

ORIGINAL ARTICLE

# Automated brain tumor detection using novel hybrid vision transformer and EfficientNetB4 models with comparative analysis

S. Selva Karthik<sup>1</sup> · V. Achuthan<sup>1</sup> · A. Dennis Ananth<sup>1</sup> · R. Venkatesan<sup>1</sup> · M. Rajakumaran<sup>1</sup> · S. Markkandeyan<sup>1</sup>

Received: 12 December 2024 / Accepted: 4 May 2025

© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2025

## Abstract

A brain tumor is an abnormal growth or mass of cells in or around the brain. Early detection of brain tumors is imperative, impacting the quality of life and potential fatality. Prolonged undetected brain tumors can cause irreversible brain damage. Early detection enables medical intervention to prevent severe harm, preserving cognitive function and reducing permanent damage risk. These tumors come in a wide variety of sizes, locations, and other characteristics. When trying to locate cancerous tumors, magnetic resonance imaging (MRI) is a crucial tool. However, detecting brain tumors manually is a difficult and time-consuming task that might lead to inaccuracies. Many researchers investigated a variety of algorithms for detecting and classifying brain tumors that were both accurate and fast. Deep learning (DL) approaches have recently been popular in developing automated systems capable of accurately diagnosing or segmenting various tumors in less time. In this research, we propose several ways to detect brain cancer and tumors using computational intelligence and statistical image processing techniques. We use three different deep learning architecture models along with data augmentation and image processing to categorize brain MRI scan images into cancerous and non-cancerous types. We later conducted a comparative analysis of our models: EfficientNetB4, Vision Transformer (ViT) combined with EfficientNetB4 (a novel hybrid model), and a custom CNN model built from scratch. The experiment results demonstrate that all models achieved high accuracy and very low complexity rate. Specifically, EfficientNetB4 achieving 99.76%, 99.6% in Vision Transformer + EfficientNetB4, and scratch CNN achieved 97.25% accuracy. Our models require very less computational power and have much better accuracy results as compared to other pretrained models.

**Keywords** Brain tumor detection · MRI classification · EfficientNetB4 · Vision Transformer (ViT) · CNN (convolution neural network)

## 1 Introduction

A brain tumor is a growth of cells in the brain or near the brain. Brain tumors can occur in the brain tissue and nearby. Nearby locations include nerves, the pituitary gland, the pineal gland, and the membranes that cover the surface of the brain. Brain tumors can originate in the brain. These are called primary brain tumors. Sometimes, cancer spreads to the brain from other parts of the body. These tumors are secondary brain tumors, also called metastatic brain tumors.



Many different types of primary brain tumors exist. Some brain tumors are not cancerous. These are called benign brain tumors. Benign brain tumors may grow over time and press on brain tissue. Other brain tumors are brain cancers, also called malignant brain tumors. Brain cancers may grow quickly. The cancer cells can invade and destroy the brain tissue. Brain tumor treatment options depend on tumor size, type, and location. Common treatments include surgery and radiation therapy [1]. There are two types of brain tumors:

**Malignant brain tumors:** Gliomas and related brain tumors, Embryonal, Germ cell, Pineal tumors.

**Benign brain tumors:** Choroid plexus, Meningiomas, Pituitary nerve tumors [2].

With over 300,000 cases reported annually on a worldwide, brain tumors remain a pressing concern for the international medical community. While some brain tumors may be benign, many can invade normal brain and develop into brain cancer. Approximately 72% of all brain tumors are benign and 28% of all brain tumors are malignant. An estimated 67,440 cases will be non-malignant (benign) in 2024. Non-malignant meningiomas are the most commonly occurring primary non-malignant brain tumors, accounting for 39.7% of all tumors and 55.4% of all non-malignant tumors. Glioblastoma is the most commonly occurring primary malignant brain tumor, accounting for 14.2% of all tumors and 50.1% of all malignant tumors [3].

Magnetic resonance imaging (MRI) and biopsy are routinely performed for the diagnosing intracranial tumors, with biopsy considered a criterion standard for the classification of tumor types. Although a standard practice, biopsy has associated challenges because it is invasive. Therefore, identification and accurate classification of tumor subtypes from noninvasive methods like MRI are desired. However, distinguishing different types tumors from MRI images can be challenging due to the similarities in tumor appearance on scans. Therefore, accurate, reliable preoperative determination of tumor types from MRI may facilitate rapid clinical decision-making and aid in better treatment planning [4].

Deep learning (DL) is widely used in healthcare for analysis, classification, and detection. The processing capability of CNNs is derived from a computational model inspired by the structure and functioning of the human brain. Humans can see and identify things by relying on their external visual characteristics. The CNN, known for its proficiency in image processing, operates similarly. Several well-recognized CNN models include ResNet, GoogLeNet, AlexNet, and VGG. Recently, deep learning (DL) has been employed to enhance diagnostic accuracy in classification and detection jobs within biomedical engineering. Deep learning (DL) approaches have been shown to enhance performance due to their ability to extract deep characteristics, resulting in effective detection and classification. Hence, the proposed computer-aided design (CAD) system employs machine learning (ML) and deep learning (DL) methodologies to accurately classify and assess several categories of brain tumors based on brain MRI.

## 2 Literature review

In recent years, advancements in deep learning (DL) have significantly enhanced the detection and classification of brain tumors using magnetic resonance imaging (MRI) images. Various studies have explored different algorithms, datasets, and methodologies to achieve high accuracy and efficiency in diagnosis.

Akmalbek Abdusalomov et al. [5] employed YOLOv7 for brain tumor detection, achieving an impressive 99.5% accuracy on a Kaggle dataset containing 10,288 MRI images categorized into gliomas, meningiomas, pituitary tumors, and no tumor. Similarly, Sandeep Kumar Mathivanan et al. [6] utilized models such as ResNet152, DenseNet169, VGG19, and MobileNetV3, with ResNet152 achieving the highest accuracy of 99.75%. These studies underscore the effectiveness of CNNs in identifying intricate patterns in medical imaging.

Models based on transfer learning, such as EfficientNet, ResNet50, and ResNet101-CWAM, have shown exceptional results in brain tumor classification. For instance, Nishanth Rao et al. [7] reported an accuracy of 98% using EfficientNet, while ensemble techniques combining models like random forests and support vector machines (SVMs), as demonstrated by Balamurugan et al. [8] and Patel Rahulkumar Manilal et al. [9], achieved accuracies of up to 99.83%. Proprietary architectures like BrainCDNet by Rasool Reddy [10] and TumorDetNet

by Naeem Ullah et al. [11] delivered binary classification accuracy of 99.45% and multi-class accuracy of 96.78%, leveraging extensive Kaggle datasets to enhance diagnostic precision.

Shikha Jain et al. [12] highlighted the capabilities of artificial neural networks (ANNs), which achieved a remarkable 99% accuracy, outperforming other models like SVM and Vision Transformers. Similarly, Novsheena Rasool et al. [13] presented a systematic review showcasing overall accuracies exceeding 98% using datasets such as TCIA and BraTS.

