Analysis of Deep Learning based Ensemble Models for Brain Tumor Detection

Sachin Minocha
School of Computer Science Engineering and Technology
Bennett University
Greater Noida, India
sachin0111@gmail.com

Vedant Nigam
School of Computer Science Engineering and Technology
Bennett University
Greater Noida, India
nigamvedant100@gmail.com

Utsav Birla
School of Computer Science Engineering and Technology
Bennett University
Greater Noida, India
utsavbirla2004@gmail.com

Hardik Kashyap
School of Computer Science Engineering and Technology
Bennett University
Greater Noida, India
kashyap.hardik789@gmail.com

***Abstract*—Brain tumor is one of the severe diseases that can lead to human death. The probability of success for human treatment increases if the tumor is detected at an early stage. This paper works on the transfer learning models for brain tumor detection. Transfer learning is one of the most useful deep learning techniques which offers to reutilization of the pre-trained models on new tasks. This paper analyzed the performance of pre-trained models, i.e., VGG16, Xception, VGG19, and ResNet50, on three different MRI based datasets of varying sizes. On the basis of the performance of these models, this work continues the analysis using different ensembles of VGG16, ResNet50, and Xception. Overall, this work analyzes twelve ensembles, which include three weighted ensembles for VGG16+ResNet50, VGG16+Xception, Xception+ResNet50, and four weighted ensembles of VGG16+Resnet50+Xception. This work shows that the VGG16+ResNet50 ensemble with 25% weightage to ResNet50 and 75% weightage of VGG16, gives 94.04%, 98.6%, and 99.4% accuracy on dataset1, dataset2, and dataset3, respectively, to outperform all other individual and ensemble models.**

Keywords— CNN, Deep Learning, Transfer Learning, Ensemble Learning, MRI, VGG16, ResNet50, Xception, Brain Tumor.

I. INTRODUCTION

The brain is the main organ of the human body, on which many other organs and human activities depend. It controls the majority of body functions, such as walking, eating, feeling, breathing, thinking, and watching [1]. Therefore, it is necessary to maintain a healthy brain by providing it with all the necessary vitamins and nutrition through a healthy lifestyle. The brain may get affected by some bacteria or viruses, which can lead to diseases like Alzheimer's, cancer, and Parkinson's disease. In diseases like cancer, a tumor grows in the brain that interferes with brain functions, resulting in an impact on basic human activities like breathing, thinking, and even death. Brain cancer can be treated using radiation or surgery if it is detected at an early stage. Due to the complex brain structure, brain tumor detection is a tedious task. However, technological advancements in medical imaging, such as MRI and CT, have helped us achieve it.

Deep learning has revolutionized medical systems and has shown promising results for processing images [2]. Medical images can be processed using deep learning techniques to analyze brain cancer, removing any scope for human error in such a fatal disease. Various researchers are working on deep learning techniques like AlexNet, VGG16, VGG19, and ResNet to detect brain tumors. The researchers are also working on hybrid models of deep learning for better results in brain tumor detection.

Aamir et al. [1] have introduced hypermetric CNN, which fine-tunes hyperparameters like learning rate, batch size, activation function, filter size, pooling, layer count, and padding. The authors analyzed the hypermetric CNN on the MRI dataset and obtained 97% recall, accuracy, F1 Score, and precision. Abdusalomov et al. [3] have used the YOLOv7 with added layers such as the Spatial Pyramid Pooling Fast+ (SPPF+) layer, Convolutional Block Attention Module (CBAM) mechanism, and Bidirectional Feature Pyramid Network (BiFPN) to enhance its feature extraction capability. The authors analyzed the model over curated data widely available containing 3 types of tumors i.e., gliomas, meningioma, and pituitary, using data augmentation techniques.

