IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Traitement du Signal
Vol. 42, No. 6, December, 2025, pp. 3423-3433
Journal homepage: http://iieta.org/journals/ts

Deep Ensemble Model Integrating MobileNet and Swin Transformer for Robust Brain Tumor Classification

Ramasamy Shunmugaraj Karthic[1], Karupanan Raju Aravind Britto[1*], Ramachandran Ragumadhavan[1], Rayappan Vimala[2]

[1] Department of Electronics and Communication Engineering, PSNA College of Engineering and Technology, Dindigul 624622, India
[2] Department of Electrical and Electronics Engineering, PSNA College of Engineering and Technology, Dindigul 624622, India

Corresponding Author Email: krbritto1975@gmail.com

https://doi.org/10.18280/ts.420629

Received: 10 July 2025
Revised: 2 September 2025
Accepted: 19 November 2025
Available online: 31 December 2025

Keywords:
brain tumor, deep learning model, MRI data sets, VGG16, ResNet, MobileNet

ABSTRACT

As the demand for accurate early detection of brain tumors continues to grow, automated deep learning models have become increasingly important in medical image analysis. This paper presents an effective ensemble approach that integrates MobileNet, known for its compact architecture and rapid feature extraction, with Swin Transformer, a structured vision transformer capable of capturing global contextual information and temporal dependencies. The hybrid model is designed to leverage the strengths of both networks, delivering high accuracy with minimal computational cost. The proposed model was trained and evaluated on standard brain tumor MRI datasets and achieved an outstanding accuracy of 99.65%, surpassing other established models such as VGG16, ResNet variants, and standalone transformer-based architectures. The experimental results demonstrate that the ensemble model significantly enhances classification performance and exhibits strong generalization capability across different tumor types. Previous comparative studies using MobileNet, transformer-based models, and ensemble techniques on brain tumor MRI datasets have reported accuracies below 99%, highlighting the superior performance and efficiency of the proposed method in medical imaging analysis.

1. INTRODUCTION

With the rapid growth in medical imaging technology, brain tumor segmentation has significantly progressed, becoming a crucial task in the medical imaging process. Currently, researchers predict two kinds of brain tumors: primary and metastatic tumors, relying on the origin of brain tumor cells that develop directly in the brain or spread to it from other organs of the body [1]. Primary brain tumors are generated in the brain tissue, whereas metastatic brain tumors are tumors spread from different locations. Regarding histological nature, it can categorize as glioma (G), meningioma (M), and pituitary tumor (P) [2]. From these, the glioma serves as the most typically affected type of brain tumor, known for its higher mortality rate and aggressive nature. Hence, physicians must diagnose promptly and provide precise and appropriate therapy by analyzing brain images to manage patients effectively and optimize survival rates. The most widely used medical imaging technique is magnetic resonance imaging (MRI) [3]. MRI can offer four varied identical modalities and when compiled, they can develop in-depth and overall insights about the structure and operation of the brain. These four image forms are unique and these can offer extensive and comprehensive data regarding the brain anatomy and malignancies. With these detailed insights, physicians can provide a precise diagnosis and generate an effective treatment strategy [4].

MRI images play a crucial role in predicting brain tumor at an early stage. But, in real-time clinical settings, the process is mainly based on knowledge of the radiologists to predict the tumor type and site mapping manually. This approach consumes more physical and material sources and may also involve a risk of errors and oversights due to the inherent uncertainty in expert opinions. Hence, it urges having computing techniques to assist experts in accurately classifying and segmenting brain tumor. This automated technique aids experts in formulating specific treatment plans while significantly reducing their workload.

Manual segmentation of brain tumor images, which is having unsymmetrical shapes and intricate boundaries consumes more time and a risk of making errors. Hence, the researchers recently generate automated segmentation methods with achieving higher accuracy. In conventional methods, the segmentation method is performed based on thresholds, boundaries, and regions [5], however, it achieves minimal accuracy. Nowadays, the rapid advancement of artificial intelligence has significantly contributed to its integration across various domains [6-8]. An integration of autonomous computing technique with brain tumor prediction enhances the prognostic efficiency while significantly reducing the expert workload. Conventional brain tumor segmentation method prone to error with minimal accuracy.

On contrary, the techniques based on deep learning automatically extracting features from MRI images, enhanced the prediction rate of brain tumor and segmentation accuracies of their tissues. But, developing highly accurate segmentation algorithms remains a critical concern in enhancing the precision and robustness of brain tumor diagnosis.

In 2015, Long et al. [8] introduced the Fully Convolutional Network (FCN), which represented a major evolution from the conventional Convolutional Neural Network (CNN) architecture. Unlike traditional CNNs, FCNs eliminate fully connected layers in favor of convolutional operations throughout the network. They also apply up-sampling techniques to generate segmented outputs that closely mirror the original input dimensions. This design substantially enhances segmentation accuracy while minimizing computational demands. As a result, FCNs have established themselves as a foundational model in the field of deep learning-based semantic segmentation, inspiring extensive research and development. For instance, Shen et al. [9] proposed a model based on FCN principles that utilized symmetric differential images and incorporated three up-sampling structures to extract features effectively. Building on the FCN framework.

2. RELATED WORK

Medical image analysis has shown a tremendous growth in recent years, particularly in brain tumor detection from MRI data. Deep learning models have demonstrated remarkable efficacy in these tasks due to their superior ability to extract relevant features. A surge in research continues to support their potential and accuracy in detecting and segmenting brain tumors. To achieve improved segmentation results, it's essential to leverage multiple MRI modalities. For instance, Zhou [10] developed a U-Net variant that handles multimodal MRI data, integrating learning techniques that separate mixed representations and focus on tumor-relevant regions through a contrastive framework. These strategies help isolate individual tumor characteristics and enhance the learning process. The model, tested on BraTS 2018 and 2019 datasets, achieved performance exceeding many current approaches. In a related work, Zhou [11] proposed a segmentation system capable of working even when certain MRI sequences are absent. This method includes reconstructing missing modalities and learning hidden relationships across different inputs. The suggested model showed robust segmentation results when evaluated on the BraTS 2018 dataset. Likewise, Zhu et al. [12] introduced a 3D segmentation model structured around three integrated modules: (1) border shape correction (BSC), (2) spatial information enhancement (SIE), and (3) modality information extraction (MIE). The model was benchmarked on the BraTS datasets from 2017 to 2019, reaching average Dice scores of 0.821, 0.858, and 0.853, respectively. Ranjbarzadeh et al. [13] proposed a segmentation framework built on convolutional neural networks, utilizing four types of MRI sequences (T1, T2, T1ce, FLAIR). In the early stage, potential tumor regions are estimated. Feature extraction is carried out using a bio-inspired optimization technique (an improved chimp-based algorithm), and classification is done through a supervised learning method, widely known for its effectiveness in small-scale datasets. These features are then passed into the CNN for final segmentation. The model's hyperparameters were optimized using the same algorithm. On the BraTS 2018 dataset, it delivered impressive precision (97.41%), recall (95.78%), and Dice score (97.04%). To strike a balance between speed and accuracy, Montaha et al. [14] introduced a compact 2D U-Net variant that analyzes 2D slices from 3D MRI volumes. This approach retains spatial coherence by using skip connections and preprocessing techniques such as image rescaling and normalization. Trained on the BraTS2020 dataset, it reached a Dice score of 93.1% and accuracy of 99.41%. Feng et al. [15] presented MLU-Net, a compact model that uses frequency-based representations and dense multilayer learning techniques to address feature degradation often observed during segmentation. This model, designed for efficient computation, reduced the number of learnable parameters and processing load by significant margins compared to conventional U-Net models, while continuing to enhance segmentation performance. Specifically, the Dice and overlap metrics were improved by 3.37% and 3.30%.

