

Leveraging Transfer Learning Using Convolutional Neural Networks to Enhance Brain Tumor Images Classification

Dian Kurniasari, Muhtarom Ahkam Maulana, Favorisen Rosyking Lumbanraja, and Warsono

Abstract— Effective treatment relies on accurately diagnosing brain tumors, which are characterized by abnormal cell proliferation. Artificial Intelligence (AI) offers a promising alternative to traditional diagnostic methods, which are frequently error-prone. This study aims to enhance the precision of brain tumor image classification using a Transfer Learning (TL) approach with Convolutional Neural Networks (CNNs). A dataset of 7,020 images were categorized into four categories: glioma, meningioma, pituitary tumor, and no tumor. This dataset was used to test several pre-trained models, including DenseNet121, InceptionResNetV2, MobileNetV2, NasNetMobile, and ResNet50V2. Performance was measured using accuracy, precision, sensitivity, and specificity metrics. The most effective of these was ResNet50V2, which achieved an accuracy of 97.70% and a loss of 0.066. A confusion matrix analysis of the results highlighted the model's exceptional performance, with sensitivity (97.70%), specificity (99.30%), and precision (97.80%). This research significantly contributes to medical image analysis, improving diagnostic accuracy using AI technology. The application of TL enhances early detection reduces and reduces the misdiagnosis by lowering the need for large datasets and minimizing errors. Furthermore, the model's efficiency in analyzing large numbers of MRI images significantly offers time-saving advantages for healthcare professionals, allowing them to prioritize more complex cases. This study advances the role of AI in medical diagnostics, particularly in brain tumor classification, with the potential to revolutionize early diagnosis, treatment strategies, and expand access to quality healthcare in underserved areas. By improving diagnostic accuracy, this model could contribute to reducing treatment delays, ultimately saving more lives.

Index Terms—Classification, Transfer learning, Brain tumor, Convolutional Neural Network

Manuscript received September 6, 2024; revised August 31, 2025.

D. Kurniasari is an Associate Professor of the Department of Mathematics at the Faculty of Mathematics and Natural Sciences, Universitas Lampung, Lampung, Indonesia (corresponding author to provide phone: 08127-7334-4561; email: dian.kurniasari@fmipa.unila.ac.id).

M. A. Maulana is Bachelor's Graduate from the Department of Mathematics at the Faculty of Mathematics and Natural Sciences, Universitas Lampung, Lampung, Indonesia (e-mail: muhtaromahkam0@gmail.com).

F. R. Lumbanraja is an Associate Professor of the Department of Computer Science at the Faculty of Mathematics and Natural Sciences, Universitas Lampung, Lampung, Indonesia (e-mail: favorisen.lumbanraja@fmipa.unila.ac.id).

Warsono is an Associate Professor of the Department of Mathematics at the Faculty of Mathematics and Natural Sciences, Universitas Lampung, Lampung, Indonesia (e-mail: warsono.1963@fmipa.unila.ac.id).

I. INTRODUCTION

Brain tumors present a significant medical challenge requiring accurate and timely detection to ensure the most optimal and possible treatment. These tumors are typically divided into two categories: primary and secondary. Primary brain tumors originate within the brain, whereas secondary brain tumors generally metastasize from other body regions. Tumors are further classified as malignant, which is cancerous, or benign, which is non-cancerous. Unlike benign brain tumors, malignant brain tumors exhibit a high proliferation rate and have the potential to invade multiple areas of the brain. Gliomas, meningiomas, and pituitary tumors are among the most prevalent types of brain tumors [1, 2].

Computer Vision, Artificial Intelligence (AI), and Machine Learning (ML) play an essential role in the early detection of cancer through advanced medical imaging techniques specifically Computed Tomography (CT) and Magnetic Resonance Imaging (MRI). Among these, MRI is particularly distinguished by its exceptional capability to depict various cerebral conditions comprehensively. This includes the detection of potential primary brain tumors, precise brain tumors location, and evaluation of related conditions particularly oedema, hemorrhage, and hydrocephalus [3]. Leveraging AI, ML, and computer vision technologies allows for the enhanced processing of MRI images, offering radiologists supplementary analytical insights and alternative perspectives [4, 5].

Vimala et al. [6] assert that Deep Learning (DL) has significantly advanced computer vision, particularly in medical image processing. Convolutional Neural Networks (CNNs) are frequently used and are highly effective in this field. CNN offers the ability to automatically extract critical features from images, thereby eliminating the need for manual feature engineering, which is typically required by traditional ML methodologies. However, CNNs require substantial data and computational resources to perform optimally. To mitigate this limitation, Transfer Learning (TL) can be employed by CNNs, which allows for the use of fewer training samples and shorter training times.

The TL is an advanced technique in ML and DL in which a model, initially trained on one dataset, is adapted for use on different datasets or tasks [7–9]. This strategy enables the model to leverage insights gained from the training data, enhancing its learning capabilities and applicability to novel or similar scenarios [10]. Numerous TL models have been developed using CNNs, including notable architectures such as Google Perception Net, ResNet, VGG, Xception, AlexNet, Inception, and DenseNet [11–17].

This paper compares various TL models, specifically DenseNet121, InceptionResNetV2, MobileNetV2, NasNetMobile, and ResNet50V2.

II. RELATED WORK

Narin et al. [18] suggest the application of five pre-trained CNN architectures—ResNet50, ResNet101, ResNet152, InceptionV3, and Inception-ResNetV2—for diagnosing coronavirus pneumonia through the analysis of chest X-ray images. The patients were categorized into four groups: normal, COVID-19, viral pneumonia, and bacterial pneumonia. The efficacy of the models was validated using k -fold cross-validation. Results indicated that the ResNet50 model demonstrates superior accuracy across three separate datasets, achieving accuracy rates of 96.1% for Dataset-1, 99.5% for Dataset-2, and 99.7% for Dataset-3. Hu et al. [19] evaluated the MobileNet and RseSK models against other advanced networks, specifically MobileNetV3 and Ghost, using the HAM10000 dataset. Their approach enhanced skin disease classification accuracy by 1.7% compared to existing methods, with the MobileNet and RseSK models attaining an accuracy of 85% on the test set, outperforming MobileNetV3.

Triyadi et al. [20] conducted a cataract classification study utilizing various features, including blood vessels and optical discs visible in retinal fundus images. With a primary focus on enabling early detection, they set out to create a software application that could reliably classify cataracts into two distinct classes: normal and cataracts. The classification was performed using VGG-19, ResNet50, and ResNet101 models, yielding accuracies of 91.06%, 93.50%, and 93.50%, respectively. In a separate study, Alruwaili and Gouda [21] developed a system leveraging TL to improve the classification of mammography images from the Mammographic Image Analysis Society (MIAS) dataset. The accuracy of their model achieved an accuracy of 89.5% evaluated by the ResNet50 architecture, while the NasNet-Mobile network attained 70% accuracy. With limited training datasets the results showed that the TL model offers considerable improvements in efficacy and efficiency.

Saputra et al. [22] demonstrated the DenseNet architecture to predict rice leaf diseases, employing three specific variants: DenseNet121, DenseNet169, and DenseNet201. The study's findings revealed that DenseNet121 achieved a precision rate of 91.67%, DenseNet169 attained a precision rate of 90%, and DenseNet201 reached a precision rate of 88.33%. Notably, the training time for these models was remarkably brief, requiring 24 seconds, highlighting the efficiency of the training process. Similarly, Hou et al. [23] developed a highly accurate and efficient image recognition model based on DenseNet architecture. Their performance evaluations showed that the model not only enhanced the efficiency of the model parameters but also maintained high levels of accuracy in image recognition tasks. Specifically, DenseNet200 achieved an accuracy of 97.2%, DenseNet100 reached 95.4%, and DenseNet50 obtained 92.3%. The findings indicate that augmenting the depth of the DenseNet model improves both the accuracy of pattern identification and the rate at which the model reaches convergence.

III. MATERIALS AND METHOD

The study followed a multi-phase approach outlined as follows: It utilized a dataset of 7,020 MRI brain images from Kaggle, categorized into four types of brain cancer. The data were pre-processed through grayscale conversion, noise reduction, and morphological operations. This dataset was divided into training, testing, and validation sets. Data augmentation was applied using the Image Data Generator. A CNN model was developed, with hyperparameter tuning performed on the DenseNet121, the InceptionResNetV2, the MobileNetV2, the NasNetMobile, and the ResNet50V2 architectures, with and without the TL. The model's efficacy was evaluated based on accuracy, precision, sensitivity, and specificity, providing a comparative analysis with previous research.

A. Brain Tumor Dataset

The dataset used in this study is from Kaggle and contains 7,020 MRI brain scans, categorized into four classes: meningioma, glioma, pituitary, and non-tumor. The data, which also includes contributions from Figshare, SARTAJ, and Br35H, were organized by disease type into separate directories, with each image having a resolution of 512 x 512 pixels and a 24-bit color depth in JPG format. Table 1 presents the distribution of images across these categories.

