step Test Accuracy = 0.94 Test Precision = 0.97 Test Recall = 0.96
Test		94	97	96	359/359 [===========================] - 5s 15ms/step

In terms of overall metrics, the RDA performs better than the WOA but still falls short of the SFLA and the MVO. The RDA provides low recall (89% in recall tests) but good precision (97% in validation), with validation accuracy at 89% and test accuracy at 91%. The RDA has good precision, but it may not be the best option in situations where recall is important, like in medical diagnosis, where it could be expensive to overlook positive cases. Lastly, the MVO algorithm performs better than the other approaches on nearly all criteria. It achieves good precision (97%) and recall (95% validation, 96% test) as well as the highest validation accuracy (93%) and test accuracy (94%). The MVO is the most dependable optimization technique for brain MRI classification because of its capacity to strike a compromise between precision and recall, minimizing false positives and false negatives. This method is the most appropriate optimization technique for this problem because of its performance stability across validation and test sets, which indicates that it is the best option for obtaining both high accuracy and robustness.

5. Conclusions and Outlook

This study has successfully revealed the application of advanced DL techniques in the classification of brain MRI images into tumor and non-tumor categories. By integrating two robust CNN architectures, ResNet50 and VGG16, and optimizing them with the MVO, the research has paved the way for more accurate and efficient diagnostic approaches in medical imaging. The preprocessing phase, including image resizing, grayscale conversion, and Gaussian blurring, has proven to be effective in preparing the MRI images for analysis. The optimization of data augmentation parameters and trainable layers utilizing the MVO has further enhanced the models' performance, resulting in high precision, accuracy, and recall. These findings underscore the potential of combining multiple CNN architectures and advanced optimization techniques in tackling the complexities of medical image classification.

Looking forward, there are several avenues for expanding upon this study. Future studies could explore the integration of additional neural network architectures and compare their effectiveness in brain tumor classification. The applicability of the suggested methodology to other types of medical imaging data, such as PET images or CT scans, is another area worth investigating. Additionally, incorporating more advanced forms of metaheuristic algorithms could lead to even better optimization results. As ML and DL continue to evolve, their application in healthcare promises to bring significant advancements. The techniques developed in this study contribute to this ongoing progression and open up new possibilities for improving diagnostic accuracy and patient care in the fields of neurology and oncology.

One potential limitation in the dataset selection for this study is that the brain MRI scans are derived from a singular medical source or demographic group, thus failing to encompass the diversity prevalent in worldwide populations. This may affect the model's capacity to generalize well when utilized on MRI scans from diverse institutions, scanners, or patient demographics with differing attributes, such as age, family history, or preexisting health issues. Moreover, the dataset predominantly emphasizes binary classification (tumor versus non-tumor), which may restrict the model's utility in more intricate situations necessitating the distinction among diverse tumor types or grades. Consequently, although our suggested method exhibits superior performance on the chosen dataset, further validation and testing on more varied and multicenter datasets are essential to ascertain its resilience and dependability over a broader spectrum of clinical applications.

Author Contributions: N.T.S.: Software, Conceptualization, Methodology, Investigation; S.S.: Software, Conceptualization, Methodology, Investigation; M.K.: Methodology, Software, Investigation; M.A.: Writing—Reviewing and Editing, Methodology, Formal analysis; S.J.G.: Supervisor, Validation, Formal analysis; R.R.: Supervisor, Writing—Reviewing and Editing, Validation. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Dataset available at: https://www.kaggle.com/datasets/sartajbhuvaji/brain-tumor-classification-mri/data (accessed on 4 May 2024).