Hybrid methodologies combining CNNs, UNET, and Gabor filters, as explored by researchers like Soheila Saeedi et al. [14] and Buchade et al. [15], achieved accuracies around 98%. These studies demonstrate the power of integrating deep learning with traditional image processing methods. For segmentation tasks, Md Kamrul Hasan Khan et al. [16] used nnU-Net, reporting dice scores (a metric that measures the similarity between predicted and actual regions, ranging from 0 to 1) of up to 0.931 on BraTS datasets.

Emerging models continue to improve classification accuracy while focusing on computational efficiency and cost-effectiveness. Ghauri et al. (2024) [17] developed a clean-energy platform using 2D CNNs with precision rates of 96.8%. These innovations pave the way for real-world deployment in clinical settings.

## 2.1 Brain tumor detection using machine learning techniques

Recent advancements in machine learning (ML) and deep learning (DL) have enabled significant progress in the field of brain tumor detection and classification. Here is a summarized overview of methodologies, algorithms, accuracies, and datasets used in various studies. Seyed Matin Malakouti et al. [18] demonstrated the effectiveness of transfer learning by using GoogLeNet, which achieved an impressive accuracy of 99.3% on a Kaggle dataset [19]. In the same study, LightGBM was employed, yielding a lower accuracy of 95.7%, highlighting the superior performance of deep learning models in this field.

Traditional machine learning (ML) techniques like SVM and clustering algorithms continue to play a critical role. For instance, Javaria Amin et al. [20] applied feature extraction and SVM classification, achieving 99% accuracy on the BraTS challenge dataset. Similarly, Keerthana et al. [21] utilized K-means clustering combined with SVM for brain tumor detection. Deep learning models, particularly CNNs, consistently achieve high accuracies in tumor classification. Manav Sharma et al. [22] reported 97.79% accuracy using CNNs and KNN on the Brats 2015 dataset. Rajan Hossain et al. [23] employed a combination of ANN, CNN, random forests, and SVM, achieving an accuracy of 95% with a dataset collected from MP MRI & CT Scan Centre. M. Aarthilakshmi et al. [24] integrated CNNs and Fuzzy C-means (FCM) algorithms, achieving 91% accuracy for binary classification of MRI scans, distinguishing between tumor and non-tumor cases. Meanwhile, Asma Parveen and Dr. Kamalakkannan [25] combined support vector machine (SVM) with CNN, achieving 92.29% accuracy using MRI data obtained from the BIDS website [26].

Clustering algorithms like K-means and fuzzy C-means (FCM) are also widely used for tumor segmentation. Studies by Siddhant Barshile et al. [27] and Alakuntha Raja Shekar et al. [28] integrated these methods with models like CNNs and random forests, reporting dice similarity coefficient metrics as high as 0.88, reflecting their effectiveness in accurately segmenting tumor regions. The datasets used in these studies range from publicly available ones to proprietary collections. Notably, the BraTS Challenge datasets (2013–2015) were frequently utilized, as seen in the studies by Manav Sharma et al. [22]. Additionally, proprietary datasets, such as those collected from NSCB Medical College, were used by Rajan Hossain et al. [23] to evaluate model performance in real-world scenarios.

## 2.2 Brain tumor detection using hybrid model

The integration of hybrid deep learning (DL) architectures has further revolutionized brain tumor detection, offering high accuracy and robust performance. These models combine the strengths of various algorithms and

techniques, enabling more precise classification and detection. Below is an overview of recent studies highlighting methodologies, algorithms, accuracies, and datasets used in brain tumor detection.

Mahfuz Ahmed et al. [29] combined Vision Transformers (ViT) with gated recurrent units (GRUs), achieving 97% accuracy on a Kaggle dataset containing 3,264 MRI images categorized into glioma, meningioma, pituitary tumors, and no tumor. Similarly, Priya and Vasudevan [30] utilized an AlexNet-GRU hybrid model on a dataset of 7,023 images, also achieving 97% accuracy, showcasing the effectiveness of GRUs in enhancing traditional deep learning architectures. EfficientNet variants have also demonstrated outstanding performance in tumor classification tasks. Baiju Babu Vimala et al. [31] achieved 99.06% accuracy by leveraging EfficientNetB0 to EfficientNetB4 combined with Grad-CAM for enhanced interpretability, using the CE MRI Figshare dataset. Additionally, Jose Dixon et al. [32] employed a hybrid model integrating ResNet101, DenseNet121, and EfficientNetB0, achieving 99.18% accuracy on the Figshare dataset [33] and 97.24% on a Kaggle dataset [34].

Transfer learning models have proven remarkably effective as well. Rohini et al. [35] implemented a hybrid VGG19 model, achieving an impressive 99.43% accuracy on a Kaggle brain tumor dataset. Similarly, Saeed Mohsen et al. [36] combined ResNet101 with VGG19, reaching an exceptional 99.98% accuracy on a glioma and pituitary tumor dataset. Hybrid classifiers have been instrumental in improving classification accuracy. For example, Balamurugan and Gnanamanoharan [37] proposed a CNN-LuNet hybrid model, achieving 99.7% accuracy. Likewise, Gehad Abdullah Amran et al. [38] developed DeepTumorNet, integrating CNN and GoogLeNet, and reported 99.51% accuracy using the Br35H dataset from Kaggle [39].

For smaller datasets, simpler models have shown notable results. Ebrahim Mohammed Senan et al. [40] combined AlexNet and ResNet18, achieving 95.10% accuracy on a dataset of 253 MRI images. Similarly, AlTahhan et al. [41] utilized an AlexNet-KNN hybrid model, achieving 98.6% validation accuracy on a composite dataset of 2880 T1-weighted MRI images from Figshare, SARTAJ, and Br35H sources. The datasets used across these studies, including Kaggle, Figshare, CE MRI, and proprietary sources, include various tumor types such as glioma, meningioma, pituitary tumors, and non-tumor cases. These diverse datasets provide a robust foundation for model evaluation and highlight their potential for real-world clinical applications, as demonstrated by studies like those by Gehad Abdullah Amran et al. [38] and Md. Mahfuz Ahmed et al. [42].

### 3 Materials and methods

The methodology includes multiple stages. The dataset was split by 80% training, 20% testing. We checked four types of brain tumors: glioma, meningioma, no tumor, and pituitary tumors. To verify our findings, we evaluated accuracy, specificity, and sensitivity.

#### 3.1 Experimental setup

The proposed architectures are implemented using Python 3.12 software on Intel(R) Xeon(R) Silver 4310 CPU @ 2.10 GHz 256 GB RAM with NVIDIA A100 Tensor Core (40 GB Memory) GPU.