Moreover, Zhang et al. [16] introduced ETUNet, which integrates transformer layers into the U-Net architecture to extract broader feature dependencies and improve feature representations in brain tumor segmentation. On BraTS 2018 and 2020 datasets, it achieved average DSC scores of 0.854 and 0.862 and reported Hausdorff distances (HD95) of 6.688 and 5.455, showing notable improvements. To overcome the challenge of limited labeled medical images, Hammer Håversen et al. [17] introduced QT-UNet, a self-supervised model that learns without the need for large annotated datasets. The approach incorporates a querying mechanism that directs the model's discovery of significant patterns in unlabeled data. On BraTS 2021, it achieved a Dice score of 88.61 and ahaus Dorff Distance of 4.85 mm. Several researchers have further addressed challenges like indistinct tumor borders and overlapping intensities.

Hussain and Shouno [18] introduced a parallel-deep learning architecture that combines multiple convolution layers with advanced training and preprocessing to improve accuracy. Cui et al. developed a cascaded architecture that uses a localization network for tumor detection and a classification network for sub-region analysis. Additionally, the use of attention layers and residual blocks helped refine the segmentation results. Verma et al. [19] proposed RR-U-Net, which incorporated skip-connected residual blocks into the base U-Net, boosting its ability to recognize fine-grained tumor features. Gayathri et al. [20] similarly enhanced U-Net with residual layers for brain tissue segmentation using FLAIR sequences, though this increased model complexity. Cinar et al. merged DenseNet121 into U-Net, improving feature reuse and segmentation performance. However, resolution limitations hindered the model's ability to capture subtle tumor structures.

3. PROPOSED METHODOLOGY

Figure 1 presents the proposed architecture, designed as a complete system for brain tumor detection from MRI scans. The approach integrates traditional image processing techniques with deep learning models to achieve accurate classification and diagnosis. The system is divided into two main modules: the Image Processing Module and the Deep Learning Module. In the Image Processing Module, MRI images are first collected and passed through a series of preprocessing steps to optimize feature extraction. The process

begins with Gaussian denoising to remove noise while preserving edges, followed by skull stripping to eliminate non-cerebral tissue. The images are then normalized to maintain consistent intensity values across the dataset. Otsu's thresholding converts grayscale images into binary format, supporting segmentation. This step is further refined with region growing, watershed, K-means clustering, and Canny edge detection, all of which help emphasize tumor boundaries and generate clean inputs for the deep learning stage. The Deep Learning Module begins with dataset preprocessing, including cleaning, resizing, and augmentation to improve model performance. The dataset is then split into training and testing sets to ensure fair evaluation. Several deep learning models—DenseNet, MobileNet, and Swin Transformer—are trained individually, after which an ensemble combines their strengths to improve prediction accuracy and robustness. Model performance is evaluated using Accuracy, Precision, Recall, and F-Measure, ensuring reliability across diverse MRI data. The integration of conventional preprocessing with modern deep learning provides a robust pipeline for automated brain tumor detection. Figure 2 illustrates the sequential preprocessing stages, showing (a) the original MRI, (b) Gaussian denoised, (c) skull-stripped, (d) normalized, (e) Otsu's thresholded, (f) region grown, (g) watershed segmented, (h) K-means clustered, and (i) Canny edge-detected outputs. These steps progressively refine the input, enabling accurate segmentation and effective feature learning. DenseNet, MobileNet, and Swin Transformer are then employed as core classifiers. Each learns distinct feature hierarchies: DenseNet for dense connectivity, MobileNet for lightweight convolutional efficiency, and Swin Transformer for global context via attention. Their predictions are integrated through an ensemble strategy, delivering superior accuracy and generalization compared to individual models.

Figure 1 shows the proposed architecture. This method is an overall system for brain tumor detection from MRI scans, incorporating modern image processing techniques with deep learning algorithms to achieve precise classification and diagnosis. A comprehensive system is separated into two major sections: the Image Processing Module and the Deep Learning Module. The image processing module starts with gathering MRI images and is fed into a sequence of preprocessing procedures to attain an effective feature extraction with optimized quality. The workflow begins with the original image, and then unnecessary noise is eradicated by employing Gaussian denoising while protecting edge insights. Next, the skull stripping process eliminates non-cerebral tissue from the brain region. Subsequently, the images are generalized to retain persistent intensity values among the dataset. Next, Otsu's Thresholding technique is employed to modify grayscale images into binary format, facilitating the segmentation process. The segmentation process is further improved by utilizing region growing, watershed algorithm, K-means clustering, and Canny edge detection. These methods provide effective images for deep-learning algorithms and appropriately highlight the tumor outliners. The processed image is then subjected to the Deep Learning Module. It starts with data preprocessing, which includes cleaning, resizing, and augmenting the dataset to enhance the model's performance.

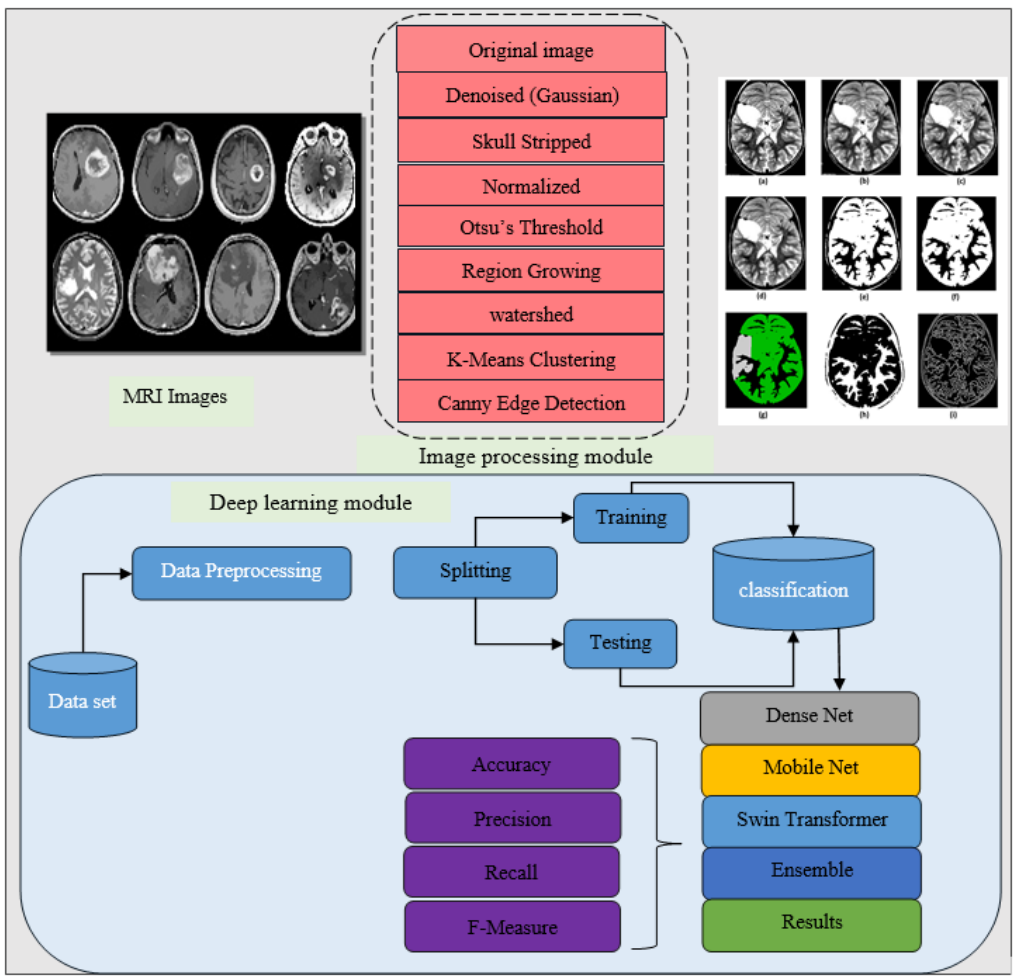

Figure 1. Proposed architecture

Subsequently, the dataset is split into training and testing sets to assure that the model performs effectively when applied to unknown data. The model efficiently learns trends and features using the preprocessed insights in the training stage, whereas the model's efficiency is evaluated in the testing phase. Subsequently, the classification process starts with employing multiple deep learning models like DenseNet, MobileNet, and Swin Transformer. These models are separately trained and evaluated, after which an ensemble approach combines their potentials, thereby improving prediction accuracy and improving overall robustness. The final result was evaluated in terms of the following metrics Accuracy, Precision, Recall, and F-Measure. These characteristics enable the model's overall performance, demonstrating its effectiveness in accurately detecting brain tumors and ensuring its reliability over a range of MRI datasets. A notable development in automated brain tumor detection is offered by the combination of traditional image processing with modern deep learning techniques, associated with an ensemble model.