Each MRI image has a resolution of 512 x 512 pixels and is in JPG format with 24-bit colour depth. The images were organized by disease type, with each category stored in separate directories named after the corresponding condition. Fig. 1 illustrates the images included in the dataset.

TABLE I
QUANTITY OF BRAIN TUMOR IMAGES.

Classification of Brain	Total	Proportional
Glioma	1620	23,08%
Meningioma	1644	23,42%
Pituitary	1756	25,02%
No tumour	2000	28,48%
Total	7020	100%

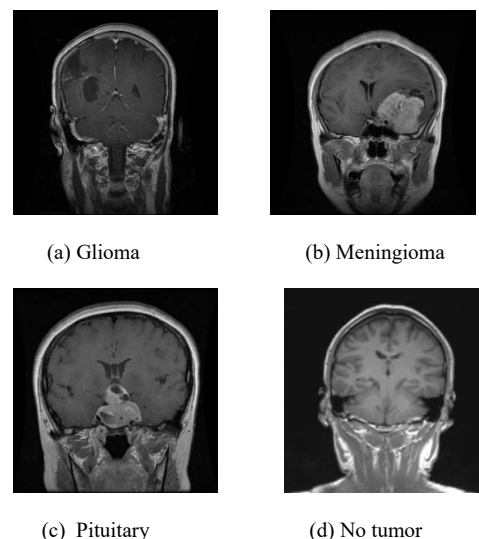


Fig. 1. Sample images representing each category.

B. Pre-processing

Pre-processing medical images is essential for improving their quality and interpretability, which affect diagnostic accuracy and treatment planning effectiveness [24]. This study harnessed several pre-processing techniques, including grayscale conversion, gaussian blur filter application, binary thresholding, erosion and dilation, contour extraction, image cropping, bone color map application, and image resizing.

The process began with grayscale conversion to simplify the image and focus on brightness intensity. The gaussian blur filter then reduces noise by generating new pixel values based on gaussian distribution [25]. The image was then converted to binary format using binary thresholding to distinguish the object from the background [26]. Erosion removes minor details and reduces image size, while dilation bridges pixel gaps, removes slight noise, fills gaps, and improves the object's shape [27]. Contouring, cropping, and bone color map techniques enhance bone structure visualization in MRI images. Finally, resizing optimizes the DL model's performance. Fig. 2 illustrates the outcomes after these pre-processing steps, showing noise reduction and object contour enhancement improvements.

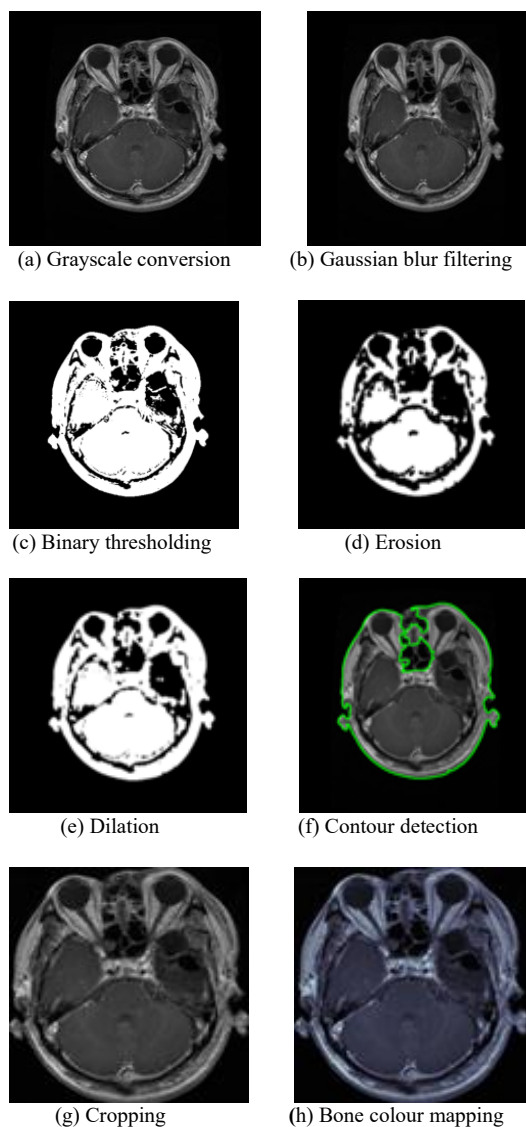


Fig. 2. Pre-processing steps.

C. Data Splitting

After the pre-processing process, the data were divided into two main components: modeling and testing. Based on prior research that suggests this ratio is ideal for performance, the data were separated with a 90% allocation for training and 10% for testing [28]. The training dataset was subdivided into two segments, maintaining a 90% training and 10% validation ratio. An illustrative representation of the dataset partitioning is shown in Fig. 3.

In ML, a 90/10 split is commonly used for training and testing, with a similar split for validation. This division is essential for developing effective models. Initially, the training data helps the model learn patterns related to brain tumors. The validation set then assesses the model's performance on unseen data, allowing hyperparameter tuning and preventing overfitting. Finally, the test data comprehensively evaluates the model's performance.

Proper dataset segmentation ensures the model can accurately classify brain tumor images, enhancing its robustness and relevance when faced with new data. Therefore, data partitioning is essential for training ML models in brain tumor research.

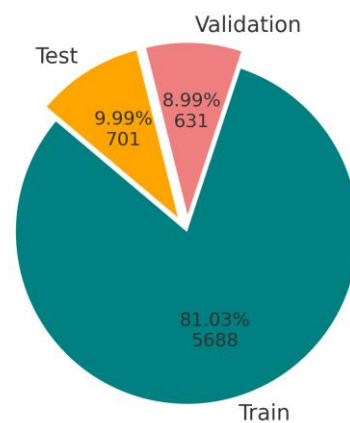


Fig. 3. Splitting dataset for MRI images.

D. Data Augmentation

Data augmentation involves generating synthetic data samples to expand the training dataset, especially for image datasets with limited samples [29]. This technique enhances dataset size, improves model performance and generalization, and reduces overfitting [30, 31]. This study employs data augmentation to enhance the model's ability to detect and classify brain tumor images. Various augmentation techniques are applied using the Image Data Generator from the Keras library, with specific parameters detailed in Table 2.

E. Classification Modeling

The study adopted a TL strategy during classification by using CNN architecture pre-trained on large datasets. This approach allows the model to leverage the existing knowledge from these datasets, thus reducing the time and resources needed to train a model from scratch. The research covers several pre-trained CNN models, detailed in sections 3.5.1 to 3.5.6, and includes a hyperparameter optimization process for fine-tuning the chosen model's parameters.

TABLE II
STAGES OF AUGMENTATION AND PARAMETER VALUES EMPLOYED

Technique	Parameter value	Description
Rescale	1/255	Normalizes the pixel values of an image from the range of 0 to 255 to a range of 0 to 1
Rotation Range	10	Applies a random rotation to the image within a predefined range of angles
Width Shift	0,002	Displaces the image horizontally within a defined fractional range relative to the image's width
Height Shift	0,002	Modifies the image's vertical position within a designated fractional range of the image's total height
Shear	12,5	Implement a sliding transformation on the image across a defined range of angles.
Horizontal Flip	True	Perform horizontal flipping of images in a randomized manner
Fill Mode	Nearest	Details the method for addressing gaps left by image transformation. The term 'nearest' refers to the technique of populating these empty pixels with the value of the closest neighbouring pixel.

1. Basic CNN

The CNN model is a sophisticated deep learning algorithm designed to extract relevant data features autonomously. CNNs consist of sequential layers: a convolutional layer that identifies features specifically edges and textures, a pooling layer that reduces resolution and parameters, and a fully connected layer that generates classification outputs. They are widely used in various fields, including computer vision, speech processing, and facial recognition. The effectiveness of CNNs in interpreting image data has led to their integration into applications like object detection, image segmentation, and autonomous driving. A sample illustration of a basic CNN architecture used in image classification can be found in [32] by Alzubaidi et al.

2. Mobile Network v2 (MobileNetV2)

The MobileNetV2 model is designed for optimal performance on resource-constrained devices, utilizing depth wise separable convolutions to enhance efficiency. Its streamlined structure enables the development of highly efficient models, ideal for mobile applications. Key advantages include improved memory efficiency during processing and broad compatibility, as it integrates seamlessly with all neural network frameworks. A representative example of the MobileNetV2 architecture is available in [33].

3. Densely Connected Convolutional Networks 121 (DenseNet121)

DenseNet121 is a variant of the DenseNet architecture proposed by Huang et al. [34] and consists of 121 layers. This model includes three principal classes: Bottleneck, Transition, and DenseNet. The Bottleneck class refers to the bottleneck layers within the DenseNet framework, while the Transition class refers to transition layers designed to reduce the spatial dimensions of the feature maps. The DenseNet class encompasses the complete DenseNet architecture [35]. Details of the DenseNet121 architecture can be found in [36].

4. Inception Residual Network v2 (InceptionResNetV2)

The InceptionResNet model integrates residual connections with the Inception architecture to enhance computational accuracy and efficiency. The InceptionResNetV2 variant, featuring 164 layers, is trained

on over a million images from the ImageNet database and can classify images into 1000 distinct object categories. Further information on the InceptionResNetV2 architecture illustration can be found in [37].