#### 3.2 Dataset

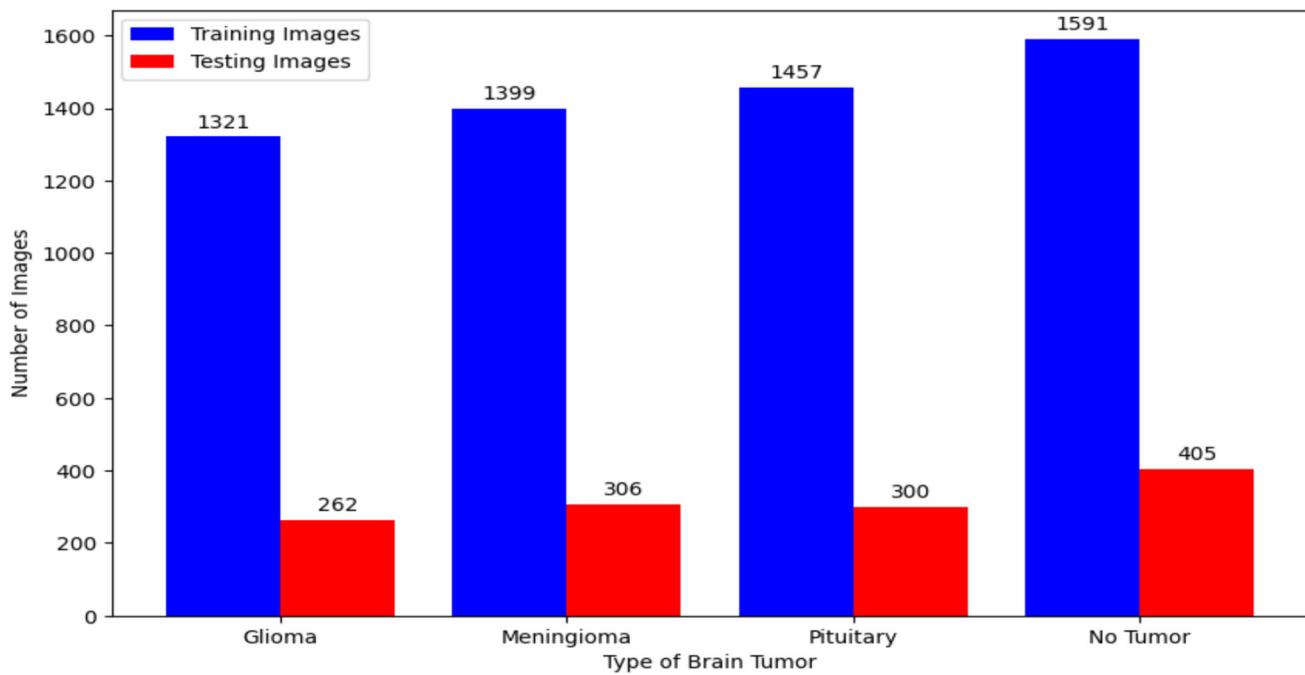
The brain tumor classification dataset was sourced from kaggle.com [43], and we summarize the dataset structure in Table 1.

This dataset contains **7041** human brain MRI images which are classified into 4 classes: glioma, meningioma, no tumor, and pituitary (Fig. 1).

**Table 1** A comprehensive overview of the dataset's structure

Type of brain tumor	No. of training images	No. of testing images
Glioma	1321	262
Meningioma	1399	306
Pituitary	1457	300
No Tumor	1591	405
Total	<b>5768</b>	<b>1273</b>

Bold values indicate the best performance results among the compared models or methods for each evaluation metric



**Fig. 1** Distribution of dataset among the four classes

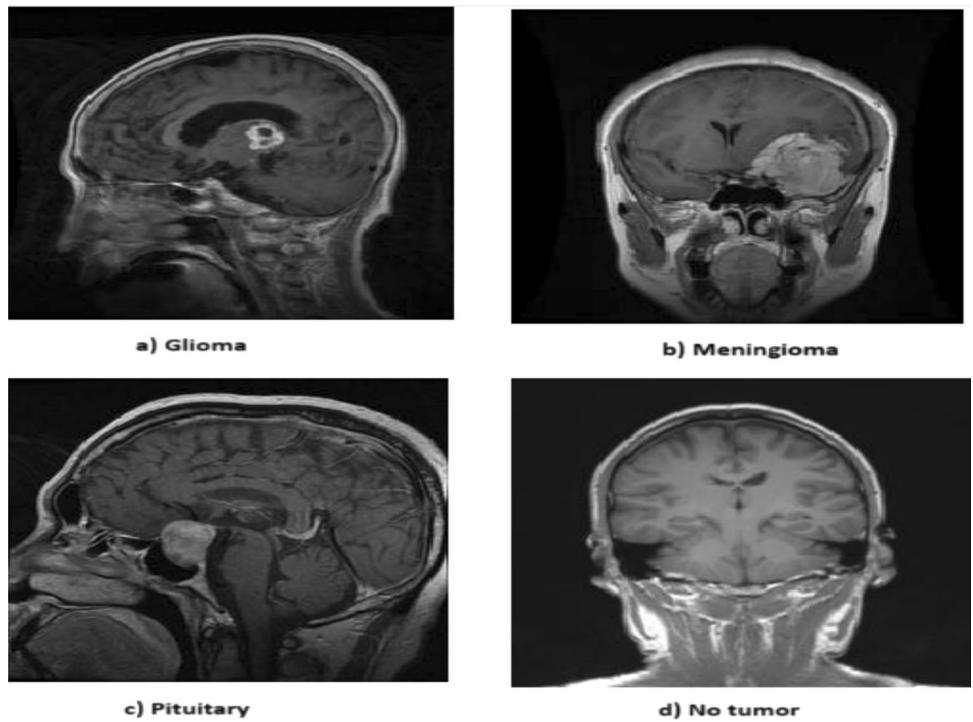
### 3.3 Proposed methodology

EfficientNet models are a series of convolutional neural networks (CNNs) developed by Google AI researchers to achieve high accuracy with fewer computational resources [44]. Introduced in 2019, these models utilize a novel concept called compound scaling, which uniformly scales a model's width, depth, and resolution. This balanced scaling results in highly efficient and scalable architectures, capable of delivering state-of-the-art performance on various computer vision tasks while requiring significantly fewer floating-point operations per second (FLOPS). The EfficientNet family, ranging from EfficientNetB0 to EfficientNetB7, offers various trade-offs between accuracy and computational efficiency, making them ideal for a wide range of applications.

#### 3.3.1 Why EfficientNetB4?

The EfficientNetB4 model architecture was chosen for its efficient scaling and high performance with relatively few parameters. Compared to smaller variants like EfficientNetB0, which may struggle with capturing the complexity of medical imaging data, EfficientNetB4 offers higher accuracy without the excessive computational overhead associated with larger variants like EfficientNetB7. EfficientNetB4 has been shown to provide strong performance across various vision tasks, making it suitable for complex medical image classification.