Anantharajan et al. [4] used the Gray-Level Co-occurrence Matrix (GLCM) for extracting features like energy, mean and entropy. The classification is done by the Ensemble Deep Neural Support Vector Machine(EDN-SVM) classifier. They achieved 97.93% accuracy, 92% sensitivity, and 98% specificity in recognizing the brain with tumor and nontumor MRI images. Ullah et al. [5] introduced the TumorDetNet model, having 48 convolution layers with ReLU activation function, leaky ReLU, dropout layer, and average pooling and finally added one fully connected layer and softmax layer in the TumorDetNet framework. The model gives an accuracy of 99.83% for brain tumor detection. Khan et al.[6] proposed two deep learning models using MRI images, achieving up to 97.8% and 100% accuracy over two datasets. A 23-layered CNN is used for large datasets, while transfer learning with VGG16 addresses overfitting in smaller datasets. Ahmed [7] has proposed a hybrid ViT-GRU (Vision Transformer with Gated Recurrent Unit) model. They extracted features via a vision transformer and used GRU to identify the relationship between the features. The model training has been done using various optimizers like SGD, Adam, and AdamW to achieve accuracy of 81.66%, 96.56%, and 98.97%, respectively.

Abu et al. [8] proposed a model similar to VGG16 but the final max pooling layer is replaced by an average pooling layer. This layer performs spatial pooling of the feature map to reduce overfitting in the model. They achieved 0.93 precision, 1.00 sensitivity, 0.97 F1 Score, 0.93 average precision-recall score, and 0.91 Cohen's Kappa. CELIK et al. [9] use MobileNetV3 and EfficientNetB7 for feature extraction, integrate the attention mechanism in their model, and finally classify the image using the softmax function. This model gives better results with an accuracy of 98.94% on the brain MRI dataset.

The existing models have shown significant performance on different datasets for brain tumor detection. However, the accuracy of datasets having images with different modalities is less due to the limited use of ensemble learning, which is a vast research area. That's why this work analyses different ensemble methods of three deep learning models, i.e., VGG16, ResNet50, and Xception, for efficient detection of brain tumors. Overall, the main contributions of this work are as follows:

- Analysis of transfer learning models using VGG16, ResNet50, VGG19, and Xception over three different MRI image datasets of small, medium, and large sizes.
- Designing twelve ensemble transfer learning models using VGG16, ResNet50, and Xception. This includes four weight based ensembles of VGG16+RestNet50+ Xception.
- Classifying the brain tumor MRI images using the proposed architecture and comparison with different state-of-the-art transfer learning techniques to prove the effectiveness of the proposed work

The rest of the paper is divided into four sections. Section 2 discusses the model architecture of VGG16, ResNet50, and Xception. The next section, i.e., section 3, describes the ensemble model in detail. Section 4 presents the results and corresponding discussion. Finally, section 5 concludes the whole study.

II. BACKGROUND

This section discusses the already available transfer learning models, i.e., VGG16, ResNet50, and Xception.

A. VGG16

VGG16 is a 16-layer CNN architecture developed by Simonyan and Zisserman in 2015 [10].This particular model has been used to provide solutions to the problems posed by the ILSVRC 2014 competition [11].

VGG16 architecture, as shown in Fig. 1, has 5 blocks. Each block consists of a number of convolutional layers and a max pooling layer. Fig. 1 shows the input RGB image of size 224x224 to the first block convolution layers having 64 batch size followed by max pooling, which gives an output image of size 112x112. This process is repeated for each block with increased batch size, leading to a 7x7 image from block 5.

The latter 7x7 size image tensor is fed to three subsequent fully connected dense layers. The first two dense layers contain 4096 units and the ReLU activation function for performing non-linear transformation. The last layer consists of 1000 units and uses softmax activation to classify the image in any one of the ILSVRC's thousand classes. VGG19 is an extension of VGG16 containing 3 more layers.