5. Residual Network 50 v2 (ResNet50V2)

The ResNet architecture emphasizes the residual function—representing the difference between the input and output—over direct input-output mapping. This approach addresses performance degradation typically seen in deep neural networks [38]. ResNet50V2, an advanced version of ResNet50, excels in image classification, object detection, and image segmentation tasks. An example of the ResNet50V2 architecture is available in [39].

6. Neural Architecture Search Network Mobile (NasNetMobile)

The NasNetMobile model is a cutting-edge neural network developed using the Neural Architecture Search (NAS) methodology, a leading technique in ML for discovering optimal network structures. NAS consists of three main elements: the search space, which explores different configurations like convolutional and pooling layers; the search strategy, which uses random search and reinforcement learning to pinpoint effective architectures; and synergy estimation, which assesses these architectures based on resource usage and time efficiency. A representation of the NasNetMobile model's architecture is provided in [40].

F. Hyperparameter Tuning

Hyperparameters, particularly the learning rate, batch size, and neuron count in neural networks, are required to be carefully tuned before training an ML model [41]. This study focused on optimizing two key hyperparameters: the dropout rate, which controls how often neurons are randomly excluded during training, and the dense unit, which determines the number of neurons in the fully connected layer. A grid search strategy was implemented to identify the best combination of these hyperparameters to enhance model performance and minimize overfitting. Initial parameters, namely the learning rate and batch size were set using standard values to ensure a robust foundation for training and evaluation. Table 3 provides a detailed summary of the hyperparameter configurations used in this process.

TABLE III
HYPERPARAMETER SETTINGS

	Hyperparameter	Value
No tuning Required	Learning Rate	0,0001
	Activation Function	ReLu, Softmax
	Epochs	15
	Loss Function	Categorical cross-entropy
	Batch Size	128
	Steps Per Epoch	49
Requires tuning	Validation Step	4
	Dropout Rate	{0,15; 0,2}
	Dense Units	{64; 128}

G. Fitting Models with the Best Hyperparameters

After completing hyperparameter tuning, the next step in model development is to fit the model using the identified optimal hyperparameter configuration. The results of this tuning process are presented in Table 4.

 TABLE IV
BEST PARAMETER COMBINATION

Model	Hyperparameters after tuning	
	Dropout rate	Dense units
Basic CNN	0,15	64
DenseNet121	0,20	128
InceptionResNetV2	0,20	128
MobileNetV2	0,15	128
NasNetMobile	0,20	128
ResNet50V2	0,15	64

H. Model Evaluation

Confusion matrix is a tool to evaluate the performance of classification models' applicable to binary and multi-class contexts. It visually compares predicted outcomes against actual values with dimensions of $N \times N$, where N represents the number of classes. The matrix comprises four key elements: True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). These elements are integral in calculating accuracy, precision, sensitivity, and specificity metrics. For multi-class classification, these metrics are averaged across all classes [42, 43].

- Accuracy: Measures the proportion of correct predictions (both positive and negative) relative to the total dataset. It is widely used due to its simplicity and is calculated as the average accuracy across all classes. The average accuracy for each class in multi-class classification is calculated using Equation (1).

$$acc_{avg} = \frac{\sum_{i=1}^l TP_i}{\sum_{i=1}^l TP_i + FP_i + FN_i + TN_i} \quad (1)$$

- Precision: Indicates the ratio of TP to the total positive predictions, aiming to reduce FP. In multi-class classification, precision is averaged across all classes and can be calculated using Equation (2).

$$p_{C_i} = \frac{TP_i}{TP_i + FP_i}$$

$$p_{avg} = \frac{\sum_{i=1}^l p_{C_i}}{l} \quad (2)$$

- Sensitivity: Also known as the true positive rate, it measures the proportion of correctly identified positives (TP) out of all actual positives, focusing on minimizing FN. The mean sensitivity value can be derived using Equation (3).

$$r_{C_i} = \frac{TP_i}{TP_i + FN_i}$$

$$r_{avg} = \frac{\sum_{i=1}^l r_{C_i}}{l} \quad (3)$$

- Specificity: Ratio of TN to the total number of genuinely negative instances, focusing on minimizing FP. The mean specificity is calculated using Equation (4).

$$s_{C_i} = \frac{TN_i}{TN_i + FP_i}$$

$$s_{avg} = \frac{\sum_{i=1}^l s_{C_i}}{l} \quad (4)$$

IV. RESULT AND ANALYSIS

A. Testing Without TL

The CNN model was initially evaluated without employing TL, utilizing a custom-designed architecture explicitly tailored for brain tumor classification. The model's performance was depicted through a loss graph, illustrating the progression of loss and accuracy throughout the training phase, as presented in Figs. 4 and 5.

As depicted in Figs. 4 and 5, the accuracy and loss graphs reveal initial signs of overfitting emerging between epochs 11 and 12. This situation was characterized by stagnation in the reduction of training loss and a simultaneous increase in validation loss, indicating a decline in the model's learning efficiency. Classification metrics, as previously outlined, were employed to assess the model's performance. The evaluation results were further illustrated through a confusion matrix and accuracy-loss visualizations. The confusion matrix for the CNN model without TL is presented in Fig. 6.

Fig. 6 highlights the performance variability of the model in classifying four tumor classes. For the no-tumor class, the model accurately classified 191 out of 200 images, resulting in a classification error rate of 4.5%. In the pituitary tumor class, the model correctly classified for 158 out of 175 images, corresponding to an error rate of 9.71%. However, the model encountered notable challenges with the glioma class, correctly identifying only 127 out of 162 images, yielding a higher error rate of 21.60%. The meningioma class posed the greatest difficulty, with the model successfully classifying only 108 out of 164

images, leading to a substantial error rate of 34.15%. Overall, the CNN model demonstrated an average accuracy of 83.30%, along with an average precision of 83.05%, an average sensitivity of 82.49%, and an average specificity of 92.71% (as summarized in Table 5).

The findings indicate that while the CNN model demonstrates a reasonable level of accuracy in detecting brain tumors, there is still potential for enhancement, particularly in sensitivity, which slightly lags behind accuracy and specificity. The application of TL is anticipated to further elevate model performance by harnessing knowledge from larger, more varied datasets, enabling the identification of more intricate patterns and improving classification, particularly for more challenging classes. These results also offer an initial assessment of a basic CNN model's ability to classify brain tumor images without TL. Such insights will provide a benchmark for comparing performance with models incorporating TL, helping to assess the potential for improved accuracy through the broader utilization of knowledge from extensive pre-existing datasets.



Fig. 4. CNN model accuracy graph.

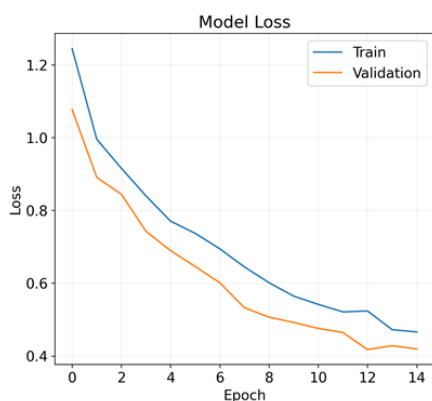


Fig. 5. CNN model loss graph.

TABLE V
CNN MODEL PERFORMANCE

Category	Precision	Sensitivity	Specificity
Glioma	76.50	78.40	92.76
Meningioma	84.37	65.85	96.27
No-Tumor	92.71	95.50	97.00
Pituitary	78.60	90.28	91.82

		Confusion Matrix			
True	Glioma	127	15	3	17
	Meningioma	24	108	9	23
	Notumor	3	3	191	3
	Pituitary	12	2	3	158
		Predicted			
		Glioma	Meningioma	Notumor	Pituitary

Fig. 6. CNN confusion matrix.

B. Testing with TL

Further evaluation was conducted by implementing TL with several pre-trained models, including DenseNet121, InceptionResNetV2, MobileNetV2, NasNetMobile, and ResNet50V2.

1. DenseNet121 Model

This study employed a TL methodology using the DenseNet121 pre-trained model. Initially, the model was loaded with weights derived from training on the ImageNet dataset, which were then fine-tuned for the classification of brain tumor diseases. The training process visualization, as depicted in Figs. 7 and 8, demonstrates a positive trend with increasing accuracy and decreasing losses during the early epochs. However, signs of overfitting became apparent between epochs 13 and 14, where the validation accuracy declined while the training losses continued to decrease, indicating that the model was starting to lose its convergence.

After conducting the training process, the model's performance is assessed using evaluation metrics for each class category, with the results visualized in the form of a confusion matrix, as illustrated in Fig. 9.

In the classification of the no-tumor category, the model demonstrated perfect performance by correctly identifying all 200 images without any errors. For the glioma category, the model accurately classified 156 out of 162 images, yielding an error rate of 3.70%. In the pituitary category, it achieved a success rate of 172 out of 175 images, resulting in an error rate of 1.71%. In the meningioma category, the model correctly identified 153 out of 164 images, corresponding to an error rate of 6.71%. Overall, DenseNet121 excels in accurately classifying the no-tumor class with no errors and exhibits a low error rate for glioma and pituitary classifications.