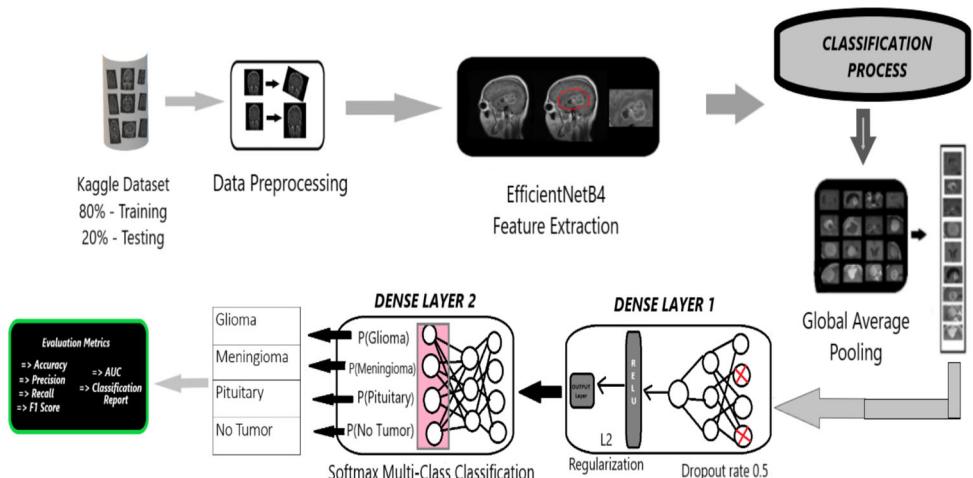
**Fig. 2** Sample MRI images from the dataset



Additionally, it strikes an ideal balance for fine-tuning, allowing for effective adaptation to the dataset while preventing overfitting. Given its proven success in medical imaging applications, EfficientNetB4 is well suited for handling the intricate details of brain MRI or CT scan images, offering superior generalization and scalability as the dataset grows. Thus, EfficientNetB4 was selected as the most effective model for this task, ensuring both accuracy and computational efficiency.

We used a dataset of MRI brain images categorized into glioma, meningioma, pituitary, no tumor. As shown in Fig. 3, we split the dataset into 80% training set and 20% testing set, where the images were resized to  $240 \times 240$  pixels, for uniformity. Images were normalized to a range of 0 to 1 for convergence improvement. In order to reduce overfitting and enhance generalization of our model, we included data augmentation involving random rotation (up to  $20^\circ$ ), random width and height shifts (up to 20%), shear transformation (up to 20%), and zooming (up to 20%) involving vertical and horizontal flipping following similar strategies as used in [45]. The results

**Fig. 3** Schematic architecture diagram for EfficientNetB4



demonstrated a validation accuracy of 99.76%, consistent with findings from prior studies using EfficientNet variants [46] (Fig. 2).

The base model used EfficientNetB4 pretrained on ImageNet by Deng et al. [47] for its scale efficiency and high performance. Custom layers included global average pooling to reduce feature map dimensionality, a dense layer with 1024 units and ReLU activation function for higher-level representation. We also included L2 Regularization to mitigate overfitting and a final dense layer with 4 units and softmax activation for multi-class classification.

For model training, categorical cross-entropy was used as the loss function for multi-class classification, and the Adam optimizer with a learning rate of  $10^{-3}$  was used. Early stopping mechanism was implemented for halting the training if validation loss did not improve for 5 epochs, model checkpointing was used to save the best model based on validation accuracy, and the learning rate was reduced by a factor of 0.2 if the validation loss did not improve for 3 epochs.

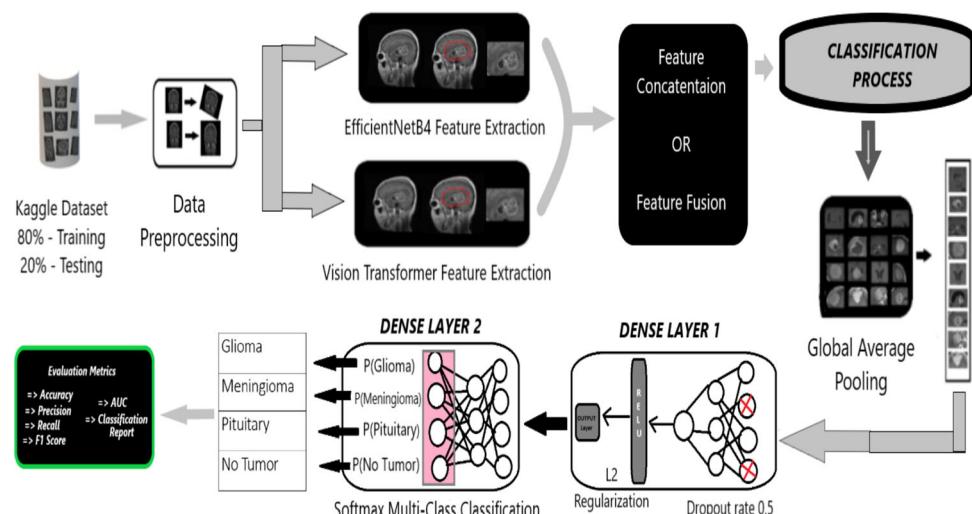
Through these meticulous steps as shown in Fig. 3, the model was refined to achieve remarkable accuracy, demonstrating its effectiveness in classifying brain tumors. Detailed performance metrics and comparative analysis are discussed in the results and discussion section, where the model's superior generalization capabilities are highlighted.

### 3.3.2 Vision Transformers + EfficientNetB4

The Vision Transformer (ViT) combined with EfficientNetB4 architecture leverages the strengths of both models to handle very high-dimensional image data and achieve robust performance. We used the same MRI brain images dataset described earlier, categorized into glioma, meningioma, pituitary, and no tumor. Images were preprocessed to  $224 \times 224$  pixels [48]. Images were normalized by rescaling them to  $(1/255)$ , and data augmentation techniques were applied to the training set to enhance model generalization and reduce overfitting. The techniques include random rotation (up to  $20^\circ$ ), random width and height shifts (up to 20%), shear transformation (up to 20%), zooming (up to 20%), horizontal and vertical flipping, and brightness adjustments (range 0.8–1.2).

The base model utilized the ViT B-32 architecture, pretrained on ImageNet, for its ability to capture long-range dependencies in image data [49]. EfficientNetB4 was integrated for its efficiency and superior performance [50]. Custom layers included global average pooling to reduce the dimensionality of the ViT output. A dense layer with 1024 units and ReLU activation captured higher-level representations. Dropout rate of 0.5 and L2 regularization was used to prevent overfitting. The final dense layer included 4 units (representing brain tumor categories) with softmax activation for multi-class classification as in Fig. 4.

**Fig. 4** Schematic architecture diagram for hybrid model



As per model training, categorical cross-entropy was used as the loss function for multi-class classification, and the Adam optimizer with a learning rate of  $10^{-4}$  was used. Early stopping was implemented for halting the training if the validation loss did not improve for 5 epochs. Model checkpointing was used to save the best model based on validation accuracy, and the learning rate was reduced by a factor of 0.2 if validation loss did not improve for next 3 epochs. The model was trained for a total of 50 epochs with a batch size of 32.