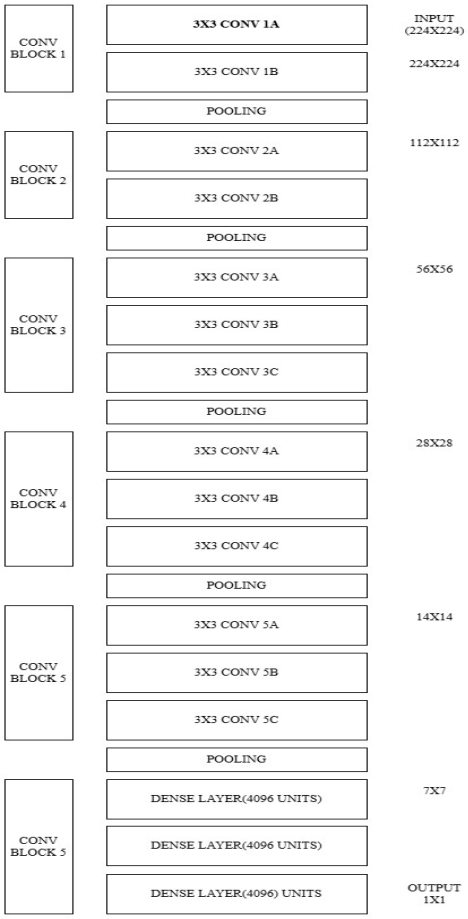

Fig. 1. VGG16 Architecture

B. ResNet50

Residual network is a deep convolutional neural network that achieved leading results in the classification segment of the ILSVRC 2015 competition [12]. This paper utilizes compact versions of residual network namely ResNet50, a model with 50 layers. ResNet50 shares the foundational structure of VGG but incorporates additional layers and residual learning, enabling each layer's input to merge with its output [13].

The 50-layer ResNet comprises 5 convolution blocks as shown in Fig. 2. The input format is an image of size 224x224x3. The first block consists of a single 7x7 convolution layer, producing a reduced 112x112 output image, which is processed through a max-pooling layer. The next stage, Block 2, has three convolution layers with configuration: 1x1 with 64 filters, 3x3 with 64 filters, and 1x1 with 256 filters. This block's output is a 56x56 image. Block 3 features four layers per configuration, utilizing 1x1 with 128 filters,3x3 with 128 filters, and 1x1 with 512 filters, yielding a total of 12 layers in this block and a 28x28 output image. Block 4 receives the 28x28 input and includes six layers having configuration, 1x1 256 filters, 3x3 with 256 filters, and 1x1 with 1024 filters, and produces a 14x14 output. The final block, Block 5, has three layers per configuration with 1x1, 3x3, and 1x1 convolutions set to 512, 512, and 2048 filters respectively. It produces a 7x7 output. Unlike VGG, ResNet50 enhances model learning with its deeper architecture. The 7x7 image undergoes average

pooling, followed by a fully connected dense layer with softmax activation and 1000 units, aligning with the number of classes for image prediction in the ILSVRC 2015 challenge.