Furthermore, the performance results of the model, evaluated through precision, recall, and specificity metrics, are provided in Table 6, indicating consistently high performance across all categories.

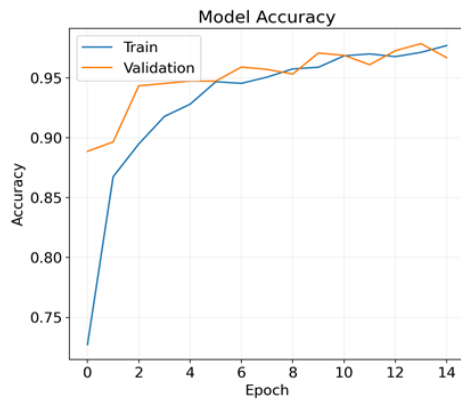


Fig. 7. DenseNet121 model accuracy graph.

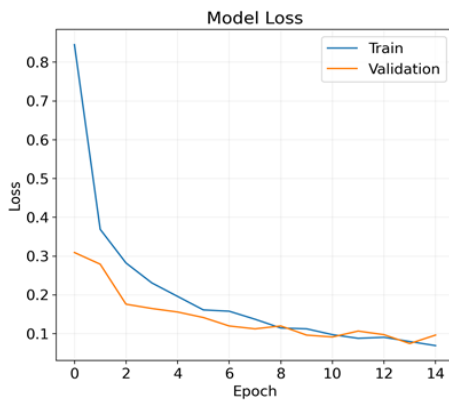


Fig. 8. DenseNet121 model loss graph.

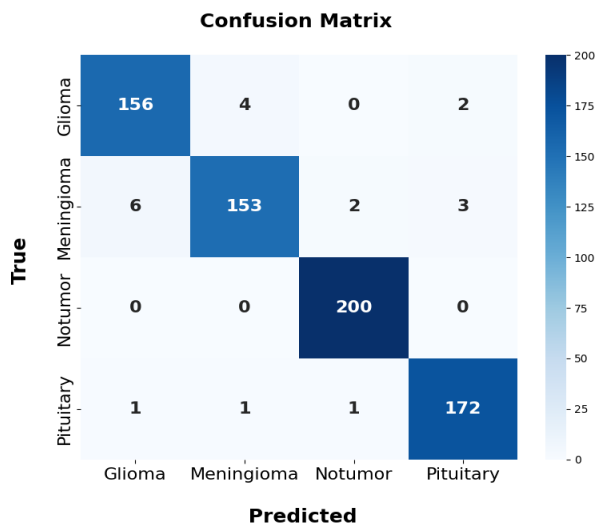


Fig. 9. DenseNet121 confusion matrix.

TABLE VI
DENSENET121 MODEL PERFORMANCE

Category	Precision	Sensitivity	Specificity
Glioma	95.70	96.30	98.70
Meningioma	96.80	93.30	99.10
No-Tumor	98.50	100	99.40
Pituitary	97.20	98.30	99.00

2. InceptionResNetV2 Model

The process of testing with TL on the InceptionResNetV2 model involves several stages, beginning with the initialization of the model using pre-trained weights derived from the large ImageNet dataset. Subsequently, the model is fine-tuned for the brain tumor classification task. By leveraging the pre-trained weights from ImageNet, the model is able to build upon previously acquired knowledge, enabling it to capture the unique features of brain tumors better. This approach aims to enhance classification accuracy by utilizing insights gained during the initial training phase.

The model's performance evaluation was conducted through an analysis of accuracy and loss graphs, as well as the confusion matrix. The visualizations depicting the accuracy and loss trends throughout the training process are presented in Figs. 10 and 11, which illustrate the evolution of these metrics across the epochs. From these figures, it is evident that the model converged around the 15th epoch, showing signs of halting further learning as its performance began to stabilize. That suggests that the training process had reached an optimal state, with no substantial improvements observed after that.

The confusion matrix shown in Fig. 12 outlines the model's classification performance across four categories: glioma, meningioma, no tumor, and pituitary. The model exhibits notable accuracy, especially in identifying glioma, no-tumor, and pituitary images. Specifically, in the glioma class, it accurately classified 157 out of 162 images, resulting in an error rate of approximately 3.09%. The performance in the meningioma class was marginally lower, with 21 out of 164 images misclassified, leading to an error rate of around 12.80%. In the no-tumor category, the model demonstrated excellent accuracy by correctly classifying 198 out of 200 images, achieving a minimal error rate of approximately 1%. Lastly, in the pituitary class, the model maintained strong performance, with an error rate of 1.14%, having correctly classified 173 out of 175 images.

A comprehensive analysis of the performance metrics, including precision, recall, and specificity, is provided in Table 7. For instance, in the glioma category, the model demonstrated a precision of 92.30%, a recall of 96.90%, and a specificity of 97.60%. These results underscore the model's accuracy and reliability in the classification of brain tumors.

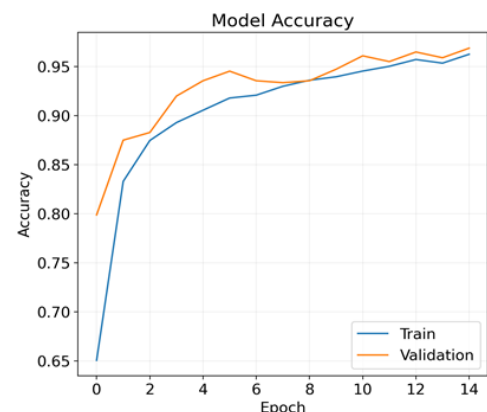


Fig. 10. InceptionResNetV2 model accuracy graph.

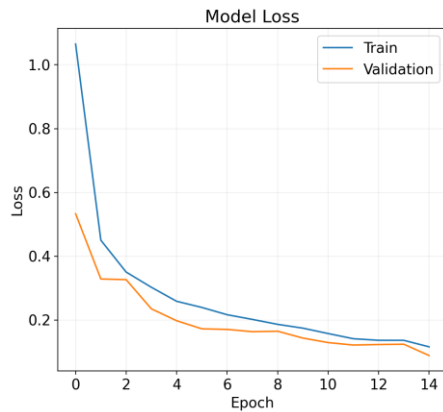


Fig. 11. InceptionResNetV2 model loss graph.

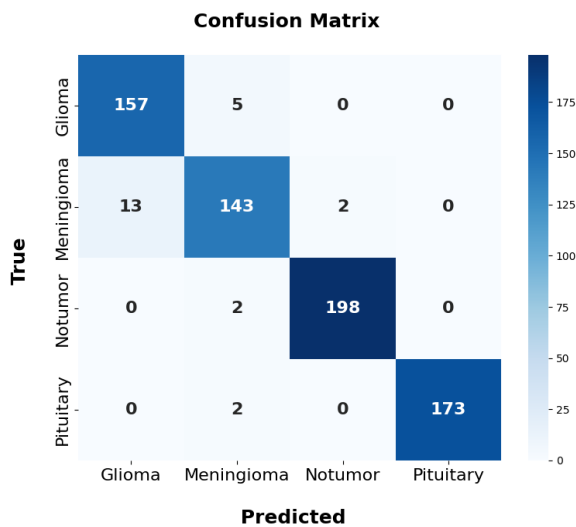


Fig. 12. InceptionResNetV2 confusion matrix.

TABLE VII
INCEPTIONRESNETV2 MODEL PERFORMANCE

Category	Precision	Sensitivity	Specificity
Glioma	92,30	96,90	97,60
Meningioma	94,10	87,20	98,30
No-Tumor	99,00	99,00	99,60
Pituitary	96,70	98,90	98,90

3. MobileNetV2 Model

The subsequent experiment employed TL with the MobileNetV2 model. This process began with initializing the pre-trained model and weights from the ImageNet dataset. The model is then fine-tuned to a specialized dataset focused on brain tumor diseases, with the objective of enhancing classification accuracy across four categories: glioma, meningioma, no-tumor, and pituitary.

Figs. 13 and 14 present the performance graphs of the model, which illustrate accuracy and loss throughout the training process. These graphs demonstrate that the model converges around the ninth epoch. The growing disparity between training and validation accuracy after the tenth epoch suggests a potential overfitting issue. Exploring regularization methods or implementing early stopping

can effectively mitigate this issue, ensuring the model achieves optimal performance while maintaining its generalization capabilities.

The classification performance is visualized through a confusion matrix, which illustrates the prediction accuracy for each class, as demonstrated in Fig. 15.

Based on the results presented in Fig. 15, the MobileNetV2 model exhibits varying classification performance across different tumor classes. In the glioma class, the model accurately identified 148 out of 162 images, yielding an error rate of approximately 8.64%. In contrast, for the meningioma class, the model demonstrated higher accuracy, correctly classifying 154 out of 164 images with an error rate of 6.10%. For the no-tumor class, the model showed exceptional performance, classifying 199 out of 200 images with only a single misclassification, resulting in a low error rate of 0.50%. Finally, in the pituitary class, the model achieved perfect classification, correctly identifying all 175 images without any errors, highlighting its flawless detection capability for pituitary tumors.