### 3.3.3 Custom 2D CNN

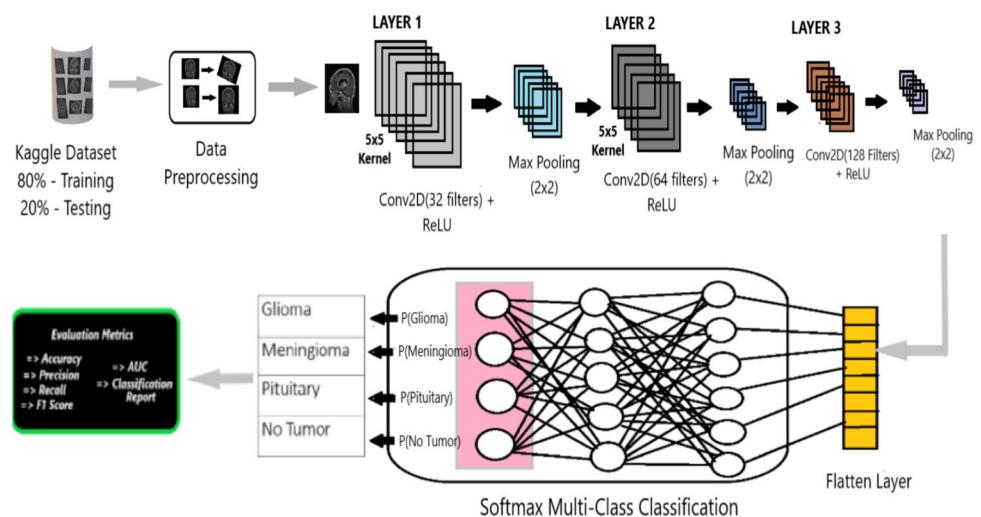
For the Custom 2D CNN model, we developed a unique architecture specifically tailored for brain tumor classification. The same MRI brain images dataset described earlier, categorized into glioma, meningioma, pituitary, and no tumor, was used. Images were preprocessed to a uniform size of  $224 \times 224$  pixels. Normalization (rescaling by 1/255) was applied to all the images, and data augmentation techniques were utilized to enhance model generalization and mitigate overfitting. These augmentation techniques included random rotation (up to  $20^\circ$ ), width and height shifts (up to 20%), shear transformation (up to 20%), zooming (up to 20%), and horizontal and vertical flipping [14].

The base model was a custom-designed 2D convolutional neural network. The architecture included the following custom layers: the first convolution layer with 32 filters and  $5 \times 5$  kernel size followed by a ReLU activation function, a  $2 \times 2$  max-pooling layer for down-sampling, a second convolutional layer with 64 filters and a  $5 \times 5$  kernel size followed by ReLU activation function, another  $2 \times 2$  max-pooling layer, a third convolutional layer with 128 filters and a  $5 \times 5$  kernel size followed by ReLU activation and a final  $2 \times 2$  max-pooling layer. A flatten layer converted the 3D output to a 1D array, followed by a fully connected layer with 128 units and a sigmoid activation function to capture complex representations. The output layer consisted of 4 units with softmax activation for multi-class classification as in Fig. 5 [51].

The model was trained using categorical cross-entropy as the loss function for multi-class classification and the Adam optimizer with a learning rate of  $10^{-3}$ . Training included early stopping to prevent overfitting if validation loss did not improve 5 epochs, model checkpointing to save the best model based on validation accuracy, and learning rate reduction by a factor of 0.2 if validation loss did not improve for 3 epochs. The training process lasted for 100 epochs with a batch size of 32.

The detailed performance metrics and comparative analysis of this model, including its final test and validation accuracies, are discussed in the results and discussion section.

**Fig. 5** Schematic architecture diagram for 2D CNN



### 3.4 Confusion matrix

The system model's effectiveness was evaluated using confusion matrix, which categorize accurate and erroneous prognostications into four distinct classifications.

True positive (TP) occurs when both the predicted and actual outcomes are positive.

False positive (FP) occurs when a forecast prediction predicts a positive outcome, but the actual outcome is negative.

True negative (TN) occurs when both the observed outcome and prognostications are negative.

False negative (FN) occurs when a prediction incorrectly predicts a negative outcome, despite the actual result being positive (Fig. 6).

### 3.5 Performance metrics

Evaluation metrics should be consistently applied, utilizing the system's all open elements to assess the viability of brain tumor discovery

$$\text{Accuracy(ACC)} = \frac{\text{True Positive (TP)} + \text{True Negative (TN)}}{\text{True Positive (TP)} + \text{True Negative (TN)} + \text{False Positive (FP)} + \text{False Negative (FN)}} \quad (1)$$

$$\text{Specificity(SPC)} = \frac{\text{True Negatives (TN)}}{\text{True Negatives (TN)} + \text{False Positives (FP)}} \quad (2)$$

$$\text{Sensitivity(SEN) or Recall(REC)} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Negatives (FN)}} \quad (3)$$

$$\text{Precision(PREC)} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Positives (FP)}} \quad (4)$$

$$\text{F1 - Score} = 2 \times \frac{\text{Precision (PREC)} \times \text{Sensitivity (SEN)}}{\text{Precision (PREC)} + \text{Sensitivity (SEN)}} \times 100\% \quad (5)$$

**Fig. 6** Confusion metrics  
[52]

		Actual Values	
		Positive	Negative
Predicted Values	Positive	TP	FP
	Negative	FN	TN

$$\text{Area under the Cuurve (AUC)} = \int_0^1 \text{Sensitivity (SEN)} * d(1 - \text{Specificity (SPC)}) \quad (6)$$

## 4 Results and discussion

This section demonstrates the experimental results of the above three approaches. Additional tests such as F1-score, recall, precision, and accuracy using Eq. (5), (3), (4) and (1), respectively, are utilized to ascertain the experimental outcomes. Furthermore, the AUC score per class for all three models proposed is calculated using Eq. (6). Let's examine each methodology and their validation accuracies.

### 4.1 EfficientNetb4

The EfficientNetB4 model demonstrated remarkable performance in classifying MRI brain images into glioma, meningioma, pituitary, and no tumor categories. The model's architecture, leveraging efficient scaling and high performance even with fewer parameters as shown in Table 4, proved to be highly effective. The preprocessing steps, including normalization and data augmentation, played a crucial role in enhancing the model's generalization capabilities. By applying various augmentation techniques such as random rotation, width and height shifts, zooming, and flipping, we ensured the model could handle diverse real-world variation in the data.

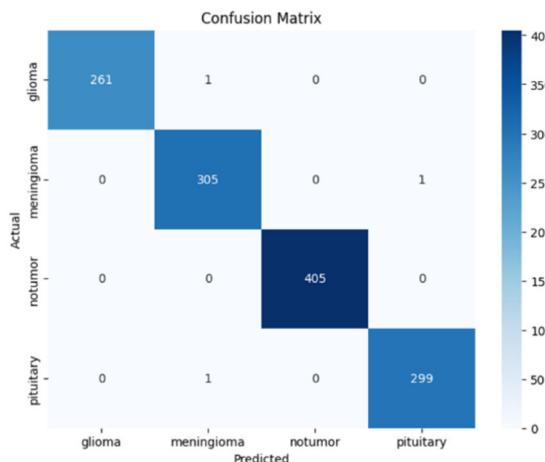
During the evaluation phase, the EfficientNetB4 model achieved significant accuracy and generalization. The detailed performance metrics highlighting the precision, recall, F1-score across all four categories are given in Table 2 and Fig. 7. As we can clearly see in Fig. 7a, the confusion matrix shows minimal mis-predictions by our model. Figure 7c and d shows that the ROC curve and precision vs recall curve is predominantly toward the top left corner and top right corner, respectively, demonstrating strong model performance. Looking at the training vs validation accuracy, even though we observed initial overfitting issues for earlier epochs, the model generalizes pretty well to real-world data.

The model consistently maintained high scores in these metrics, indicating its robustness and reliability in classification tasks. For instance, the precision and recall for glioma, meningioma, no tumor, and pituitary categories were nearly perfect, showcasing the model's ability to correctly classify each type with minimal errors as shown in Table 2.