The above image series highlights the sequential image processing steps employed in MRI scans, which is crucial for accurate tumor segmentation and interpretation. The Figure 2 shows (a) the Original Image, a fresh MRI scan image that may comprise unwanted noise and unrelated structural features. In (b) Gaussian filter is employed to eradicate those noises while protecting crucial structural edges, enhancing image quality, and producing a clear, denoised image. Following that, the image processed by (c) Skull Stripping technique to eliminate the non-brain regions such as the skull and scalp from the brain region reduces false positives in further analysis. The next image shows (d) Normalized image ensures that the pixel intensity values are normalized to a constant limit, ensuring stable input for deep learning models. Next, (e) Otsu's Thresholding is applied to transform the grayscale image into a binary format automatically by evaluating the maximum threshold value and providing a distinction between basic components (e.g., brain tissues or tumor regions) and the background. Subsequently, (f) Region Growing, a segmentation technique that extends a selected area depending upon the unique intensity values, enables accurate detection and delineation of tumor boundaries.

The segmentation is further enhanced by employing the (g) Watershed algorithm, which assumes the image as a topographic surface and segments the region based on gradient intensity, enabling more accurate segmentation of intricate patterns. Next, (h) K-Means Clustering is employed to split the image into multiple clusters based on pixel intensity values, providing the distinguished image of tumor regions from the non-tumor sites. Ultimately, (i) Canny Edge Detection is applied to emphasize sharp intensity transitions, effectively outlining the contours and edges of brain structures and potential tumor regions with high precision. This effective preprocessing procedure offers a clear image for the deep learning models to improve the prognostic accuracy and model performance.

Figure 3 depicts the DenseNet–based deep learning architecture specially constructed for brain tumor detection using MRI images. Initially, the brain's original MRI image is given as input for analysis. This raw image is first processed by the integration of convolutional and pooling layers. The convolutional layers are utilized to extract the crucial basic features such as edges and textures while pooling layers mitigate the spatial dimension of the feature maps, which enhances computing efficacy and reduces overfitting.

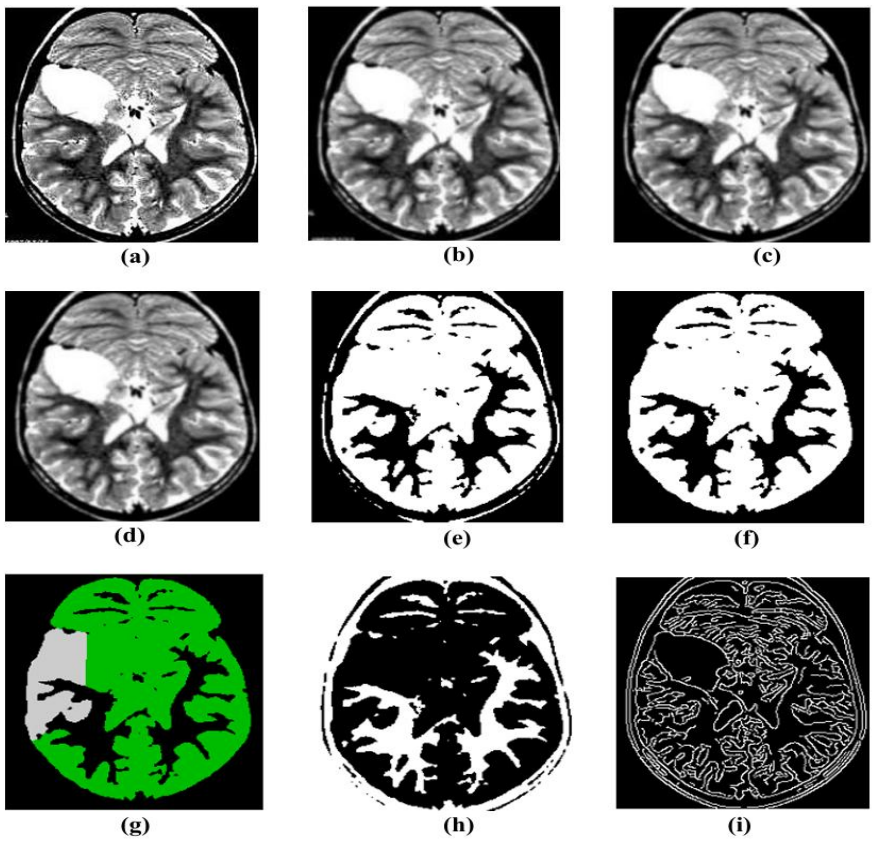

Figure 2. (a) Original Image (b) Denoised (Gaussian) (c) Skull Stripped (d) Normalized (e) Otsu's Threshold (f) Region Growing (g) watershed (h) K-Means Clustering (i) Canny Edge Detection

After this data processing stage, the input is passed through dense blocks, which consist of a series of densely connected convolutional layers. Every layer in the dense block, shown as receiving input X2X_2X2 is connected in a feed-forward manner, enabling each layer receive input from the preceding layer. This peculiar architecture optimizes training efficiency and accuracy by providing superior feature reuse and ensures the free movement of gradients during backpropagation. Following the completion of the initial dense block, the result is transferred to a transition layer. A critical transition layer mitigates the feature map's count to reduce the network and implement spatial down-sampling using convolution and pooling functions. This process mitigates the intricacies of the model and acts as a generalization method.

This process is carried out by the second and third dense blocks, each of which improves and builds upon the features extracted from the previous layers. The third block input denoted as X3X_3X3, knows more abstract and intricate features crucial for distinguishing between healthy tissue and potential tumors. A transition layer is added following these blocks to preserve the network's depth and dimensionality. There are several inputs to the final blocks, especially integrating specific features from the earlier blocks (X1X_1X1 and X3X_3X3), to generate overall feature representation. This integration of features from various levels of the network ensures that both low-level data and high-level abstract patterns are effectively captured and utilized, primarily enhancing the potential of the model to identify delicate signs of tumors. The outcome is given as input into the classification layer after all the specific features are captured and integrated. This layer generally comprises fully connected layers, and then a softmax or sigmoid activation function is available to produce the final prediction values. The model completes the brain tumor detection process by analyzing the parameters to examine whether the tumor is present or not.