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Badža, M.M.; Barjaktarović, M. Segmentation of Brain Tumors from MRI Images Using Convolutional Autoencoder. *Appl. Sci.* **2021**, *11*, 4317. [CrossRef]
2. Choi, S.-G.; Sohn, C.-B. Detection of HGG and LGG Brain Tumors using U-Net. *Medico-Legal Updat.* **2019**, *19*, 560–565. [CrossRef]
3. Wijethilake, N.; Meedeniya, D.; Chitraranjan, C.; Perera, I.; Islam, M.; Ren, H. Glioma Survival Analysis Empowered With Data Engineering—A Survey. *IEEE Access* **2021**, *9*, 43168–43191. [CrossRef]
4. Ranjbarzadeh, R.; Caputo, A.; Tirkolaee, E.B.; Ghoushchi, S.J.; Bendechache, M. Brain tumor segmentation of MRI images: A comprehensive review on the application of artificial intelligence tools. *Comput. Biol. Med.* **2022**, *152*, 106405. [CrossRef]
5. Dasanayaka, S.; Silva, S.; Shantha, V.; Meedeniya, D.; Ambegoda, T. Interpretable Machine Learning for Brain Tumor Analysis Using MRI. In Proceedings of the ICARC 2022—2nd International Conference on Advanced Research in Computing: Towards a Digitally Empowered Society, Belihuloya, Sri Lanka, 23–24 February 2022; pp. 212–217. [CrossRef]
6. Raj, R.; Luostarinen, T.; Pursiainen, E.; Posti, J.P.; Takala, R.S.K.; Bendel, S.; Konttila, T.; Korja, M. Machine learning-based dynamic mortality prediction after traumatic brain injury. *Sci. Rep.* **2019**, *9*, 17672. [CrossRef]
7. Karkehabadi, A.; Bakhshi, M.; Razavian, S.B. Optimizing Underwater IoT Routing with Multi-Criteria Decision Making and Uncertainty Weights. May 2024. Available online: https://arxiv.org/abs/2405.11513v1 (accessed on 6 June 2024).
8. Karkehabadi, A.; Homayoun, H.; Sasan, A. FFCL: Forward-Forward Net with Cortical Loops, Training and Inference on Edge Without Backpropagation. *arXiv* **2024**, arXiv:2405.12443. [CrossRef]

9. Zhang, L.; Tan, J.; Han, D.; Zhu, H. From machine learning to deep learning: Progress in machine intelligence for rational drug discovery. *Drug Discov. Today* **2017**, *22*, 1680–1685. [CrossRef]

10. Saryazdi, S.M.E.; Etemad, A.; Shafaat, A.; Bahman, A.M. Data-driven performance analysis of a residential building applying artificial neural network (ANN) and multi-objective genetic algorithm (GA). *Build. Environ.* **2022**, *225*, 109633. [CrossRef]

11. Anari, S.; Sarshar, N.T.; Mahjoori, N.; Dorosti, S.; Rezaie, A. Review of Deep Learning Approaches for Thyroid Cancer Diagnosis. *Math. Probl. Eng.* **2022**, *2022*, 5052435. [CrossRef]

12. Ranjbarzadeh, R.; Keles, A.; Crane, M.; Anari, S.; Bendechache, M. Secure and Decentralized Collaboration in Oncology: A Blockchain Approach to Tumor Segmentation. In Proceedings of the 2024 IEEE 48th Annual Computers, Software, and Applications Conference (COMPSAC), Osaka, Japan, 2–4 July 2024; pp. 1681–1686. [CrossRef]

13. Pandey, B.; Pandey, D.K.; Mishra, B.P.; Rhmann, W. A comprehensive survey of deep learning in the field of medical imaging and medical natural language processing: Challenges and research directions. *J. King Saud Univ. Comput. Inf. Sci.* **2022**, *34*, 5083–5099. [CrossRef]

14. Conneau, A.; Kiela, D.; Schwenk, H.; Barrault, L.; Bordes, A. Supervised Learning of Universal Sentence Representations from Natural Language Inference Data. EMNLP 2017—Conference on Empirical Methods in Natural Language Processing, Proceedings, pp. 670–680. May 2017. Available online: http://arxiv.org/abs/1705.02364 (accessed on 15 June 2020).

15. Sarshar, N.T.; Mirzaei, M. Premature Ventricular Contraction Recognition Based on a Deep Learning Approach. *J. Heal. Eng.* **2022**, *2022*, 1450723. [CrossRef] [PubMed]

16. Sarshar, N.T.; Abdossalehi, M. Automated Cardiovascular Arrhythmia Classification Based on Through Nonlinear Features and Tunable-Q Wavelet Transform (TQWT) Based Decomposition. *Rev. Comput. Eng. Stud.* **2021**, *8*, 35–41. [CrossRef]

17. Gambella, C.; Ghaddar, B.; Naoum-Sawaya, J. Optimization problems for machine learning: A survey. *Eur. J. Oper. Res.* **2021**, *290*, 807–828. [CrossRef]