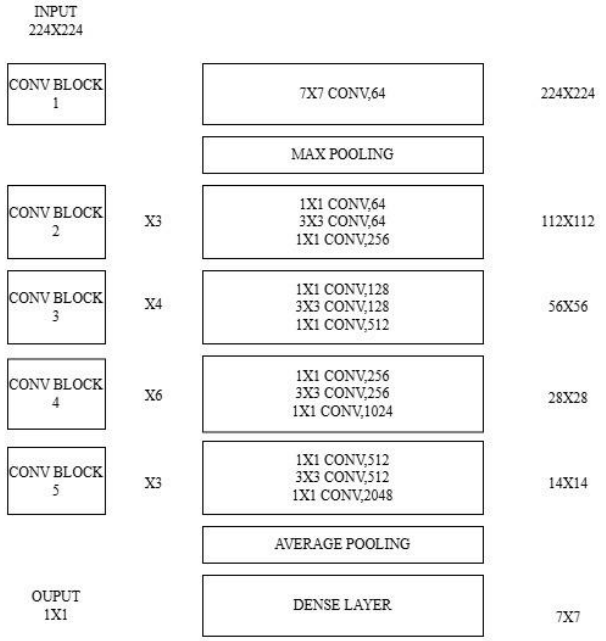

Fig. 2. ResNet50 Architecture

C. Xception

Xception is an advanced type of deep convolutional neural network that has been designed on Inception and depth-wise separable convolution principles [14]. The Xception model features a completely separable convolution architecture, which allows it to be efficiently utilized by separating cross-channel and spatial information [15].

Xception starts with a 299x299 input RGB image, which is decreased to 149x149 by a max pooling layer and 3x3 convolution. This model consists of 36 depth-wise separable convolutional layers organized into three primary modules: Entry, Middle, and Exit. Three convolutional blocks are part of the Entry module, which starts the feature extraction process. Each block includes batch normalization after depth-wise separable convolution and ReLU activation. After this module, the feature map shrinks to 75x75 through pooling operations.

The middle module, the central component of the model, consists of eight depth-wise separable convolutional blocks. Without altering the image size, each block employs a sequence of 3x3 depth-wise separable convolutions.

In the Exit module, a global average pooling layer follows 2 additional convolutional blocks with 1x1 and 3x3 depth-wise separable convolutions. This module captures high-level abstract features by reducing the feature map to 10x10. This model can classify images and outperform Inception on a number of benchmarks, achieving high accuracy while retaining efficiency [15].

III. Proposed Work

This work aims to improve the detection of Brain tumors through transfer learning. Transfer learning enables us to use a pre-trained model as the starting point for a new task. Based on experimental data, this work selects VGG16, Xception, and ResNet50 models to define new ensemble models to improve learning and outcomes. This work proposes twelve weighted ensemble methods, as shown in Table 1, where each model in the ensemble has a weight that defines its contribution towards the output.

TABLE I. Proposed Ensemble Methods

Ensemble Method	Weights
ResNet50*w_1+ VGG16*w_2 + Xception*w_3	$w_1, w_2, w_3 = 0.33$
	$w_1 = 0.5, w_2 = 0.5, w_3 = 0.5$
	$w_1 = 0.25, w_2 = 0.5, w_3 = 0.25$
	$w_1 = 0.25, w_2 = 0.25, w_3 = 0.5$
ResNet50*w_1+VGG16*w_2	$w_1 = 0.5, w_2 = 0.5$
	$w_1 = 0.75, w_2 = 0.25$
	$w_1 = 0.25, w_2 = 0.75$
VGG16*w_1+Xception*w_2	$w_1 = 0.5, w_2 = 0.5$
	$w_1 = 0.75, w_2 = 0.25$
	$w_1 = 0.25, w_2 = 0.75$
ResNet50*w_1+Xception*w_2	$w_1 = 0.5, w_2 = 0.5$
	$w_1 = 0.75, w_2 = 0.25$
	$w_1 = 0.25, w_2 = 0.75$

Different ensemble methods presented in Table 1 are designed by giving equal weightage to each model or high weight to one model and splitting the remaining weight among other models. Moreover, the last fully connected layer of the ensemble model can be removed or customized according to the classification needs. Further subsections give a detailed discussion of two and three model based ensembles.

A. Ensemble VGG16 & ResNet50

The ensemble architecture of the VGG16 and ResNet50 shown in Fig. 3 consists of a combination of features from these models. By ensembling these two models, our hybrid architecture can capture a diverse set of features, from small details to large structural patterns, making the overall model more robust for brain tumor classification.