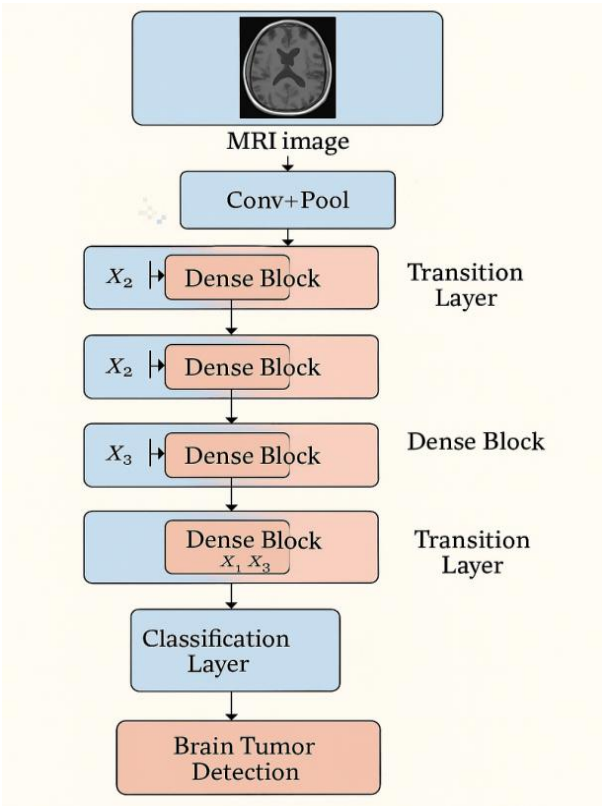

Figure 3. Dens Net architecture for brain tumor detection

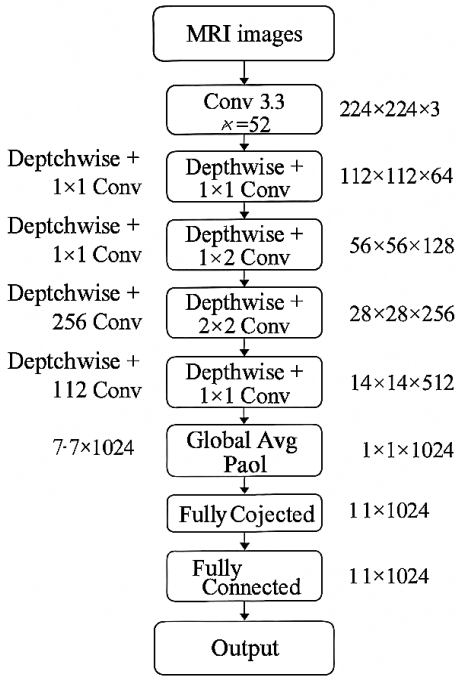

Figure 4. MobileNet models for brain tumor detection

The Figure 4 is shown in the MobileNet architecture is a compact deep-learning framework which offers higher computational efficiency and accuracy, making it more useful for brain tumor detection. It is specially designed for mobile and embedded devices but is just as useful for tasks involving the classification of medical images. The basic criteria for developing this MobileNet is its depth wise separable convolutions, which primarily reduce the attributes count and computing resources while maintaining the performance. Unlike standard convolutions, MobileNet factorizes them into two simpler functions: depthwise convolution and point wise convolution. The depthwise convolution filters all input channels individually, while the pointwise convolution (a 1×1 convolution), compiles the outcomes of the depthwise layer. This infrastructure significantly mitigates computing expenses. The architecture starts with an initial standard convolution layer and, then series of depthwise separable convolutional blocks. Each block generally comprises a 3×3 depthwise convolution, then batch normalization and a ReLU6 activation, followed by 1×1 pointwise convolution, again followed by batch normalization and activation. At the last stage of the network, a global average pooling layer mitigates the spatial dimensions and, then a fully connected layer that outputs class possibilities through a softmax function. In the aspects of brain tumor detection, this ultimate output layer classifies input MRI images into tumor types like glioma, meningioma, or pituitary tumors. The MobileNet maintained a balance between the model size and accuracy, ensuring highly suitable for fast and precise tumor diagnosis, specifically in systems having limited resources and in real-time diagnosis.

Figure 5 shows the Swin Transformer (Shifted Window Transformer) architecture for brain tumor segmentation. It shows the remarkable developments in vision transformer architectures, providing hierarchical representation learning via a unique window-based self-attention mechanism. Rather than traditional CNNs or early Vision Transformers process overall image patches globally, the Swin Transformer splits an image into non-overlapping local windows and evaluates self-

attention in these windows. An innovative shifted windowing technique is proposed to enable cross-window connections and improve the receptive field. This type of architecture enables the Swin Transformer more reliable and effective for high-resolution images like brain MRIs. Initially, the framework starts with separating input images into patches, which are directly integrated into patch tokens. These tokens transfer through the various hierarchical–based Swin Transformer blocks, each stage comprises shifted window attention layers and, then multilayer perceptrons (MLPs). Among these stages, patch merging operations mitigate spatial resolution and enhance the channel dimension, resembling CNN's feature pyramid structure. This hierarchical framework enables Swin Transformer to extract both local and global features efficiently. To detect the brain tumor, the output of a model is generally fed into the classification head (such as a fully connected softmax layer) that classifies the tumor based on learned features. Since it has the great potential to represent the relevant data and manage varied input resolution, the Swin Transformer exhibits higher accuracy in segmenting and classifying varied types of brain tumors in recent studies. It is especially suited for systems that require accurate positioning and segmentation of tumors in intricate medical images.

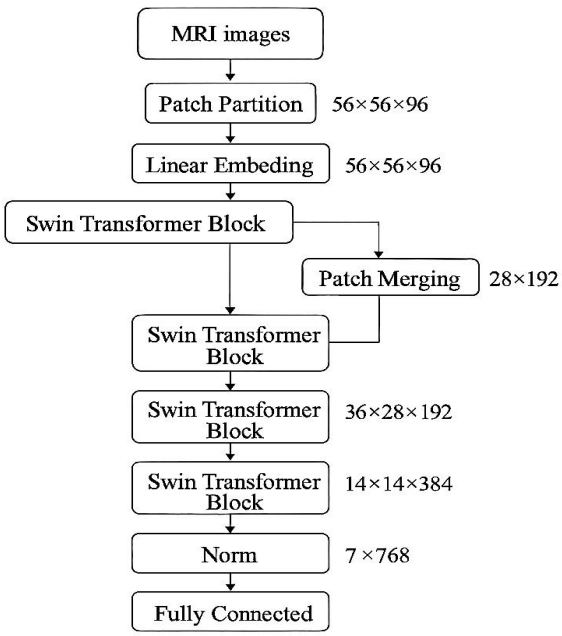

Figure 5. Swin Transformer brain tumour segmentation

Figure 6 shows an ensemble learning architecture, that integrates MobileNet and Swin Transformer models, specially designed for brain tumor detection. This hybrid model highlights the strength of individual models: MobileNet's compact, effective convolutional layers and Swin Transformer's hierarchical vision-based attention mechanism. This process starts with input MRI images that are pre-processed to eliminate noise and improve contrast. These images are transferred through both MobileNet and Swin Transformer branches at the same time. MobileNet utilizes depth-wise separable convolutions to manage the spatial feature efficiently with reduced computing intricacies. While the Swin Transformer extracts global-related data through shifted window-based self-attention mechanisms. A future representation is produced by each model and merged in a fusion layer. This layer compiles the localized feature extraction from MobileNet and long-range dependencies from Swin Transformer. These fused features are then transferred through fully connected layers for classification. The final prediction image is produced by the softmax layer, indicating whether the brain tumor is benign, malignant, or absent. This ensemble approach utilizes the individual model's potential by adjusting its weaknesses and enhancing classification accuracy and robustness. The architecture image also includes numerical annotations on all blocks, representing the number of layers, filters, or windows used in each stage, providing an extensive, structured visualization of the entire framework. This enables the model can be suited for both scholarly presentation and real-world application.