18. Ver Berne, J.; Saadi, S.B.; Politis, C.; Jacobs, R. A deep learning approach for radiological detection and classification of radicular cysts and periapical granulomas. *J. Dent.* **2023**, *135*, 104581. [CrossRef]

19. Liu, Z.; Tong, L.; Chen, L.; Jiang, Z.; Zhou, F.; Zhang, Q.; Zhang, X.; Jin, Y.; Zhou, H. Deep learning based brain tumor segmentation: A survey. *Complex Intell. Syst.* **2022**, *9*, 1001–1026. [CrossRef]

20. Sailunaz, K.; Alhajj, S.; Özyer, T.; Rokne, J.; Alhajj, R. A survey on brain tumor image analysis. *Med. Biol. Eng. Comput.* **2024**, *62*, 1–45. [CrossRef]

21. Chahal, P.K.; Pandey, S.; Goel, S. A survey on brain tumor detection techniques for MR images. *Multimedia Tools Appl.* **2020**, *79*, 21771–21814. [CrossRef]

22. Dasanayaka, S.; Shantha, V.; Silva, S.; Meedeniya, D.; Ambegoda, T. Interpretable machine learning for brain tumour analysis using MRI and whole slide images. *Softw. Impacts* **2022**, *13*, 100340. [CrossRef]

23. Vadhnani, S.; Singh, N. Brain tumor segmentation and classification in MRI using SVM and its variants: A survey. *Multimedia Tools Appl.* **2022**, *81*, 31631–31656. [CrossRef]

24. Raza, R.; Bajwa, U.I.; Mehmood, Y.; Anwar, M.W.; Jamal, M.H. dResU-Net: 3D deep residual U-Net based brain tumor segmentation from multimodal MRI. *Biomed. Signal Process. Control.* **2022**, *79*, 103861. [CrossRef]

25. Zhu, Z.; He, X.; Qi, G.; Li, Y.; Cong, B.; Liu, Y. Brain tumor segmentation based on the fusion of deep semantics and edge information in multimodal MRI. *Inf. Fusion* **2022**, *91*, 376–387. [CrossRef]

26. Parhizkar, M.; Amirfakhrian, M. Car detection and damage segmentation in the real scene using a deep learning approach. *Int. J. Intell. Robot. Appl.* **2022**, *6*, 231–245. [CrossRef]

27. Safavi, S.; Jalali, M. RecPOID: POI Recommendation with Friendship Aware and Deep CNN. *Futur. Internet* **2021**, *13*, 79. [CrossRef]

28. Ranjbarzadeh, R.; Crane, M.; Bendechache, M. The Impact of Backbone Selection in Yolov8 Models on Brain Tumor Localization. 2024; preprint. [CrossRef]

29. Vankdothu, R.; Hameed, M.A. Brain tumor MRI images identification and classification based on the recurrent convolutional neural network. *Meas. Sensors* **2022**, *24*, 100412. [CrossRef]

30. Siddiqi, M.H.; Alsayat, A.; Alhwaiti, Y.; Azad, M.; Alruwaili, M.; Alanazi, S.; Kamruzzaman, M.M.; Khan, A. A Precise Medical Imaging Approach for Brain MRI Image Classification. *Comput. Intell. Neurosci.* **2022**, *2022*, 6447769. [CrossRef]

31. Ullah, Z.; Farooq, M.U.; Lee, S.-H.; An, D. A hybrid image enhancement based brain MRI images classification technique. *Med. Hypotheses* **2020**, *143*, 109922. [CrossRef]

32. Alrashedy, H.H.N.; Almansour, A.F.; Ibrahim, D.M.; Hammoudeh, M.A.A. BrainGAN: Brain MRI Image Generation and Classification Framework Using GAN Architectures and CNN Models. *Sensors* **2022**, *22*, 4297. [CrossRef]

33. Khan, M.A.; Khan, A.; Alhaisoni, M.; Alqahtani, A.; Alsubai, S.; Alharbi, M.; Malik, N.A.; Damaševičius, R. Multimodal brain tumor detection and classification using deep saliency map and improved dragonfly optimization algorithm. *Int. J. Imaging Syst. Technol.* **2022**, *33*, 572–587. [CrossRef]