After feature extraction, a fully connected dense layer (with 4069 units and ReLU activation) is added, as shown in Fig. 3, to learn the complex pattern from the combined feature vectors. This extra layer increases the depth of the model and allows for further refinement of the features, leading to better performance in the final classification task. The two models run in parallel without their internal architecture being disturbed and the final layer predicts the output for the two classes using the softmax activation function. The final prediction $Pred_P$ is calculated using Eq. (1).

$$Pred_P = w_1 \times Pred_{ResNet50} + w_2 \times Pred_{VGG16} \quad (1)$$

where w_1 and w_2 are the respective weights assigned to the predictions from the Resnet50 and VGG16 models. The proposed ensemble model is analyzed upon three combinations of weights, as shown in in Table 1.

Fig. 3. ResNet50 and VGG16 Ensemble Architecture

Similarly, the final predictions for VGG16+Xception and ResNet50+Xception are computed using Eqs. (2) and (3), respectively.

$$Pred_P = w_1 \times Pred_{VGG16} + w_2 \times Pred_{Xception} \quad (2)$$

$$Pred_p = w_1 \times Pred_{ResNet50} + w_2 \times Pred_{Xception} \quad (3)$$

where w_1 and w_2 are the respective weights assigned to the predictions.

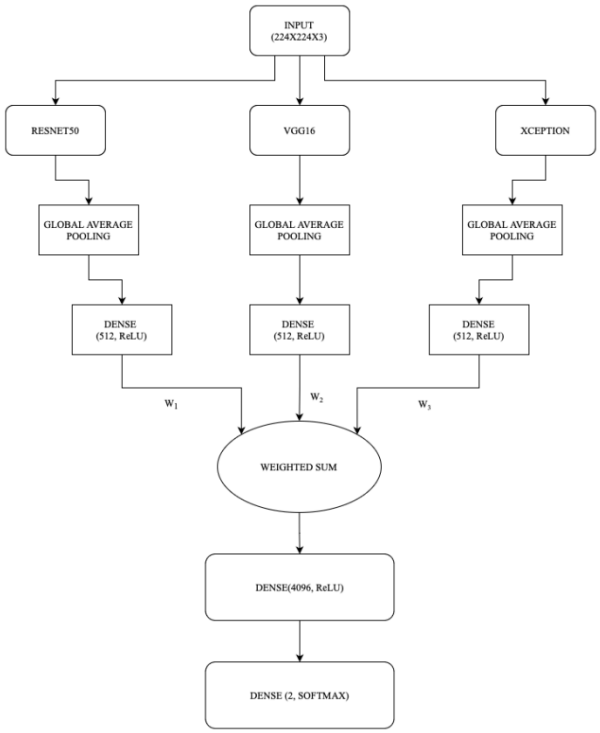

Fig. 4. ResNet50, Xception and VGG16 Ensemble Architecture

B. Ensemble ResNet50, Xception and VGG16

The ensemble architecture of the VGG16, Xception, and ResNet50 is shown in Fig. 4 consists of a combination of features from these models. The final prediction is computing using Eq. (4).

$$Pred_P = w_1 \times Pred_{ResNet50} + w_2 \times Pred_{Xception} + w_3 \times Pred_{VGG16} \quad (4)$$

where w_1, w_2 and w_3 are weights for resNet50, Xception and VGG16 models respectively. Ensembles with four combinations of weights (w1, w2, w3), i.e., equal (0.33,0.33,0.33), Xception dominance (0.25,0.5,0.25), VGG16 dominance (0.25,0.25,0.5), and ResNet50 dominance (0.5,0.25,0.25) are analyzed. The corresponding performance analysis is discussed in the next section.

IV. Results and Discussions

This section discusses the dataset used and the analysis of individual and proposed ensemble transfer learning models. Different hybrid models have been analyzed over three different datasets using a system with the M2 chip, the macOS Sequoia 15.0, and 8 GB memory. The analysis has been done on three publicly available MRI datasets obtained from Kaggle.com. Three datasets, say dataset1 [16], dataset2 [17], and dataset3 [18] have 253, 3060, and 21,673 samples, respectively. Each dataset has two classes i.e., Brain Tumor and No Tumor. The dataset of different sizes ensures comprehensive analysis. The whole analysis is shown in two phases, firstly, the analysis of the individual transfer learning models followed by the analysis of the ensemble methods.