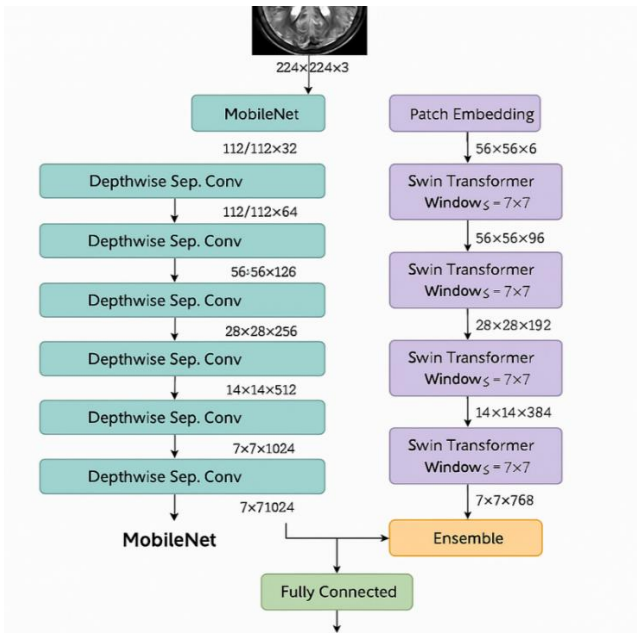

Figure 6. Ensemble classifier MobileNet and Swin Transformer

3.1 Fusion strategy of MobileNet and Swin Transformer

Our model fuses MobileNet and Swin Transformer to balance accuracy with efficiency. MobileNet captures fine-grained local patterns through lightweight convolutions, while Swin Transformer models global context using hierarchical self-attention. The outputs are projected to the same dimension and combined through a learnable gating block that adaptively weights local and global cues before classification. Unlike ResNet+Transformer or EfficientNet+Transformer, which require heavier backbones, our design achieves competitive accuracy with fewer parameters and lower computational cost, making the fusion both novel and practical.

4. EXPERIMENTAL RESULTS

4.1 Accuracy

An accuracy is calculated by the division of precise prediction and overall prediction. The first step starts with image extraction and then the extracted insights are compared with the overall dataset using the below mentioned mathematical expressions. while calculating the accuracy percentage (%), the two major factors considered are data quality and errors.

$$\text{Accuracy} = \frac{(TPV + TNV)}{(TPV + TNV + FPV + FNV)} \quad (1)$$

where, True Negative (TNV), True Positive (TPV), False positive (FPV), and False Negative (FNV).

4.2 Sensitivity

The sensitivity is evaluated by determining the values of true positives and false negatives from the datasets. The true positive and false negativity is calculated by adding the count values to the true positive. The quantity of positive outcomes is stated based on the calculation and the sensitivity is indicated from the output values. The sensitivity is calculated by the following mathematical notation in percentage (%).

$$\text{Sensitivity} = \frac{TPV}{(TPV + FNV)} \quad (2)$$

4.3 Specificity

The specificity is defined as the implementation result of the proposed model which is identified based on the impact of prediction and any variations from the original datasets. The specificity is determined by correctly analyzed negative counts and is expressed in percentage (%). It is the comprehensive count of negative values to the summation of true negative and false positive values. The mathematical representation of Specificity is as follows.

$$\text{Specificity} = \frac{TNV}{(TNV + FPV)} \quad (3)$$

The above Figure 7 shows the performance of the machine learning model in terms of two key attributes accuracy and loss across various epochs. In left graph, it shows the accuracy of training and validation across epochs ranging from 0.0 to 4.0. Likewise, the training and validation loss also represents a similar epoch range. The accuracy graph shows that the training and validation accuracy grows as the number of epochs increases, indicating that the model learns effectively. The loss range also reduces over increased epochs, illustrating that the model reduces the errors during training. But, without exact numerical values, it is very challenging to identify appropriate performance or detect possible concerns like overfitting or underfitting. The steady alignment of training and validation metrics indicates that the model's generalization ability is being closely monitored, highly significant for maintaining robust performance on unknown data. In general, the graph offers a clear visual depiction of the learning process of the model. The dataset used in the study is taken from [21].

To further validate the effectiveness of the proposed MobileNet–Swin Transformer fusion, we conducted comparative experiments with other CNN–Transformer ensembles, namely ResNet50+Swin and EfficientNetB1+Swin. The results show that while ResNet50+Swin and EfficientNetB1+Swin achieved competitive accuracy, they required considerably more parameters and higher computational cost. In contrast, the MobileNet–Swin ensemble delivered comparable or better accuracy with significantly fewer parameters and reduced inference complexity. This highlights the uniqueness of our approach, as it balances high performance with efficiency, making it more suitable for practical and resource-constrained clinical environments.

Table 1 presents the performance metrics of the classification process, effectively distinguishing between "tumor" and "non-tumor" cases. In case of both groups, the model offers high recall, precision, and F1-scores of 0.98, exhibiting optimized true positive detection accuracy with reducing false positives and false negatives. The model's robustness is further confirmed by achieving an overall accuracy of 0.98. Both macro average (equal weight each class) and weighted average (class-weighted) are constantly 0.98 across all metrics, demonstrating the performance balanced among classes with no bias. The efficient performance of this model highlights its reliability for medical diagnostic tasks, especially in brain tumor detection.

Figure 8 shows MobileNet performance of model in terms of two key attributes like accuracy and loss across various epochs. In accuracy graph, the accuracy starts at 75% and steadily rises to about 90%, ensuring steady progression. Since both curves retain closely matched throughout the training phase, this increasing nature across both validation and training accuracy shows the model's effective learning potential with no overfitting. Likewise, both training and validation loss values reduces steadily from 0.7 to 0.5, highlighting the model's learning potential to mitigate errors. As the model performs similar movement on both training and validation data, the simultaneous movement of these metrics indicates high generalization capacity. These outcomes demonstrate effective model convergence within the observed epoch range, while also suggesting potential for further optimization and performance gains with extended training. The persistent performance among all metrics, demonstrating model's integrity in its designated task for brain tumor segmentation.