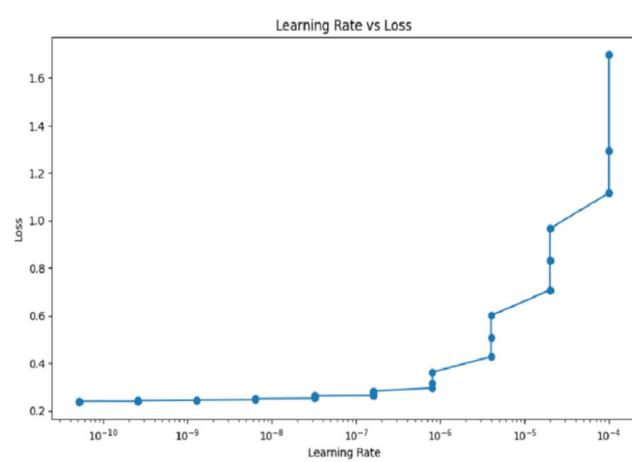
When compared with the Vision Transformer + EfficientNetB4 and Custom 2D CNN models, EfficientNetB4 stood out for its balanced performance across all classes (Table 3). While the hybrid model also showed strong performance, EfficientNetB4's simpler architecture and fewer parameters made it more efficient in terms of computational resources as shown in Table 4. The custom 2D CNN, although specifically designed for this task, did not achieve the same level of accuracy as EfficientNetB4, highlighting the latter's superior capabilities in handling complex image classification tasks.

**Table 2** Classification report for EfficientNetB4 model

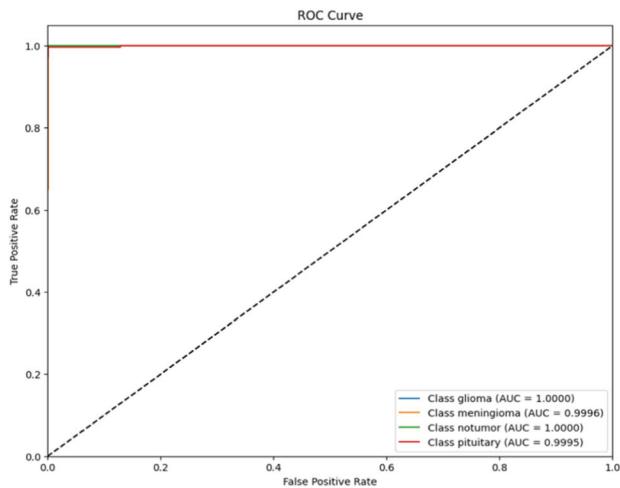
	Accuracy	Precision	Recall	F1-Score	Support
Glioma	0.99	1.00	1.00	1.00	262
Meningioma	0.99	0.99	1.00	1.00	306
No Tumor	1.00	1.00	1.00	1.00	405
Pituitary	0.99	1.00	1.00	1.00	300
Macro-Average	0.99	1.00	1.00	1.00	1273
Weighted Average	0.99	1.00	1.00	1.00	1273



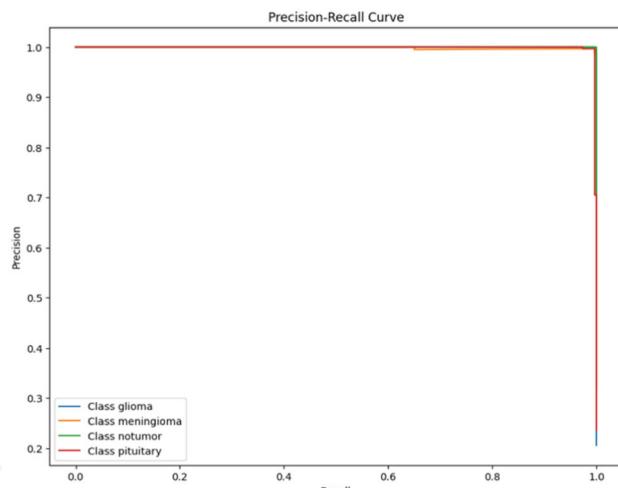
(a)



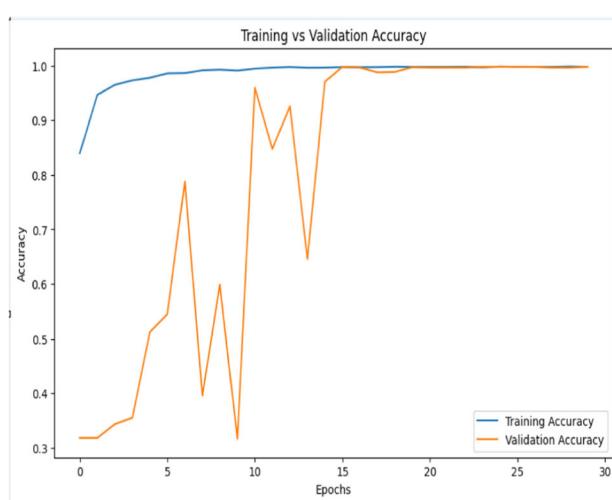
(b)



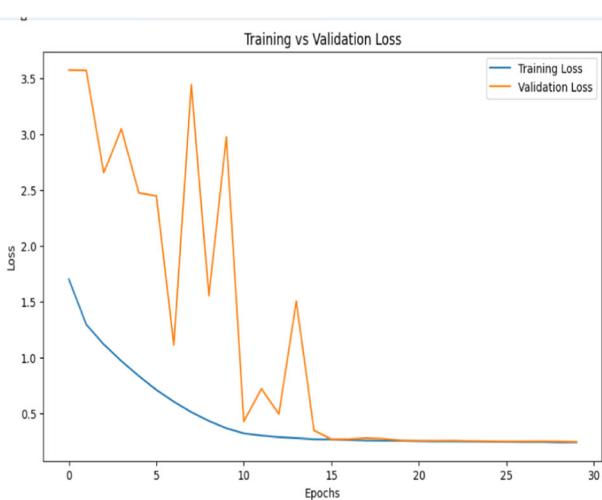
(c)



(d)



(e)



(f)

◀ Fig. 7 EfficientNetB4 performance analysis **a** Confusion matrix, **b** learning rate vs loss, **c** ROC curve, **d** precision–recall curve, **e** training vs validation accuracy, **f** training vs validation loss

**Table 3** AUC score per class

Class	AUC score
Glioma	1.000
Meningioma	0.9996
No Tumor	1.000
Pituitary	0.9995

**Table 4** Parameters of EfficientNetB4 model

Parameters type	No. of parameters	Total size
Total params	19,513,955	74.44 MB
Trainable params	19,388,748	73.96 MB
Non-trainable params	125,207	489.09 KB

#### 4.1.1 Training journey

During the initial training phase, we set the hyperparameters as follows: rotation range (40), width and height shift range (0.4), shear range (0.3), and zoom range (0.4), including vertical flipping. The learning rate was set to  $10^{-5}$  and the number of epochs to 10. Our model achieved a validation accuracy of 54.83% with a validation loss of 1.0321 after the 3rd epoch.