The performance is summarized in Table 8, which presents the precision, recall, and specificity metrics for each category. For instance, in the no-tumor class, both precision and specificity achieved a perfect score of 100%, while the pituitary class demonstrated a flawless recall of 100%, highlighting exceptional detection capabilities without overlooking any target images.

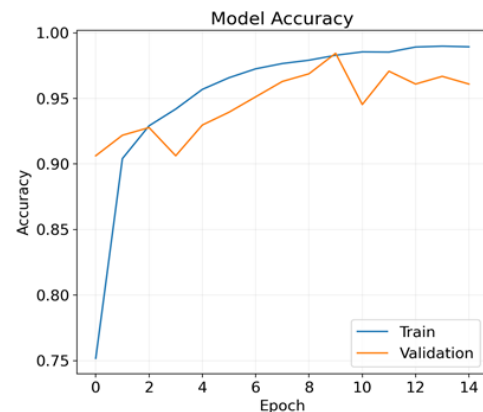


Fig. 13. MobileNetV2 model accuracy graph.

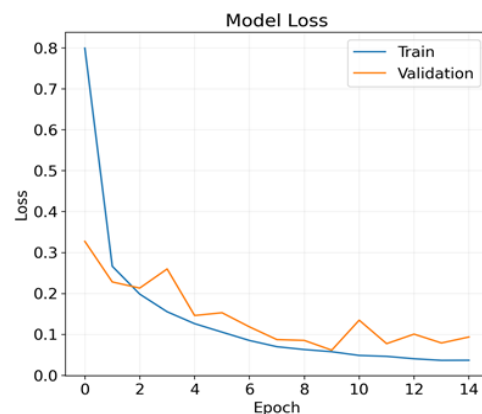


Fig. 14. MobileNetV2 model loss graph.

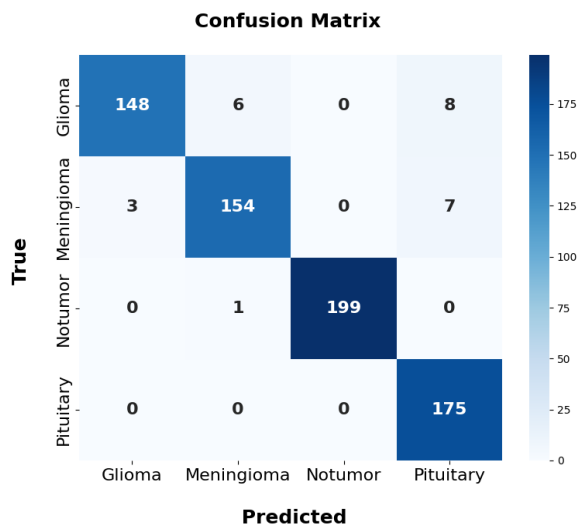


Fig. 15. MobileNetV2 confusion matrix.

TABLE VIII
MOBILENETV2 MODEL PERFORMANCE

Category	Precision	Sensitivity	Specificity
Glioma	98,00	91,40	99,40
Meningioma	95,60	93,90	98,70
No-Tumor	100	99,50	100
Pituitary	92,10	100	97,20

4. NasNetMobile Model

This study employed a TL approach utilizing the NasNetMobile model, which was initially trained on the ImageNet dataset. The pre-trained model is subsequently fine-tuned for the detection of brain tumors across several categories, including glioma, meningioma, pituitary tumors, and tumor-free classes. During the evaluation phase, the performance of the NasNetMobile classifier is assessed through accuracy and loss curves generated throughout the training process, as illustrated in Figs. 16 and 17.

Analyzing these two graphs reveal that the model exhibits unstable convergence on the validation data, with a significant reduction in loss values observed during each epoch, particularly between epochs 10 and 15. The findings indicate that although the model achieves high accuracy in certain classes, it faces considerable challenges in attaining optimal convergence.

The performance of this model was further illustrated using a confusion matrix, as shown in Fig. 18. Analysis of the confusion matrix reveals that the model demonstrates a high degree of accuracy in categorizing different types of tumors.

In the glioma class, the model accurately identified 146 out of 162 images, yielding an error rate of approximately 9.88%. For the meningioma class, the model recognized 158 out of 164 images, achieving a lower error rate of 3.66%. In the tumor-free class, the model demonstrated exceptional performance, correctly classifying 197 out of 200 images with an error rate of just 1.50%. For the pituitary class, the model correctly identified 173 out of 175 images, resulting in an error

rate of 1.14%. These findings indicate that the NasNetMobile model exhibits strong classification capabilities, particularly in the meningioma, tumor-free, and pituitary classes.

Table 9 presents the detailed performance of the model, including the precision, recall, and sensitivity values for each class. The model achieved exceptional precision in the tumor-free class (99.50%), outstanding recall in the pituitary class (98.90%), and the highest specificity in the tumor-free class (99.80%).

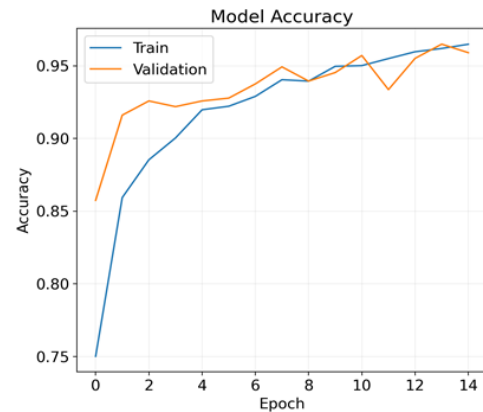


Fig. 16. NasNetMobile model accuracy graph.

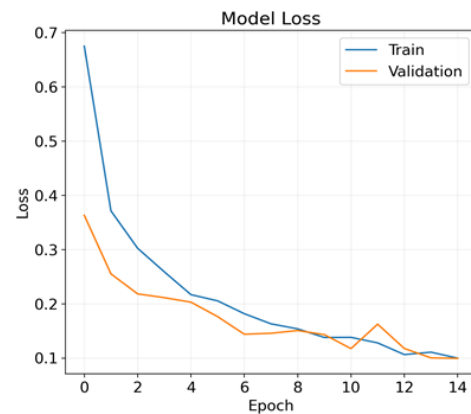


Fig. 17. NasNetMobile model loss graph.

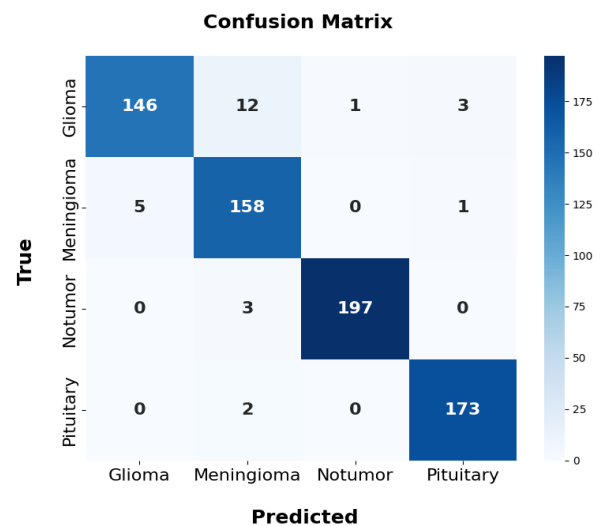


Fig. 18. NasNetMobile confusion matrix.

TABLE IX
NASNETMOBILE MODEL PERFORMANCE

Category	Precision	Sensitivity	Specificity
Glioma	96,70	90,10	99,10
Meningioma	90,30	95,80	96,80
No-Tumor	99,50	98,50	99,80
Pituitary	97,10	98,90	99,00

5. ResNet50V2 Model

The ResNet50V2 pre-trained model utilized weights obtained from training on the ImageNet dataset and was subsequently fine-tuned to classify brain tumor datasets. Graphical representations of accuracy and loss, along with the confusion matrix, were employed to evaluate the model's performance. Figs. 19 and 20 illustrate the accuracy and loss progression during training, indicating that the model converged at the 13th epoch, where loss reduction became negligible. These findings imply that the model achieved stability at this stage.

In the glioma class, the model accurately classified 153 out of 162 images, yielding an error rate of 6.17%. For the meningioma class, it successfully identified 162 out of 164 images, with an error rate of 1.22%. The model also demonstrated strong performance in the tumor-free class, classifying 198 out of 200 images, corresponding to an error rate of 1%. In the pituitary class, it correctly identified 173 out of 175 images, resulting in an error rate of 1.14%. These findings indicate that ResNet50v2 exhibited excellent accuracy across all classes, with exceptionally minimal classification errors in the meningioma, tumor-free, and pituitary classes.

Furthermore, the confusion matrix presented in Fig. 21 demonstrates the model's exceptional classification performance, particularly in accurately distinguishing the four primary categories: glioma, meningioma, tumor-free, and pituitary.