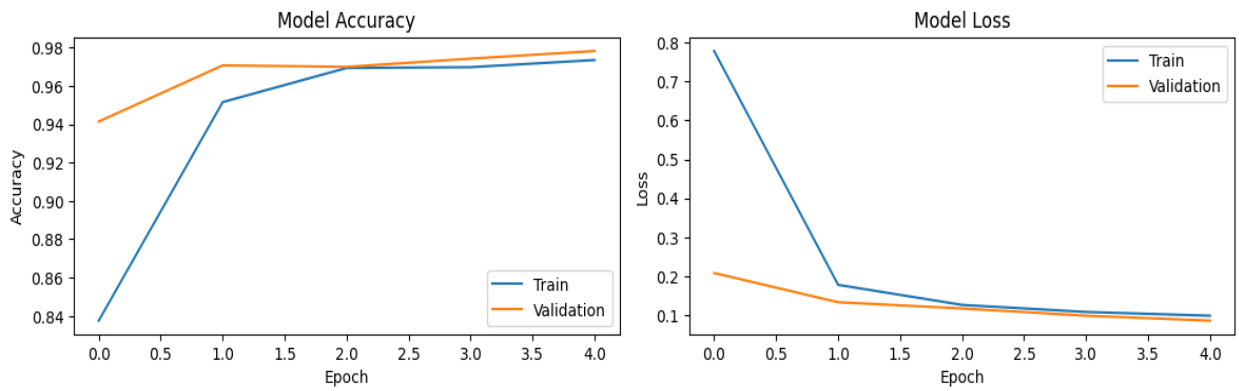

Figure 7. DenseNet accuracy and loss plots

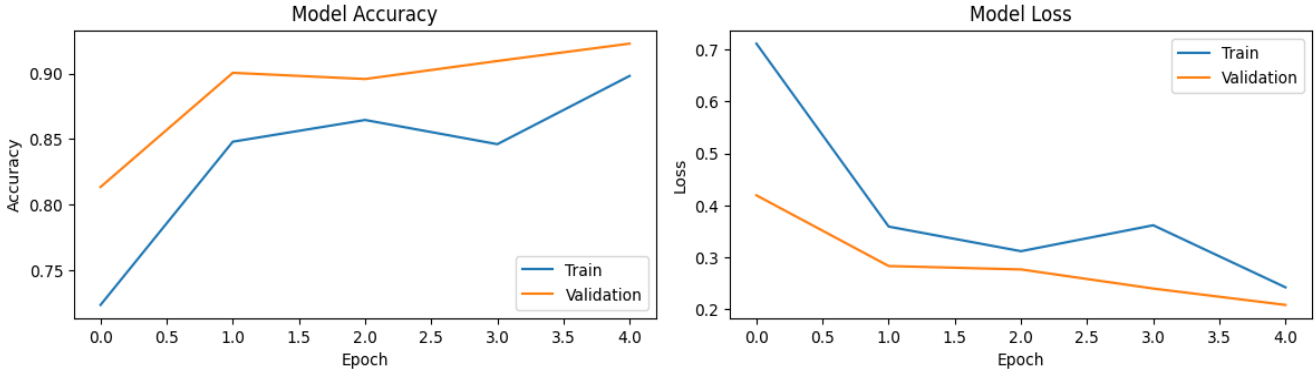

Figure 8. MobileNet accuracy and loss plot

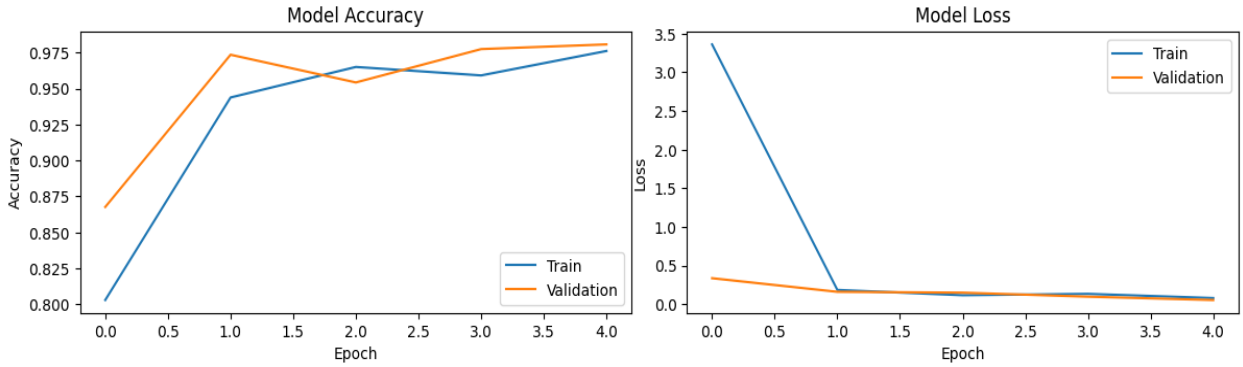

Figure 9. Swin Transformer accuracy and loss plots

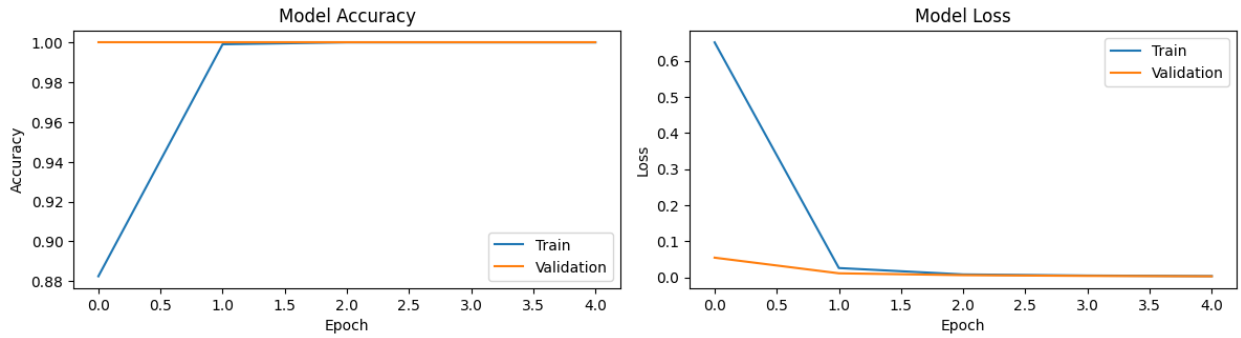

Figure 10. Ensemble accuracy and loss plots

Table 1. Dense Net classification report

Target	Precision	Recall	F1-score
Healthy	0.982	0.97	0.98
Brain Tumor	0.98	0.982	0.98
Accuracy			0.981
Macro avg	0.98	0.981	0.98
Weighted avg	0.98	0.98	0.98

Table 2. MobileNet classification report

Target	Precision	Recall	F1-score
Healthy	0.902	0.90	0.90
Brain Tumor	0.90	0.90	0.90
Accuracy			0.901
Macro avg	0.90	0.901	0.90
Weighted avg	0.90	0.90	0.90

Table 2 shows the performance of the classification task in distinguishing between tumor and non-tumor cases. Each key metric in this classification analysis achieved 0.90. There are similar precision, recall, and F1-score values across both classes, indicating that there is identical reliability in detecting true positives (tumors) with mitigating false positives/negatives. Macro and weighted average achieves 90%, illustrating the model's generalization without bias. This uniform result shows that the model offers balanced performance in diagnostic screening for both cases, making it well-suited for medical applications like detecting tumors and critical healthy cases.

Figure 9 shows the performance of Swin Transformer model in terms of two key attributes like accuracy and loss across various epochs. As per the accuracy and loss metrics plotted in the graph, it demonstrates that the model's learning efficiency across all epochs. The accuracy line shows a steady improvement, with starting at about 97.5% and slightly reduced to 90-92.5% at epoch (4). This slight decrease, along with strong similarity between the training and validation accuracy lines, demonstrates effective learning ability with no overfitting. Likewise, the loss values start declining and reaching close to zero value by the final epoch, which shows

that the model learns effectively to mitigate the errors in the entire training phase. Both training and validation loss retain in a similar declining range, further illustrates that the model has the robust learning ability and powerful generalization capacity. These outcomes shows that the model has learned the basic patterns in the data effectively with achieving outstanding performance on both training and validation subset, ensuring the reliability in its designated tasks.

Table 3. Swin Transformer classification report

Target	Precision	Recall	F1-score
Healthy	0.972	0.97	0.972
Brain Tumor	0.97	0.971	0.972
Accuracy			0.97
Macro avg	0.971	0.97	0.971
Weighted avg	0.97	0.97	0.97

Table 4. Ensemble classification report

Target	Precision	Recall	F1-score
Healthy	0.9948	0.9925	0.9963
Brain Tumor	0.9952	0.9936	0.9945
Accuracy	0.9965	0.9845	0.9865
Macro avg			0.9965
Weighted avg	0.9965	0.9945	0.9965

Table 3 shows the classification report of the swing transformer in terms of metrics like precision, recall, and F1-score. For both healthy and tumor classes, it achieves 97% across all metrics. Each metric measure is similar for all classes, demonstrating a balanced diagnostic ability. The consistency is maintained in both macro and weighted average (each at 0.97) further confirming the model's unbiased and dependable performance, irrespective of class distribution. These robust and consistent outcomes demonstrate that this model is medically appropriate for high-stakes brain tumor detection, where even the smallest performance margins are critical.

The above Figure 10 shows the ensemble performance of model in terms of two key attributes like accuracy and loss across various epochs. In the left graph, the training accuracy starts with 88% and rapidly increases to 100% by the epoch (2) and stays stable. The validation accuracy remains 100% throughout the entire epochs, indicating effective generalization. In the right graph, the training loss declines sharply from 0.6 to nearly 0, while the validation loss starts with minimal loss and remains low. These findings illustrate the effectiveness of model's learning ability and converge rapidly.

Table 4 shows the classification report of the ensemble model. Across all classes, the model highlights strong performance in distinguishing between healthy subjects and brain tumor cases, achieving near-optimal evaluation metrics. In the case of both classes, Precision, recall, and F1-scores are above 0.99, and the model achieved an outstanding accuracy of 0.9965 in correctly identifying true positives while effectively reducing errors. The macro and weighted averages also achieved 0.9965, ensuring consistent and unbiased performance among all datasets. These findings illustrate the model is significantly reliable for medical systems, providing precision and recall values, that align with stringent medical regulations. This level of performance makes the model well-suited for medical applications like brain tumor segmentation, where precision and accuracy are more critical.