To improve this, we increased the training epochs to 30 and removed vertical flipping, as we thought it was irrelevant or potentially detrimental to the dataset. These changes helped in fine-tuning the model further, minimizing overfitting and enhancing generalization. The model achieved a validation accuracy of 87.75% with a validation loss of 0.3528 during the 8th epoch.

For further improvement, we included horizontal flipping and implemented checkpoints to save the model with the best validation accuracy, reducing the learning rate to  $10^{-4}$ . This led to a model validation accuracy of 94.03% with a validation loss of 0.1809 during the 13th epoch.

We further fine-tuned the model by including a learning rate reduction mechanism, reducing the learning rate by a factor of 0.2 if the validation accuracy did not improve for 3 epochs, with a lower limit of  $10^{-6}$ . This prevented the model from oscillating too much and allowed it to converge smoothly by progressively taking smaller steps in the learning process as shown in Fig. 7b. Ultimately, we achieved a significant milestone: a validation accuracy of 99.76% with a validation loss of 0.2636 and a training accuracy of 99.73% with a training loss of 0.2657 as shown in Fig. 7e and f.

**Table 5** Classification report for hybrid model

	Accuracy	Precision	Recall	F1-Score	Support
Glioma	0.99	1.00	1.00	1.00	262
Meningioma	0.99	1.00	0.99	1.00	306
No Tumor	1.00	1.00	1.00	1.00	405
Pituitary	0.99	0.99	0.99	0.99	300
Macro-Average	0.99	1.00	1.00	1.00	1273
Weighted Average	0.99	1.00	1.00	1.00	1273

## 4.2 Hybrid model (Vision Transformers + EfficientNetb4)

The Vision Transformer + EfficientNetB4 hybrid model demonstrated a remarkable performance in classifying MRI brain images into glioma, meningioma, pituitary, and no tumor categories. By combining the strengths of Vision Transformers with the efficient scaling of EfficientNetB4, this hybrid approach proved highly effective.

Preprocessing steps include normalization and various data augmentation techniques, such as random rotation, width and height shifts, and flipping. These steps enhanced the model's generalization capabilities by ensuring it could handle diverse real-world variation in the data.

The detailed performance metrics highlighting the precision, recall, and F1-score across all four categories are presented in Table 5 and Fig. 9. In Fig. 9a, the confusion matrix for hybrid model shows us very minimal misclassification highlighting its capability for generalization. Figure 9c and d presents the ROC curve and precision–recall curve are again predominantly toward the top left and top right corner which shows strong model performance. Looking at the training vs validation loss graph in Fig. 9f, the close alignment of both the graphs indicates how well the model performs and while mitigating overfitting issues.

In comparison with other models, the hybrid model stood out for its enhanced performance across all metrics, particularly due to the combined strengths of the Vision Transformers and EfficientNetB4. It achieved an excellent accuracy with a balanced trade-off between computational resources and performance as shown in Tables 6 and 7. Its superior feature extraction and robust preprocessing helped in handling diverse real-world variation efficiently.

### 4.2.1 Training journey

During the initial phase of training, we utilized the same hyperparameters as our previous EfficientNetB4 model. However, the results were unsatisfactory. After 30 epochs, the test accuracy plateaued at approximately 86.174%, while the test loss remained high at 10.165. These metrics indicate suboptimal performance and are deemed unacceptable for our objectives.

We next set the hyperparameters as follows: rotation range (20), width and height shifts (0.2), shear range (0.2), zoom range (0.2), and included horizontal flipping. We first tested our model for 30 epochs, achieving a validation accuracy of 98.98% with a validation loss of 0.2108 and a learning rate of  $10^{-4}$  during the 25th epoch.

Further fine-tuning involved including a brightness range [0.8–1.2] in data augmentation. This adjustment improved the validation accuracy to 99.06% during the 27th epoch, with a validation loss of 0.4981 and a learning rate of  $2.00 \times 10^{-5}$ .

Curious to explore other architectures, we implemented EfficientNetB0 [53] instead of EfficientNetB4. This yielded a validation accuracy of around 99.29% and a validation loss of 0.5677 with a learning rate of  $10^{-6}$  during the 21st epoch. However, due to significant overfitting in the initial epochs as shown in Fig. 8, we reverted to the EfficientNetB4 architecture.

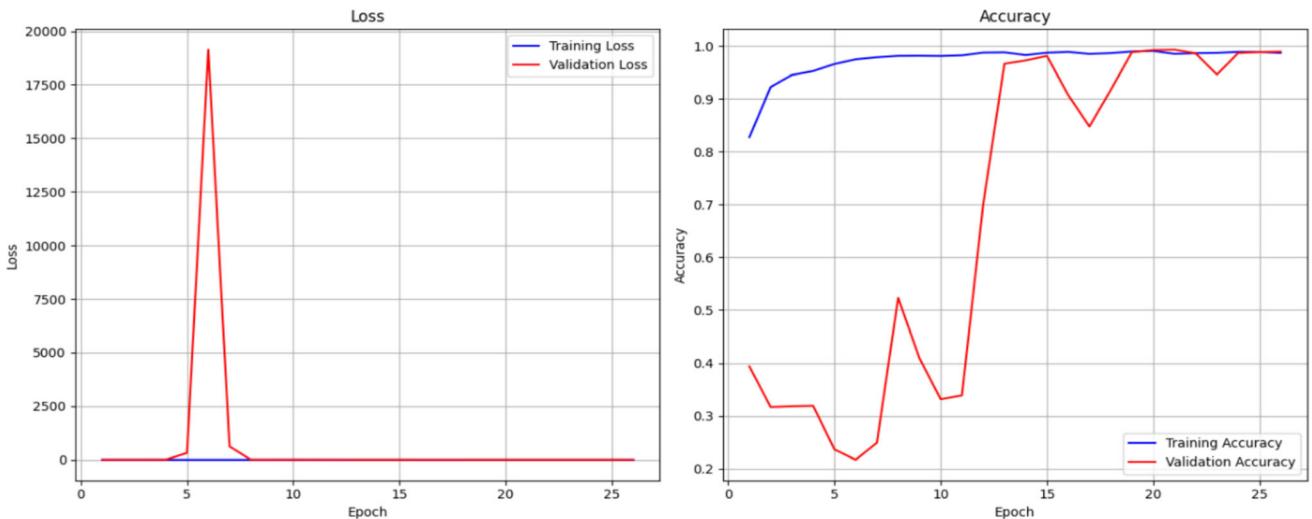
Increasing the training epochs to 50 and employing techniques such as reduced learning rate mechanism, early stopping, and checkpoints, our model achieved a validation accuracy of 99.6% with a minimal validation loss of 0.034. The training accuracy was 99.63%, with a training loss of 0.0236. Among all the models, the hybrid model exhibited the lowest validation loss as shown in Fig. 9e and f and in Fig. 16.

**Table 6** AUC score per class

Class	AUC Score
Glioma	1.000
Meningioma	0.9997
No Tumor	1.000
Pituitary	0.9998

**Table 7** Parameters of hybrid model

Parameters type	No. of parameters	Total size
Total params	108,766,283	415.02 MB
Trainable params	108,639,028	414.53 MB
Non-trainable params	127,255	501.76 KB

**Fig. 8** Training vs validation loss and accuracy of Vision Transformer + EfficientNetB0 architecture

#### 4.3 Custom 2D CNN

The custom 2D CNN model demonstrated significant potential in classifying MRI brain images into glioma, meningioma, pituitary, and no tumor categories. The architecture, tailored specifically for this task, focused on extracting features pertinent to brain tumor classification.