A. Analysis of Individual Transfer Learning Models

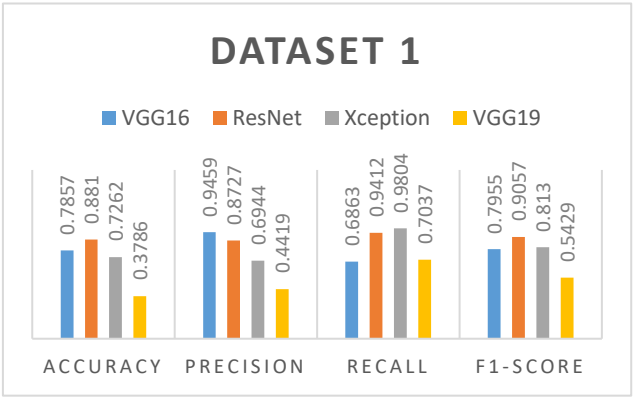

Fig. 5. Performance Analysis on Dataset1

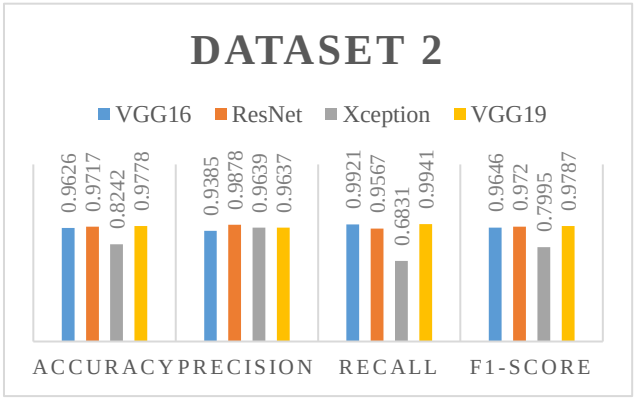

Fig. 6. Performance Analysis on Dataset2

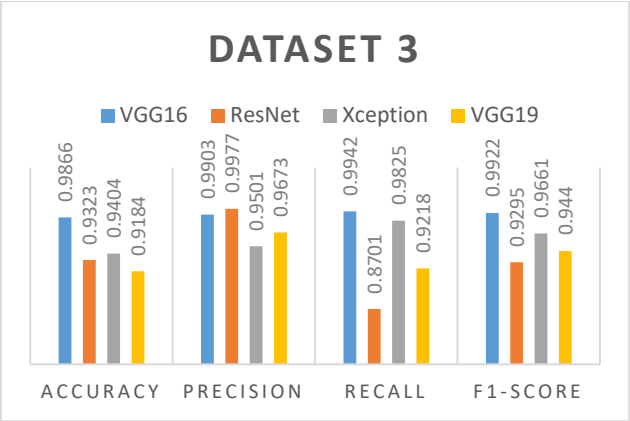

Fig. 7. Performance Analysis on Dataset2

Figs. 5, 6, and 7 present the analysis of ResNet, VGG16, Xception, and VGG19 on the three MRI datasets already described using the parameters Precision, Accuracy, F1-Score, and Recall. Fig. 5 shows that VGG19 exhibits the lowest performance for dataset1. Fig. 6 shows that the Xception model exhibits the lowest accuracy, F1-score, and recall for dataset2, while the VGG16 shows the lowest precision. In a similar fashion, Fig. 7 shows that the VGG19 model shows the lowest accuracy. The analysis of the ensemble models has been discussed in the next subsection.

B. Analysis of Ensemble methods

Table 2 shows the analysis of the proposed twelve ensemble methods. The analysis of all models in Table 2 clearly signifies that the ensemble of ResNet50 and VGG16, with 0.25 weight to ResNet50 and 0.75 weight to VGG16, i.e., the VGG16 dominated ResNet50+VGG16 ensemble, outperformed all other ensembles in terms of accuracy.