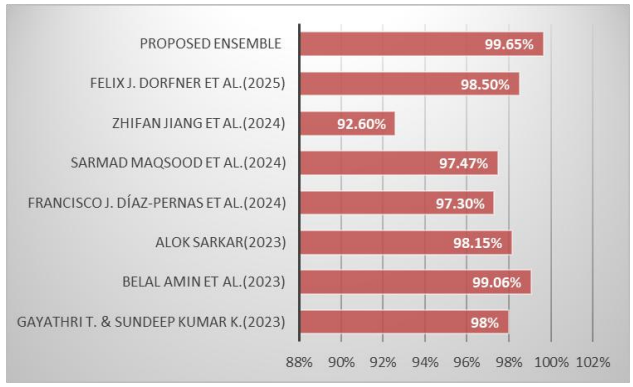

Figure 11. Comparisons between existing and proposed method

Table 5. Comparison between existing and proposed methodology

Year	Authors	Dataset	Model	Accuracy (%)	Precision (%)	Recall (%)
2023	Gayathri and Sundeep Kumar [20]	BraTS 2015, 2017, 2019	CNN–ResNeXt	98.00	97.50	97.80
2024	Wei [22]	Public MRI Dataset	EfficientNetB1 (Classification), U-Net (Segmentation)	99.06	98.73	99.13
2023	Sarkar et al. [23]	Kaggle MRI Dataset	AlexNet CNN	98.15	97.80	98.00
2021	Díaz-Pernas et al. [24]	3064 slices from 233 patients	Multiscale CNN	97.30	96.80	97.00
2021	Maqsood et al. [25]	T1-weighted contrast-enhanced MRI	MobileNetV2	97.47	96.90	97.20
2024	Capellán-Martín et al. [26]	BraTS 2024	Ensemble of State-of-the-Art Models	92.60	91.50	92.00
2023	Potadar et al. [27]	Multi-sequence MRI	Swin Transformer	98.50	98.00	98.30

Figure 11 presents the comparison results summarized in Table 5 for the accuracy of brain tumor detection models from 2023 to 2025. The proposed model achieved the highest accuracy of 99.65%, outperforming all existing approaches. In comparison, Amin et al. [28] and Dorfner et al. [29] achieved accuracies of 99.06% and 98.50%, respectively, while Jiang et al. [30] reported the lowest accuracy of 92.60%. These outcomes highlight the effectiveness of the proposed ensemble model.

Figure 12 illustrates the performance comparison of multiple brain tumor detection models in terms of accuracy, precision, and recall. Among all models, the proposed approach demonstrates superior performance, attaining 99.65% accuracy, 99.45% precision, and 99.60% rec

all, which indicates its strong predictive capability and consistency. Dorfner et al. [29] also achieved competitive results close to those of the proposed model, whereas Amin et al. [28] and Sarkar et al. [23] showed balanced and comparable performance with scores exceeding 98% across all metrics. In contrast, Jiang et al. [30] exhibited lower performance, with all metrics around 92%, indicating the need for further improvement. Overall, the results clearly demonstrate the effectiveness and reliability of the proposed ensemble model compared with existing methods.

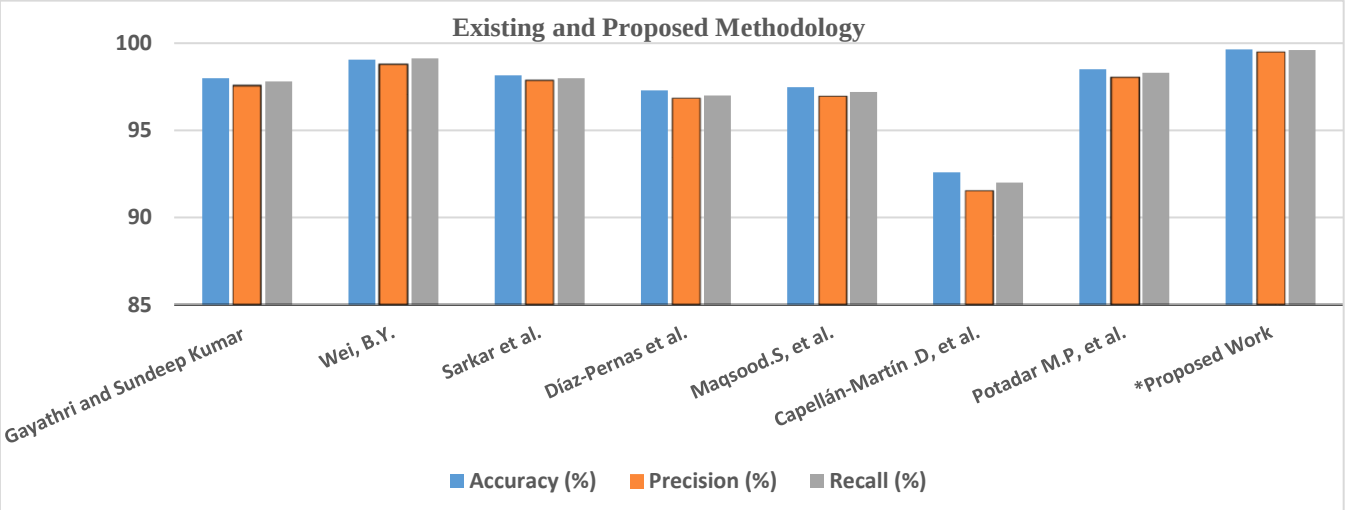

Figure 12. Comparison between the existing and proposed accuracy, precision, recall measures

5. CONCLUSIONS

In this research we present a novel ensemble deep learning framework that combines the strengths of Mobile Net and Swin Transformer architectures to improve the accurate detection of brain tumors from MRI scans. Mobile Net, known for its lightweight and fast convolutional operations, is paired with the Swin Transformer, which excels at capturing both local and global contextual information through its hierarchical self-attention mechanism. Together, these models form a hybrid system that was rigorously evaluated on benchmark brain tumor datasets, achieving an outstanding accuracy of 99.65%, significantly outperforming several traditional deep learning models. The results highlight that the proposed ensemble is not only computationally efficient but also highly dependable for real-time clinical use. It effectively balances speed and performance, making it an excellent fit for medical environments where computational resources may be limited. Looking ahead, this model could be extended to incorporate multimodal imaging data such as PET and CT scans, enabling even more detailed tumor analysis. Additionally, integrating patient demographic information and medical metadata could further strengthen the model's decision-making abilities. To enhance transparency and trust in its predictions, the model also incorporates explainable AI (XAI) techniques, allowing physicians to visualize and better understand its decision processes. Real-world deployment on edge devices and testing across datasets from diverse institutions also suggest strong potential for broad medical adoption and adaptability.

Although the proposed MobileNet–Swin Transformer fusion model demonstrated strong performance, the study has some limitations. Detailed efficiency measures such as parameter counts, FLOPs, and inference time were not included, as the main focus was on methodological validation through Python-based experiments. These metrics will be addressed in future work to better assess the framework's suitability for deployment on different hardware platforms. In addition, while the importance of explainable AI (XAI) was acknowledged, no interpretability experiments were presented in this version. Future extensions will incorporate visualization techniques such as Grad-CAM and attention heatmaps to provide greater transparency and support clinical trust in the model's predictions.