The performance metrics of the model were summarized in Table 10, which provides precision, recall (sensitivity), and specificity values for each class. For the glioma class, the model achieved a precision of 98.70%, a recall of 94.40%, and a specificity of 99.60%. The meningioma class demonstrated a precision of 93.10%, a recall of 98.70%, and a specificity of 97.80%. In the tumor-free class, the model reached a precision and specificity of 100%, alongside a recall of 99.00%. Lastly, for the pituitary class, the model reported a precision of 99.40%, a recall of 98.90%, and a specificity of 99.80%.

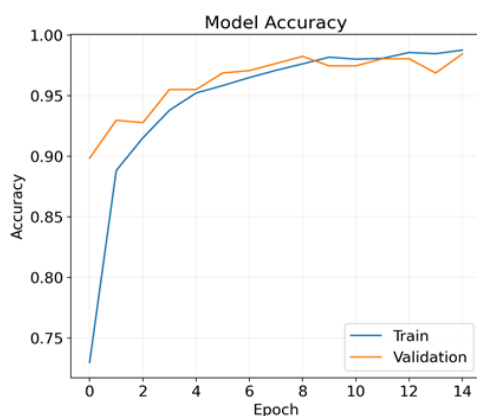


Fig. 19. ResNet50V2 model accuracy graph.

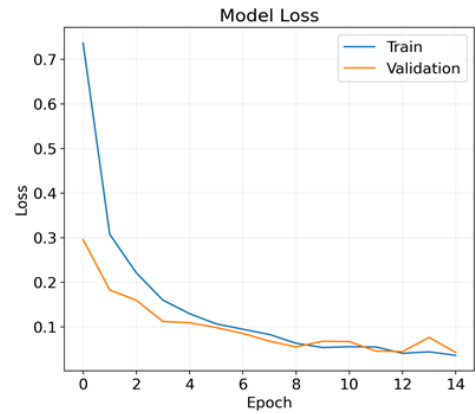


Fig. 20. ResNet50V2 model loss graph.

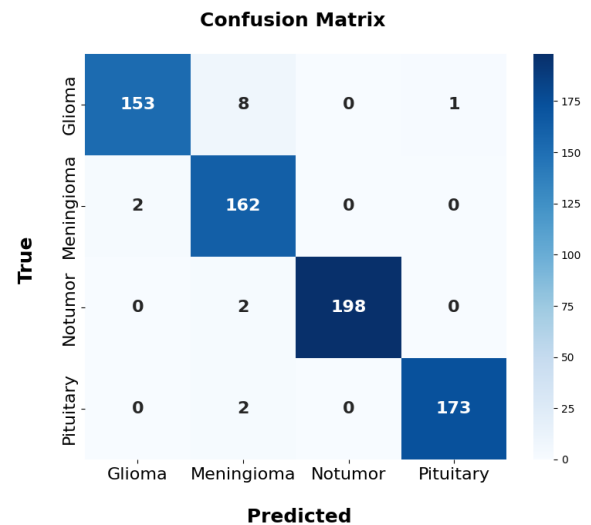


Fig. 21. ResNet50V2 confusion matrix.

TABLE X
RESNET50V2 MODEL PERFORMANCE

Category	Precision	Sensitivity	Specificity
Glioma	98,70	94,40	99,60
Meningioma	93,10	98,70	97,80
No-Tumor	100	99,00	100
Pituitary	99,40	98,90	99,80

C. Model Performance Comparative Analysis

The evaluation of the brain tumor classification results appeared in this section. The study thoroughly assessed the performance of each model and compares them to identify the most effective model for tumor classification. The comparative analysis used a range of critical metrics, including mean accuracy, mean precision, mean sensitivity, and mean specificity, with comprehensive findings detailed in Fig. 22. The analysis presented in Fig. 22 demonstrates that the ResNet50V2 model outperforms all other models in accurately classifying brain tumors, achieving an impressive mean accuracy of 97.70%. Its precision stands at 97.80%, with sensitivity and specificity rates of 97.70% and 99.30%, respectively, underscoring its exceptional performance in tumor classification. These metrics reflect the model's high reliability in detecting positive cases and its ability to differentiate between tumor types with minimal errors.

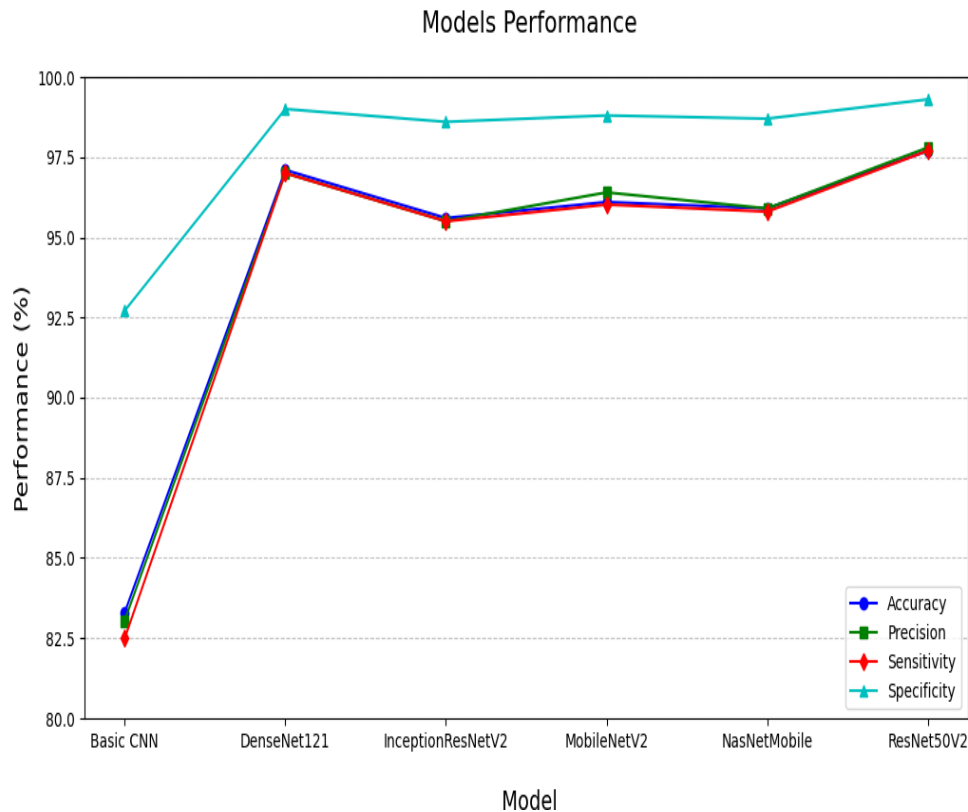


Fig. 22. Model performance comparison.

However, to fully validate the performance of the ResNet50V2 model, it is essential to conduct comprehensive evaluation of each model's loss graph. As illustrated in Fig. 19, such an analysis will offer deeper insights into the model's capability to recognize and categorize various brain tumor types accurately.

Fig. 20 illustrates that the ResNet50V2 model achieved optimal performance, effectively preventing both underfitting and overfitting. The loss curve indicates an apparent convergence between the training and validation losses, with the blue and orange lines converging at a particular point without substantial fluctuations leading up to it. Underfitting, which arises when a model is overly simplistic and unable to capture the underlying patterns in the data, was characterized by consistently high loss values in both the training and validation sets. Conversely, overfitting occurs when a model is excessively complex, capturing noise and irrelevant details in the training data, leading to excellent training performance but poor validation results. A noticeable gap between the loss reduction in the training data and the stagnation or increase of the loss in the validation data typically evidences this.

The ResNet50V2 model demonstrates a balanced trade-off between complexity and generalization, as evidenced by the intersection of the blue and orange curves, which highlights its capacity to both learn from training data and make accurate predictions on unseen data. Furthermore, this study contrasts the performance of a basic CNN model with that of a TL-based CNN model, evaluating the effectiveness of the optimal model in relation to previous research findings. A comprehensive performance comparison is presented in Fig. 23, offering a detailed assessment of the model's effectiveness in the context of prior studies.

Based on the data presented in Fig. 23, the proposed ResNet50V2 model demonstrates superior performance relative to previous TL-based CNN models, with the exception of the study by Narin et al. (2021). Narin et al. (2021) focused on binary classification, which involved only two classes, whereas this study addresses multi-class classification, encompassing more than two classes, thus presenting a greater level of technical complexity.

The ResNet50V2 model demonstrated comparable and, in some cases, superior performance across several metrics relative to previous studies. It achieved an accuracy of 97.70%, precision of 97.80%, sensitivity of 97.70%, and specificity of 99.30%. When compared to the ResNet50 model evaluated on Data-1 by Narin et al. (2021), the proposed model outperforms it in precision and specificity despite showing slightly lower accuracy and sensitivity. Moreover, in the studies by Narin et al. on Data-2 and Data-3, which report exceptionally high accuracy, the ResNet50V2 model remained highly competitive. When compared to MobileNetV3 by Hu et al. (2022) and other models like VGG-19 and various ResNet variants from the study by Triyadi et al. (2022), the ResNet50V2 model consistently outperforms all evaluation metrics. Furthermore, DenseNet models, as presented by Saputra et al. (2023) and Hou et al. (2024), demonstrate performance that is either comparable to or slightly lower than that of the proposed model. These findings collectively underscore the superior and consistent performance of the ResNet50V2 model, which is on par with or exceeds previous research, particularly in the domain of more complex multi-class classification.