TABLE II. ANALYSIS OF ENSEMBLE METHODS

Ensemble Method	Weights	Parameter	Dataset1	Dataset2	Dataset3
ResNet50*w_1+ VGG16*w_2 + Xception*w_3	$w_1, w_2, w_3 = 0.33$	Accuracy	0.9048	0.9818	0.9901
		Loss	0.2181	0.0649	0.0368
	$w_1 = 0.5, w_2 = 0.5, w_3 = 0.5$	Accuracy	0.8929	0.9808	0.9905
		Loss	0.2698	0.0717	0.0291
	$w_1 = 0.25, w_2 = 0.5, w_3 = 0.25$	Accuracy	0.9048	0.9848	0.9897
		Loss	0.2155	0.0526	0.0342
	$w_1 = 0.25, w_2 = 0.25, w_3 = 0.5$	Accuracy	0.8214	0.9838	0.9920
		Loss	0.3518	0.0508	0.0280
ResNet50*w_1+VGG16*w_2	$w_1 = 0.5, w_2 = 0.5$	Accuracy	0.9167	0.9828	0.9894
		Loss	0.2250	0.0672	0.0289
	$w_1 = 0.75, w_2 = 0.25$	Accuracy	0.9048	0.9808	0.9897
		Loss	0.2598	0.0638	0.0304
	$w_1 = 0.25, w_2 = 0.75$	Accuracy	**0.9405**	**0.9869**	**0.9941**
		Loss	0.2841	0.0360	0.0268
VGG16*w_1+Xception*w_2	$w_1 = 0.5, w_2 = 0.5$	Accuracy	0.8810	0.9667	0.9826
		Loss	0.2993	0.1016	0.0321
	$w_1 = 0.75, w_2 = 0.25$	Accuracy	0.9167	0.9818	0.9868
		Loss	0.2212	0.0664	0.0464
	$w_1 = 0.25, w_2 = 0.75$	Accuracy	0.8690	0.9727	0.8453
		Loss	0.2732	0.1010	0.2954
ResNet50*w_1+Xception*w_2	$w_1 = 0.5, w_2 = 0.5$	Accuracy	0.8810	0.9727	0.9902
		Loss	0.3008	0.0806	0.0347
	$w_1 = 0.75, w_2 = 0.25$	Accuracy	0.8452	0.9808	0.9825
		Loss	0.5580	0.0585	0.0492
	$w_1 = 0.25, w_2 = 0.75$	Accuracy	0.7738	0.9646	0.9884
		Loss	0.4247	0.0957	0.0314

V. CONCLUSIONS

This work performs the analysis of ensemble methods for brain tumor detection. The analysis shows that the combination of VGG16 and ResNet50 with 0.75 and 0.25 weight, respectively, outperforms all other methods with an accuracy of 94.05%, 98.69%, and 99.41% on dataset1, dataset2, and dataset3, respectively. This result suggests that the selective use of different CNN architectures with weighted sum can outperform individual models, creating a robust system that maximizes diagnostic accuracy and reduces error. The thought of integrating different models with their specific strengths like VGG16's detail-oriented feature extraction and ResNet50's capability to capture complex patterns with deep layers, is found well-suited for analysing complex brain imaging data. Moreover, the three-model ensemble (VGG16, ResNet50, Xception) slightly underperformed, highlighting the importance of strategic

model selection rather than increasing model complexity. For future work, more experiments on the selected hybrid model, adding different layers, and randomizing the weights may result in higher accuracy levels. Using the hybrid model method, we can extend our approach to more medical fields, making it a valuable tool for healthcare providers.