34. Balaha, H.M.; Hassan, A.E.S. A variate brain tumor segmentation, optimization, and recognition framework. *Artif. Intell. Rev.* **2023**, *56*, 7403–7456. [CrossRef]

35. Deepa, S.; Janet, J.; Sumathi, S.; Ananth, J.P. Hybrid Optimization Algorithm Enabled Deep Learning Approach Brain Tumor Segmentation and Classification Using MRI. *J. Digit. Imaging* **2023**, *36*, 847–868. [CrossRef] [PubMed]

36. Ezugwu, A.E.; Shukla, A.K.; Nath, R.; Akinyelu, A.A.; Agushaka, J.O.; Chiroma, H.; Muhuri, P.K. Metaheuristics: A comprehensive overview and classification along with bibliometric analysis. *Artif. Intell. Rev.* **2021**, *54*, 4237–4316. [CrossRef]

37. França, R.P.; Monteiro, A.C.B.; Estrela, V.V.; Razmjooy, N. Using Metaheuristics in Discrete-Event Simulation. In *Lecture Notes in Electrical Engineering*; Springer Science and Business Media Deutschland GmbH: Cham, Switzerland, 2021; Volume 696, pp. 275–292. [CrossRef]

38. Razmjooy, N.; Ashourian, M.; Foroozandeh, Z. (Eds.) Metaheuristics and Optimization in Computer and Electrical Engineering. In *Lecture Notes in Electrical Engineering*; Springer International Publishing: Cham, Switzerland, 2021; Volume 696. [CrossRef]

39. Hu, A.; Razmjooy, N. Brain tumor diagnosis based on metaheuristics and deep learning. *Int. J. Imaging Syst. Technol.* **2020**, *31*, 657–669. [CrossRef]

40. Mirjalili, S.; Mirjalili, S.M.; Hatamlou, A. Multi-Verse Optimizer: A nature-inspired algorithm for global optimization. *Neural Comput. Appl.* **2015**, *27*, 495–513. [CrossRef]

41. Son, P.V.H.; Dang, N.T.N. Solving large-scale discrete time–cost trade-off problem using hybrid multi-verse optimizer model. *Sci. Rep.* **2023**, *13*, 1987. [CrossRef]

42. Han, Y.; Chen, W.; Heidari, A.A.; Chen, H.; Zhang, X. A solution to the stagnation of multi-verse optimization: An efficient method for breast cancer pathologic images segmentation. *Biomed. Signal Process. Control.* **2023**, *86*, 105208. [CrossRef]

43. Haseeb, A.; Waleed, U.; Ashraf, M.M.; Siddiq, F.; Rafiq, M.; Shafique, M. Hybrid Weighted Least Square Multi-Verse Optimizer (WLS–MVO) Framework for Real-Time Estimation of Harmonics in Non-Linear Loads. *Energies* **2023**, *16*, 609. [CrossRef]

44. Xu, W.; Yu, X. A multi-objective multi-verse optimizer algorithm to solve environmental and economic dispatch. *Appl. Soft Comput.* **2023**, *146*, 110650. [CrossRef]

45. Shorten, C.; Khoshgoftaar, T.M. A survey on Image Data Augmentation for Deep Learning. *J. Big Data* **2019**, *6*, 60. [CrossRef]

46. Nalepa, J.; Marcinkiewicz, M.; Kawulok, M. Data Augmentation for Brain-Tumor Segmentation: A Review. *Front. Comput. Neurosci.* **2019**, *13*, 83. [CrossRef]

47. Zeiser, F.A.; da Costa, C.A.; Zonta, T.; Marques, N.M.C.; Roehe, A.V.; Moreno, M.; Righi, R.d.R. Segmentation of Masses on Mammograms Using Data Augmentation and Deep Learning. *J. Digit. Imaging* **2020**, *33*, 858–868. [CrossRef] [PubMed]

48. de la Rosa, F.L.; Gómez-Sirvent, J.L.; Sánchez-Reolid, R.; Morales, R.; Fernández-Caballero, A. Geometric transformation-based data augmentation on defect classification of segmented images of semiconductor materials using a ResNet50 convolutional neural network. *Expert Syst. Appl.* **2022**, *206*, 117731. [CrossRef]