REFERENCES

[1] Siegel, R.L., Miller, K.D., Jemal, A. (2015). Cancer statistics, 2015. CA: A Cancer Journal for Clinicians, 65(1): 5-29. https://doi.org/10.3322/caac.21254

[2] Srinivasan, S., Bai, P.S.M., Mathivanan, S.K., Muthukumaran, V., Babu, J.C., Vilcekova, L. (2023). Grade Classification of Tumors from Brain Magnetic Resonance Images Using a Deep Learning Technique. Diagnostics, 13(6): 1153. https://doi.org/10.3390/diagnostics13061153

[3] Wadhwa, A., Bhardwaj, A., Verma, V.S. (2019). A review on brain tumor segmentation of MRI images. Magnetic Resonance Imaging, 61: 247-259. https://doi.org/10.1016/j.mri.2019.05.043

[4] Liu, Z.H., Tong, L., Chen, L., Jiang, Z.H., Zhou, F.X., Zhang, Q.N., Zhang, X.R., Jin, Y.C., Zhou, H.Y. (2023). Deep learning based brain tumor segmentation: A survey. Complex & Intelligent Systems, 9: 1001-1026. https://doi.org/10.1007/s40747-022-00815-5

[5] Patil, D.D., Deore, S.G. (2013). Medical image segmentation: A review. International Journal of Computer Science and Mobile Computing, 2(1): 22-27.

[6] Shi, J.J., Chan, T.T.D., Pan, H.Y., Lok, T.M. (2023). Reconfigurable intelligent surface assisted semantic communication systems. In Proceedings of the 2023 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC),

Zhengzhou, China, pp. 1-6. https://doi.org/10.1109/ICSPCC59353.2023.10400366
[7] Enoch Sit, C.Y.E., Kong, S.C. (2024). A deep learning framework with visualisation for uncovering students' learning progression and learning bottlenecks. Journal of Educational Computing Research, 62(1): 3-29. https://doi.org/10.1177/07356331231200600
[8] Long, J., Shelhamer, E., Darrell, T. (2017). Fully convolutional networks for semantic segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(4): 640-651. https://doi.org/10.1109/TPAMI.2016.2572683
[9] Shen, H.C., Zhang, J.G., Zheng, W.S. (2017). Efficient symmetry-driven fully convolutional network for multimodal brain tumor segmentation. In 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, pp. 3864-3868. https://doi.org/10.1109/ICIP.2017.8297006
[10] Zhou, T.X. (2024). Multi-modal brain tumor segmentation via disentangled representation learning and region-aware contrastive learning. Pattern Recognition, 149: 110282. https://doi.org/10.1016/j.patcog.2024.110282
[11] Zhou, T.X. (2023). Feature fusion and latent feature learning guided brain tumor segmentation and missing modality recovery network. Pattern Recognition, 141: 109665. https://doi.org/10.1016/j.patcog.2023.109665
[12] Zhu, Z.Q., Wang, Z.Y., Qi, G.Q., Mazur, N., Yang, P., Liu, Y. (2024). Brain tumor segmentation in MRI with multi-modality spatial information enhancement and boundary shape correction. Pattern Recognition, 153: 110553. https://doi.org/10.1016/j.patcog.2024.110553
[13] Ranjbarzadeh, R., Zarbakhsh, P., Caputo, A., Tirkolaee, E.B., Bendechache, M. (2024). Brain tumor segmentation based on optimized convolutional neural network and improved chimp optimization algorithm. Computers in Biology and Medicine, 168: 107723. https://doi.org/10.1016/j.compbiomed.2023.107723
[14] Montaha, S., Azam, S., Rakibul Haque Rafid, A., Hasan, M.Z., Karim, A. (2023). Brain tumor segmentation from 3D MRI scans using U-Net. SN Computer Science, 4: 386. https://doi.org/10.1007/s42979-023-01854-6
[15] Feng, L.P., Wu, K.P., Pei, Z.Y., Weng, T.F., Han, Q., Meng, L. (2024). MLU Net: A multi-level lightweight U-Net for medical image segmentation integrating frequency representation and MLP-based methods. IEEE Access, 12: 20734-20751. https://doi.org/10.1109/ACCESS.2024.3360889
[16] Zhang, W., Chen, S.X., Ma, Y.Q., Liu, Y., Cao, X. (2024). ETUNet: Exploring efficient transformer enhanced UNet for 3D brain tumor segmentation. Computers in Biology and Medicine, 171: 108005. https://doi.org/10.1016/j.compbiomed.2024.108005
[17] Hammer Håversen, A., Bavirisetti, D.P., Hanssen Kiss, G., Lindseth, F. (2024). Qt-UNet: A self-supervised self-querying all-transformer U-Net for 3D segmentation. IEEE Access, 12: 62664-62676. https://doi.org/10.1109/ACCESS.2024.3395058
[18] Hussain, T., Shouno, H. (2024). MAGRes-UNet: Improved medical image segmentation through a deep learning paradigm of multi-attention gated residual U-Net. IEEE Access, 12: 40290-40310. https://doi.org/10.1109/ACCESS.2024.3374108
[19] Verma, A., Shivhare, S.N., Singh, S.P., Kumar, N., Nayyar, A. (2024). Comprehensive review on MRI-based brain tumor segmentation: A comparative study from 2017 onwards. Archives of Computational Methods in Engineering, 31: 4805-4851. https://doi.org/10.1007/s11831-024-10128-0
[20] Gayathri, T., Sundeep Kumar, K. (2023). A deep learning based effective model for brain tumor segmentation and classification using MRI images. Journal of Advances in Information Technology, 14(6): 1280-1288.https://doi.org/10.12720/jait.14.6.1280-1288
[21] Brain_Tumor_Detection_MRI. https://www.kaggle.com/datasets/abhranta/brain-tumor-detection-mri, accessed on Jun 24, 2025.
[22] Wei, B.Y. (2024). Brain tumor MRI segmentation method based on segment anything model. Revue d'Intelligence Artificielle, 38(2): 567-573. https://doi.org/10.18280/ria.380220
[23] Sarkar, A., Maniruzzaman, M., Alahe, M.A., Ahmad, M. (2023). An effective and novel approach for brain tumor classification using AlexNet CNN feature extractor and multiple eminent machine learning classifiers in MRIs. Journal of Sensors, 2023(1): 1224619. https://doi.org/10.1155/2023/1224619
[24] Díaz-Pernas, F.J., Martínez-Zarzuela, M., Antón-Rodríguez, M., González-Ortega, D. (2021). A deep learning approach for brain tumor classification and segmentation using a multiscale convolutional neural network. Healthcare, 9(2): 153. https://doi.org/10.3390/healthcare9020153
[25] Maqsood, S., Damasevicius, R., Shah, F.M. (2021). An efficient approach for the detection of brain tumor using fuzzy logic and U-Net CNN classification. In Computational Science and Its Applications – ICCSA 2021. ICCSA 2021. Lecture Notes in Computer Science, pp 105-118. https://doi.org/10.1007/978-3-030-86976-2_8
[26] Capellán-Martín, D., Jiang, Z.F., Parida, A., Liu, X.Y., et al. (2024). Model ensemble for brain tumor segmentation in magnetic resonance imaging. arXiv preprint arXiv:2409.08232. https://doi.org/10.48550/arXiv.2409.08232
[27] Potadar, M.P., Holambe, R.S., Chile, R.H. (2023). Design and development of a deep learning model for brain abnormality detection using MRI. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 12(1): 2250878. https://doi.org/10.1080/21681163.2023.2250878
[28] Amin, B., Samir, R.S., Tarek, Y., Ahmed, M., Ibrahim, R., Ahmed, M., Hassan, M. (2023). Brain tumor multi classification and segmentation in MRI images using deep learning. arXiv preprint arXiv:2304.10039. https://doi.org/10.48550/arXiv.2304.10039
[29] Dorfner, F.J., Patel, J.B., Kalpathy-Cramer, J., Gerstner, E.R., Bridge, C.P. (2025). A review of deep learning for brain tumor analysis in MRI. NPJ Precision Oncology, 9(1): 2. https://doi.org/10.1038/s41698-024-00789-2
[30] Jiang, Z., Capellán-Martín, D., Parida, A., Tapp, A., Liu, X., Ledesma-Carbayo, M.J., Linguraru, M.G. (2024). Magnetic resonance imaging feature-based subtyping and model ensemble for enhanced brain tumor segmentation. arXiv preprint arXiv:2412.04094. https://doi.org/10.48550/arXiv.2412.04094