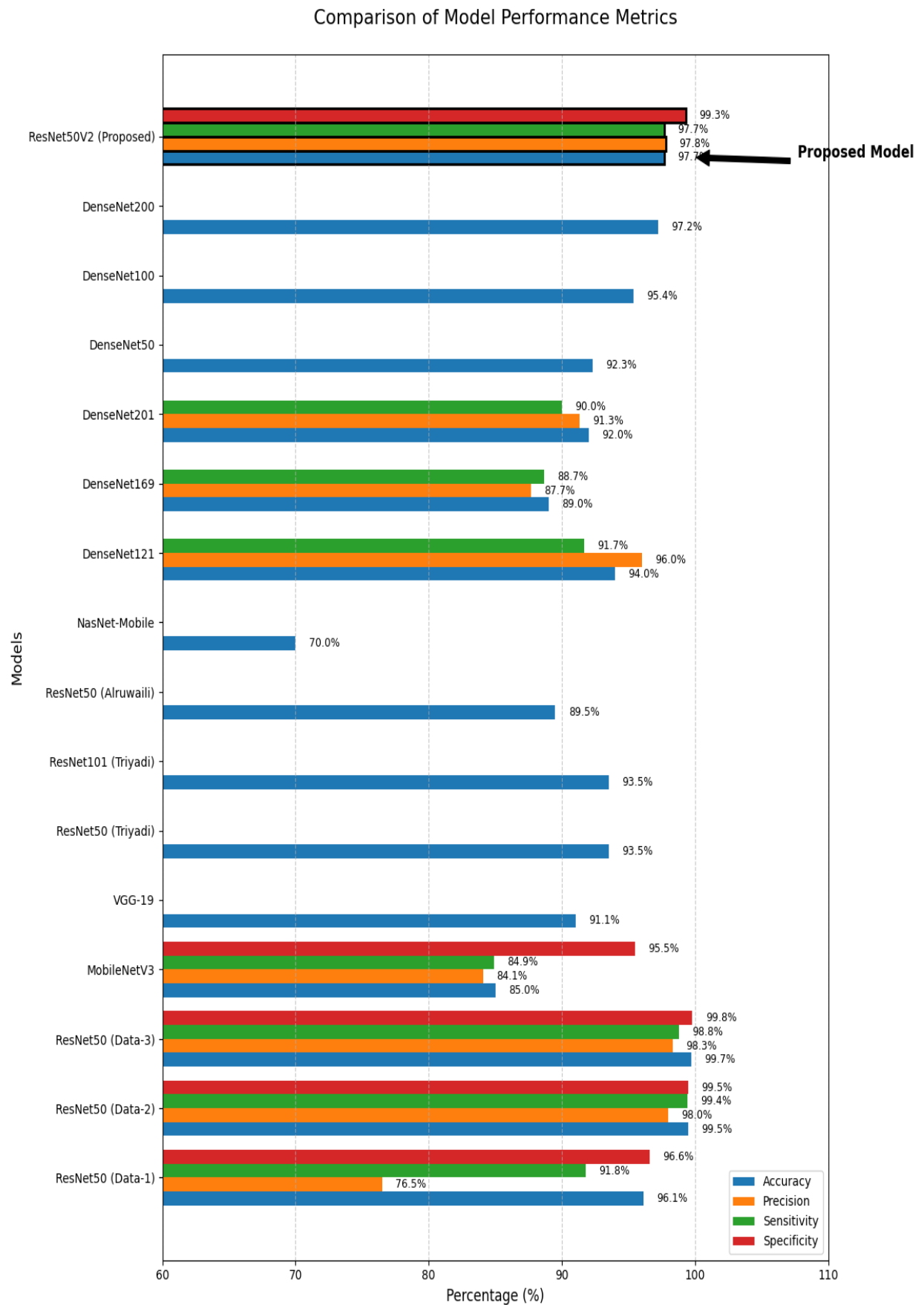


Fig. 23. Performance comparison of classification models against prior studies.

This model's performance not only surpass previous transfer learning-based approaches but also have a tangible impact on medical practice. With its high accuracy and precision, it can detect tumors at earlier stages and reduce diagnostic errors, such as false positives and negatives, thereby expediting patient care. Additionally, its ability to analyze thousands of MRI images in a short amount of time helps alleviate the workload of radiologists and reduces the risk of human error, enabling healthcare professionals to focus on more complex cases. The model's precise multi-class classification also facilitates more specific diagnoses, allowing treatments to be tailored to the tumor type, thereby enhancing therapeutic effectiveness.

The model also ensures the application in healthcare facilities with limited resources. Transfer learning techniques do not require large datasets for training, making it suitable for use in remote areas or developing countries with limited access to high-quality medical data. Its capacity to avoid overfitting and adapt to new data ensures reliable results across diverse patient conditions. Thus, ResNet50V2 not only enhances diagnostic efficiency in large hospitals but also broadens access to quality healthcare in underserved regions. Its implementation in clinical decision support systems could revolutionize brain tumor diagnosis, reduce treatment delays, and ultimately save more lives. However, this study has certain limitations that warrant consideration. The dataset used is confined to MRI images from a single source, potentially lacking the variability of data from different hospitals or MRI machines. Some models, such as MobileNetV2 and NasNetMobile, also exhibit signs of overfitting, meaning that while they perform well on training data, their accuracy diminishes on new data. Future research could expand by utilizing a more diverse dataset, as well as enhancing model performance through fine-tuning hyperparameters or improving preprocessing methods.

V. CONCLUSION

This study aims to develop and evaluate a CNN employing TL for the classification of brain tumor images. The critical stages of the approach included pre-processing, data augmentation, and hyperparameter optimization, all of which played essential roles in achieving accurate classification of brain MRI scans. The performance of the model was assessed using metrics specifically accuracy, precision, sensitivity, and specificity, with the results analyzed via a confusion matrix for each respective test class.

Among the models tested, ResNet50v2 achieved the highest performance, with an average accuracy of 97.7% and exceptional mean accuracy, sensitivity, and specificity of 97.80%, 97.70%, and 99.30%, respectively. The ResNet50v2 model's performance surpassed previous studies, confirming its effectiveness.

Future work should explore additional pre-trained models to further understand performance variations and address gaps in accuracy and loss chart depiction. Furthermore, further investigation is possible to be conducted into the model's real-world applicability, including the deployment in clinical settings with limited resources or varied equipment. Enhancements can be achieved by refining hyperparameters or improving pre-processing methods to

address challenges related to diverse data sources. This study contributes significantly to brain tumor image classification and aims to advance TL applications in this field, with the potential to drive innovations in early diagnosis and treatment optimization.