REFERENCES

[1] M. Aamir et al., "Brain Tumor Detection and Classification Using an Optimized Convolutional Neural Network," Diagnostics, vol. 14, no. 16, Aug. 2024, doi: 10.3390/diagnostics14161714.

[2] M. Li, Y. Jiang, Y. Zhang, and H. Zhu, "Medical image analysis using deep learning algorithms," Front Public Health, vol. 11, 2023, doi: 10.3389/fpubh.2023.1273253.

[3] A. B. Abdusalomov, M. Mukhiddinov, and T. K. Whangbo, "Brain Tumor Detection Based on Deep Learning Approaches and Magnetic Resonance Imaging," Cancers (Basel), vol. 15, no. 16, Aug. 2023, doi: 10.3390/cancers15164172.

[4] S. Anantharajan, S. Gunasekaran, T. Subramanian, and V. R, "MRI brain tumor detection using deep learning and machine learning approaches," Measurement: Sensors, vol. 31, Feb. 2024, doi: 10.1016/j.measen.2024.101026.

[5] N. Ullah, A. Javed, A. Alhazmi, S. M. Hasnain, A. Tahir, and R. Ashraf, "TumorDetNet: A unified deep learning model for brain tumor detection and classification," PLoS One, vol. 18, no. 9 September, pp. 1–24, Sep. 2023, doi: 10.1371/journal.pone.0291200.

[6] M. S. I. Khan et al., "Accurate brain tumor detection using deep convolutional neural network," Comput Struct Biotechnol J, vol. 20, pp. 4733–4745, Jan. 2022, doi: 10.1016/j.csbj.2022.08.039.

[7] M. M. Ahmed et al., "Brain tumor detection and classification in MRI using hybrid ViT and GRU model with explainable AI in Southern Bangladesh," Sci Rep, vol. 14, no. 1, p. 22797, Dec. 2024, doi: 10.1038/s41598-024-71893-3.

[8] M. Abu et al., "Deep Convolutional Neural Networks Model-based Brain Tumor Detection in Brain MRI Images."

[9] F. Celik, K. Celik, and A. Celik, "Enhancing brain tumor classification through ensemble attention mechanism," Sci Rep, vol. 14, no. 1, p. 22260, Dec. 2024, doi: 10.1038/s41598-024-73803-z.

[10]K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," in ICLR 2015, Sep. 2014. [Online]. Available: http://arxiv.org/abs/1409.1556

[11]S. Tammina, "Transfer learning using VGG-16 with Deep Convolutional Neural Network for Classifying Images," International Journal of Scientific and Research Publications (IJSRP), vol. 9, no. 10, p. p9420, Oct. 2019, doi: 10.29322/ijsrp.9.10.2019.p9420.

[12]Z. Wu, C. Shen, and A. van den Hengel, "Wider or Deeper: Revisiting the ResNet Model for Visual Recognition," Pattern Recognit, vol. 90, pp. 119–133, Jun. 2019, doi: 10.1016/j.patcog.2019.01.006.

[13]A. Veit, M. Wilber, and S. Belongie, "Residual Networks Behave Like Ensembles of Relatively Shallow Networks," in Advances in neural information processing systems 29 (2016)., pp. 1–9.

[14]F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," in In Proceedings of the IEEE conference on computer vision and pattern recognition, 2017. [Online]. Available: http://arxiv.org/abs/1610.02357

[15] B. Gülmez, "A novel deep neural network model based Xception and genetic algorithm for detection of COVID-19 from X-ray images," Ann Oper Res, vol. 328, no. 1, pp. 617–641, Sep. 2023, doi: 10.1007/s10479-022-05151-y.

[16] [Dataet] https://www.kaggle.com/datasets/navoneel/brain-mri-images-for-brain-tumor-detection

[17] [Dataset] https://www.kaggle.com/datasets/abhranta/brain-tumor-detection-mri/data

[18] [Dataset] https://www.kaggle.com/datasets/mohammadhossein77/brain-tumors-dataset