49. Deepa, N.; Chokkalingam, S. Optimization of VGG16 utilizing the Arithmetic Optimization Algorithm for early detection of Alzheimer's disease. *Biomed. Signal Process. Control.* **2022**, *74*, 103455. [CrossRef]

50. Zhu, F.; Li, J.; Zhu, B.; Li, H.; Liu, G. UAV remote sensing image stitching via improved VGG16 Siamese feature extraction network. *Expert Syst. Appl.* **2023**, *229*, 120525. [CrossRef]

51. Bakasa, W.; Viriri, S. VGG16 Feature Extractor with Extreme Gradient Boost Classifier for Pancreas Cancer Prediction. *J. Imaging* **2023**, *9*, 138. [CrossRef]

52. Sarker, S.; Tushar, S.N.B.; Chen, H. High accuracy keyway angle identification using VGG16-based learning method. *J. Manuf. Process.* **2023**, *98*, 223–233. [CrossRef]

53. Mpova, L.; Shongwe, T.C.; Hasan, A. The Classification and Detection of Cyanosis Images on Lightly and Darkly Pigmented Individual Human Skins Applying Simple CNN and Fine-Tuned VGG16 Models in TensorFlow's Keras API. In Proceedings of the 2023 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA), Gammarth, Tunisia, 12 June 2023; pp. 1–6. [CrossRef]

54. Zhang, Y.; Liu, Y.-L.; Nie, K.; Zhou, J.; Chen, Z.; Chen, J.-H.; Wang, X.; Kim, B.; Parajuli, R.; Mehta, R.S.; et al. Deep Learning-based Automatic Diagnosis of Breast Cancer on MRI Using Mask R-CNN for Detection Followed by ResNet50 for Classification. *Acad. Radiol.* **2023**, *30*, S161–S171. [CrossRef]

55. Sharma, A.K.; Nandal, A.; Dhaka, A.; Zhou, L.; Alhudhaif, A.; Alenezi, F.; Polat, K. Brain tumor classification using the modified ResNet50 model based on transfer learning. *Biomed. Signal Process. Control.* **2023**, *86*, 105299. [CrossRef]

56. Lee, J.-R.; Ng, K.-W.; Yoong, Y.-J. Face and Facial Expressions Recognition System for Blind People Using ResNet50 Architecture and CNN. *J. Informatics Web Eng.* **2023**, *2*, 284–298. [CrossRef]

57. Hossain, B.; Iqbal, S.H.S.; Islam, M.; Akhtar, N.; Sarker, I.H. Transfer learning with fine-tuned deep CNN ResNet50 model for classifying COVID-19 from chest X-ray images. *Informatics Med. Unlocked* **2022**, *30*, 100916. [CrossRef]

58. Brain Tumor Classification (MRI). Available online: https://www.kaggle.com/datasets/sartajbhuvaji/brain-tumor-classification-mri/data (accessed on 4 May 2024).

59. Krishnapriya, S.; Karuna, Y. Pre-trained deep learning models for brain MRI image classification. *Front. Hum. Neurosci.* **2023**, *17*, 1150120. [CrossRef] [PubMed]

60. Gore, D.V.; Sinha, A.K.; Deshpande, V. Automatic CAD System for Brain Diseases Classification Using CNN-LSTM Model. In Proceedings of the Emerging Technologies in Data Mining and Information Security. Advances in Intelligent Systems and Computing; Springer: Singapore, 2023; pp. 623–634. [CrossRef]

61. Saadi, S.B.; Sarshar, N.T.; Ranjbarzadeh, R.; Forooshani, M.K.; Bendechache, M. Investigation of Effectiveness of Shuffled Frog-Leaping Optimizer in Training a Convolution Neural Network. *J. Heal. Eng.* **2022**, *2022*, 1–11. [CrossRef] [PubMed]

62. Mirjalili, S.; Lewis, A. The Whale Optimization Algorithm. *Adv. Eng. Softw.* **2016**, *95*, 51–67. [CrossRef]

63. Fathollahi-Fard, A.M.; Hajiaghaei-Keshteli, M.; Tavakkoli-Moghaddam, R. Red deer algorithm (RDA): A new nature-inspired meta-heuristic. *Soft Comput.* **2020**, *24*, 14637–14665. [CrossRef]