REFERENCES

- [1] T. A. Sadoon and M. H. Ali, "Deep learning model for glioma, meningioma, and pituitary classification," *Int. J. Adv. Appl. Sci.*, vol. 10, no. 1, pp. 88–98, 2021, doi: 10.11591/ijaas.v10.i1.pp88-98.
- [2] M. Yildirim, E. Cengil, Y. Eroglu, and A. Cinar, "Detection and classification of glioma, meningioma, pituitary tumor, and normal in brain magnetic resonance imaging using deep learning-based hybrid model," *Iran J. Comput. Sci.*, vol. 6, no. 4, pp. 455–464, 2023, doi: 10.1007/s42044-023-00139-8.
- [3] H. Gao and X. Jiang, "Progress on the diagnosis and evaluation of brain tumors," *Cancer Imaging*, vol. 13, no. 4, pp. 466–481, 2013, doi: 10.1102/1470-7330.2013.0039.
- [4] M. M. Badža and M. C. Barjaktarović, "Classification of brain tumors from mri images using a convolutional neural network," *Appl. Sci.*, vol. 10, no. 6, 2020, doi: 10.3390/app10061999.
- [5] S. J. Prashantha and H. N. Prakash, "Feature Level Fusion Framework for Brain Mr Image Classification Using Supervised Deep Learning and Hand Crafted Features," *Jordanian J. Comput. Inf. Technol.*, vol. 8, no. 4, pp. 336–344, 2022, doi: 10.5455/jcit.71-1655376900.
- [6] B. Babu Vimala, S. Srinivasan, S. K. Mathivanan, Mahalakshmi, P. Jayagopal, and G. T. Dalu, "Detection and classification of brain tumor using hybrid deep learning models," *Sci. Rep.*, vol. 13, no. 1, 2023, doi: 10.1038/s41598-023-50505-6.
- [7] A. T. Kabakus and P. Erdogmus, "An experimental comparison of the widely used pre-trained deep neural networks for image classification tasks towards revealing the promise of transfer-learning," *Concurr. Comput. Pract. Exp.*, vol. 34, no. 24, 2022, doi: 10.1002/cpe.7216.
- [8] M. S. Başarslan and F. Kayaalp, "MBi-GRUMCONV: A novel Multi Bi-GRU and Multi CNN-Based deep learning model for social media sentiment analysis," *J. Cloud Comput.*, vol. 12, no. 1, 2023, doi: 10.1186/s13677-022-00386-3.
- [9] M. Z. Khaliki and M. S. Başarslan, "Brain tumor detection from images and comparison with transfer learning methods and 3-layer CNN," *Sci. Rep.*, vol. 14, no. 1, 2024, doi: 10.1038/s41598-024-52823-9.
- [10] Z. Zhao, L. Alzubaidi, J. Zhang, Y. Duan, and Y. Gu, "A comparison review of transfer learning and self-supervised learning: Definitions, applications, advantages and limitations," *Expert Syst. Appl.*, vol. 242, 2024, doi: 10.1016/j.eswa.2023.122807.
- [11] A. Shrestha and A. Mahmood, "Review of deep learning algorithms and architectures," *IEEE Access*, vol. 7, pp. 53040–53065, 2019, doi: 10.1109/ACCESS.2019.2912200.
- [12] M. K. Bohmrah and H. Kaur, "Classification of Covid-19 patients using efficient fine-tuned deep learning DenseNet model," *Glob. Transitions Proc.*, vol. 2, no. 2, pp. 476–483, 2021, doi: 10.1016/j.gltp.2021.08.003.
- [13] D. Zhao, J. Wang, Y. Chu, Y. Zhang, Z. Yang, and H. Lin, "Improving biomedical word representation with locally linear embedding," *Neurocomputing*, vol. 447, pp. 172–182, 2021, doi: 10.1016/j.neucom.2021.02.071.
- [14] S. Wang, D. Liu, Z. Yang, C. Feng, and R. Yao, "A Convolutional Neural Network Image Classification Based on Extreme Learning Machine," *IAENG Int. J. Comput. Sci.*, vol. 48, no. 3, pp. 799–803, 2021.
- [15] Y. Tao, J. Shi, W. Guo, and J. Zheng, "Convolutional Neural Network Based Defect Recognition Model for Phased Array Ultrasonic Testing Images of Electrofusion Joints," *J. Press. Vessel Technol. Trans. ASME*, vol. 145, no. 2, 2023, doi: 10.1115/1.4056836.
- [16] Y. Shi, J. Xi, D. Hu, Z. Cai, and K. Xu, "RayMVSNet++: Learning Ray-Based 1D Implicit Fields for Accurate Multi-View Stereo," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 11, pp. 13666–13682, 2023, doi: 10.1109/TPAMI.2023.3296163.
- [17] V. S. Mahalle, N. M. Kandoi, and S. B. Patil, "Transfer Learning by Fine-Tuning Pre-trained Convolutional Neural Network Architectures for Image Recognition," 2024, pp. 273–287. doi:

- 10.1007/978-981-99-9179-2_21.
- [18] A. Narin, C. Kaya, and Z. Pamuk, "Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks," *Pattern Anal. Appl.*, vol. 24, no. 3, pp. 1207–1220, 2021, doi: 10.1007/s10044-021-00984-y.
- [19] J. Hu, Y. Qi, and J. Wang, "Skin Disease Classification Using Mobilenet-RseSK Network," in *Journal of Physics: Conference Series*, 2022, doi: 10.1088/1742-6596/2405/1/012017.
- [20] A. B. Triyadi, A. Bustamam, and P. Anki, "Deep Learning in Image Classification using VGG-19 and Residual Networks for Cataract Detection," in *Proceedings - 2022 2nd International Conference on Information Technology and Education, ICIT and E* 2022, 2022, pp. 293–297. doi: 10.1109/ICITE54466.2022.9759886.
- [21] M. Alruwaili and W. Gouda, "Automated Breast Cancer Detection Models Based on Transfer Learning," *Sensors*, vol. 22, no. 3, 2022, doi: 10.3390/s22030876.
- [22] A. D. Saputra, D. Hindarto, and H. Santoso, "Disease Classification on Rice Leaves using DenseNet121, DenseNet169, DenseNet201," *Sinkron*, vol. 8, no. 1, pp. 48–55, 2023, doi: 10.33395/sinkron.v8i1.11906.
- [23] Y. Hou, Z. Wu, X. Cai, and T. Zhu, "The application of improved densenet algorithm in accurate image recognition," *Sci. Rep.*, vol. 14, no. 1, p. 8645, 2024, doi: 10.1038/s41598-024-58421-z.
- [24] N. T. Singh, C. Kaur, A. Chaudhary, and S. Goyal, "Preprocessing of Medical Images using Deep Learning: A Comprehensive Review," in *Proceedings of the 2023 2nd International Conference on Augmented Intelligence and Sustainable Systems, ICAISS 2023*, Aug. 2023, pp. 521–527. doi: 10.1109/ICAISS58487.2023.10250462.
- [25] N. Zin Oo, "The Improvement of 1D Gaussian Blur Filter using AVX and OpenMP," in *International Conference on Control, Automation and Systems*, 2022, pp. 1493–1496. doi: 10.23919/ICCAS55662.2022.10003739.
- [26] K. R. Raheem and H. A. Shabat, "An Otsu thresholding for images based on a nature-inspired optimization algorithm," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 31, no. 2, pp. 933–944, 2023, doi: 10.11591/ijeecs.v31.i2.pp933-944.
- [27] S. Y. Ma, A. Khalil, H. Hajjdiab, and H. Eleuch, "Quantum dilation and erosion," *Appl. Sci.*, vol. 10, no. 11, 2020, doi: 10.3390/app10114040.
- [28] D. H. N. Aini, D. Kurniasari, A. Nuryaman, and M. Usman, "Implementation of Artificial Neural Network With Backpropagation Algorithm for Rating Classification on Sales of Blackmores in Tokopedia," *J. Tek. Inform.*, vol. 4, no. 2, pp. 365–372, 2023, doi: 10.52436/1.jutif.2023.4.2.539.
- [29] Z. Yang, R. O. Sinnott, J. Bailey, and Q. Ke, "A survey of automated data augmentation algorithms for deep learning-based image classification tasks," *Knowl. Inf. Syst.*, vol. 65, no. 7, pp. 2805–2861, 2023, doi: 10.1007/s10115-023-01853-2.
- [30] N. E. Khalifa, M. Loey, and S. Mirjalili, "A comprehensive survey of recent trends in deep learning for digital images augmentation," *Artif. Intell. Rev.*, vol. 55, no. 3, pp. 2351–2377, 2022, doi: 10.1007/s10462-021-10066-4.
- [31] W. Zeng, "Image data augmentation techniques based on deep learning: A survey," *Math. Biosci. Eng.*, vol. 21, no. 6, pp. 6190–6224, Jun. 2024, doi: 10.3934/mbe.2024272.
- [32] L. Alzubaidi et al., "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," *J. Big Data*, vol. 8, no. 1, 2021, doi: 10.1186/s40537-021-00444-8.
- [33] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2018, pp. 4510–4520. doi: 10.1109/CVPR.2018.00474.
- [34] C. Huang, Y. Li, C. C. Loy, and X. Tang, "Learning deep representation for imbalanced classification," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016, pp. 5375–5384. doi: 10.1109/CVPR.2016.580.
- [35] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017, pp. 2261–2269. doi: 10.1109/CVPR.2017.243.
- [36] Q. Ji, J. Huang, W. He, and Y. Sun, "Optimized deep convolutional neural networks for identification of macular diseases from optical coherence tomography images," *Algorithms*, vol. 12, no. 3, 2019, doi: 10.3390/a12030051.
- [37] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-ResNet and the impact of residual connections on learning," in *31st AAAI Conference on Artificial Intelligence, AAAI 2017*, 2017, pp. 4278–4284. doi: 10.1609/aaai.v31i1.11231.
- [38] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.
- [39] S. P. Nikhat Raza Khan, "A Weighted-Average-Ensembling-Based Hybrid CNN Model for Improved Covid-19 Detection," *Tuijin Jishu/Journal Propuls. Technol.*, vol. 44, no. 4, pp. 6016–6033, 2023, doi: 10.52783/tjjpt.v44.i4.2033.
- [40] F. A. Shah et al., "A cascaded design of best features selection for fruit diseases recognition," *Comput. Mater. Contin.*, vol. 70, no. 1, pp. 1491–1507, 2021, doi: 10.32604/cmc.2022.019490.
- [41] E. Elgeldawi, A. Sayed, A. R. Galal, and A. M. Zaki, "Hyperparameter tuning for machine learning algorithms used for arabic sentiment analysis," *Informatics*, vol. 8, no. 4, 2021, doi: 10.3390/informatics8040079.
- [42] M. Hossin and M. . Sulaiman, "A Review on Evaluation Metrics for Data Classification Evaluations," *Int. J. Data Min. Knowl. Manag. Process*, vol. 5, no. 2, pp. 01–11, 2015, doi: 10.5121/ijdkp.2015.5201.
- [43] F. J. Ramirez-Arias et al., "Evaluation of Machine Learning Algorithms for Classification of EEG Signals," *Technologies*, vol. 10, no. 4, 2022, doi: 10.3390/technologies10040079.