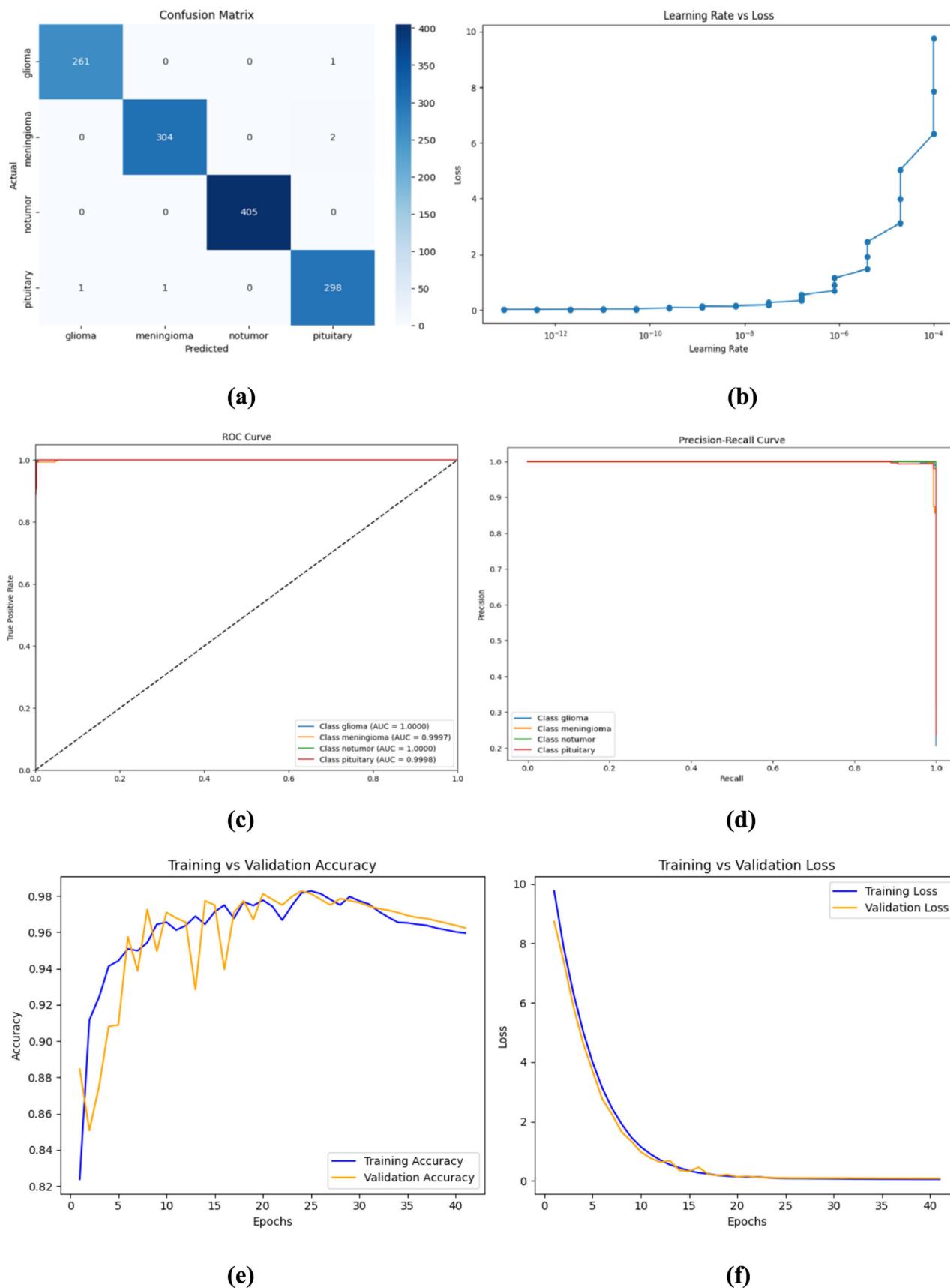
The preprocessing steps included normalization and data augmentation techniques like random rotation, width and height shifts, zooming, and horizontal flipping. These steps helped in enhancing the model's ability to generalize by training on diverse variations of the data.

Despite being a custom architecture, the 2D CNN model achieved competitive performance. The detailed performance metrics, including precision, recall, and F1-score for each category, are presented in Table 8 and Fig. 10.

While the custom 2D CNN showed strong classification capabilities, it did not outperform the Vision Transformer + EfficientNetB4 hybrid model or EfficientNetB4 model. The validation accuracy and loss metrics, as detailed in Fig. 10e and f, indicated that the custom 2D CNN model had room for improvement in terms of handling overfitting and achieving higher generalization (Table 9).

The model achieved a validation accuracy of 97.25% with a validation loss of 0.0838. Training accuracy was 95.83% with a training loss of 0.1155 as shown in Fig. 10e and f. The precision and recall for the glioma, meningioma, no tumor, and pituitary categories were respectable, but slightly lower compared to the other models.

Overall, while the custom 2D CNN model provided a robust baseline and demonstrated the effectiveness of a tailored architecture as shown in Table 10 [54], the comparative performance highlighted the superior capabilities of the hybrid and EfficientNetB4 models in handling complex image classification tasks.



◀ Fig. 9 Vision Transformer + EfficientNetB4 (hybrid model) performance analysis **a** confusion matrix, **b** learning rate vs loss, **c** ROC curve, **d** precision–recall curve, **e** training vs validation accuracy, **f** training vs validation loss

**Table 8** Classification report for custom 2D CNN

	Accuracy	Precision	Recall	F1-Score	Support
Glioma	0.91	1.00	0.92	0.96	262
Meningioma	0.95	0.93	0.95	0.94	306
No Tumor	1.00	0.97	1.00	0.99	405
Pituitary	0.99	0.98	0.99	0.99	300
Macro-Average	0.96	0.97	0.97	0.97	1273
Weighted Average	0.96	0.97	0.97	0.97	1273

#### 4.3.1 Training journey

During the initial training phase, we employed the ReLU activation function and set the kernel size to  $3 \times 3$ . The number of kernels in each layer was configured as follows: 32 kernels in the first layer, 64 in the second layer, and 128 in the third layer. Additionally, our model included one fully connected dense layer, and three max-pooling layers of  $2 \times 2$ . This setup achieved a validation accuracy of 91.59% with a validation loss of 0.2400.

Subsequently, we experimented with different kernel sizes ( $5 \times 5$  and  $7 \times 7$ ), but observed no significant improvement in validation accuracy. Next, we switched the activation function to sigmoid while keeping other hyperparameters unchanged. This adjustment increased the validation accuracy to 92.77%, with a validation loss of 0.1941.

Further attempts to fine-tune the model involved incorporating elastic net regularization. However, this led to a decrease in validation accuracy to 87.04% and an increase in validation loss to 0.6455. Finally, we explored the use of the tanh activation function without any regularization. Training the model for 50 epochs with this configuration resulted in a validation accuracy of 92.22% and a validation loss of 0.1881.

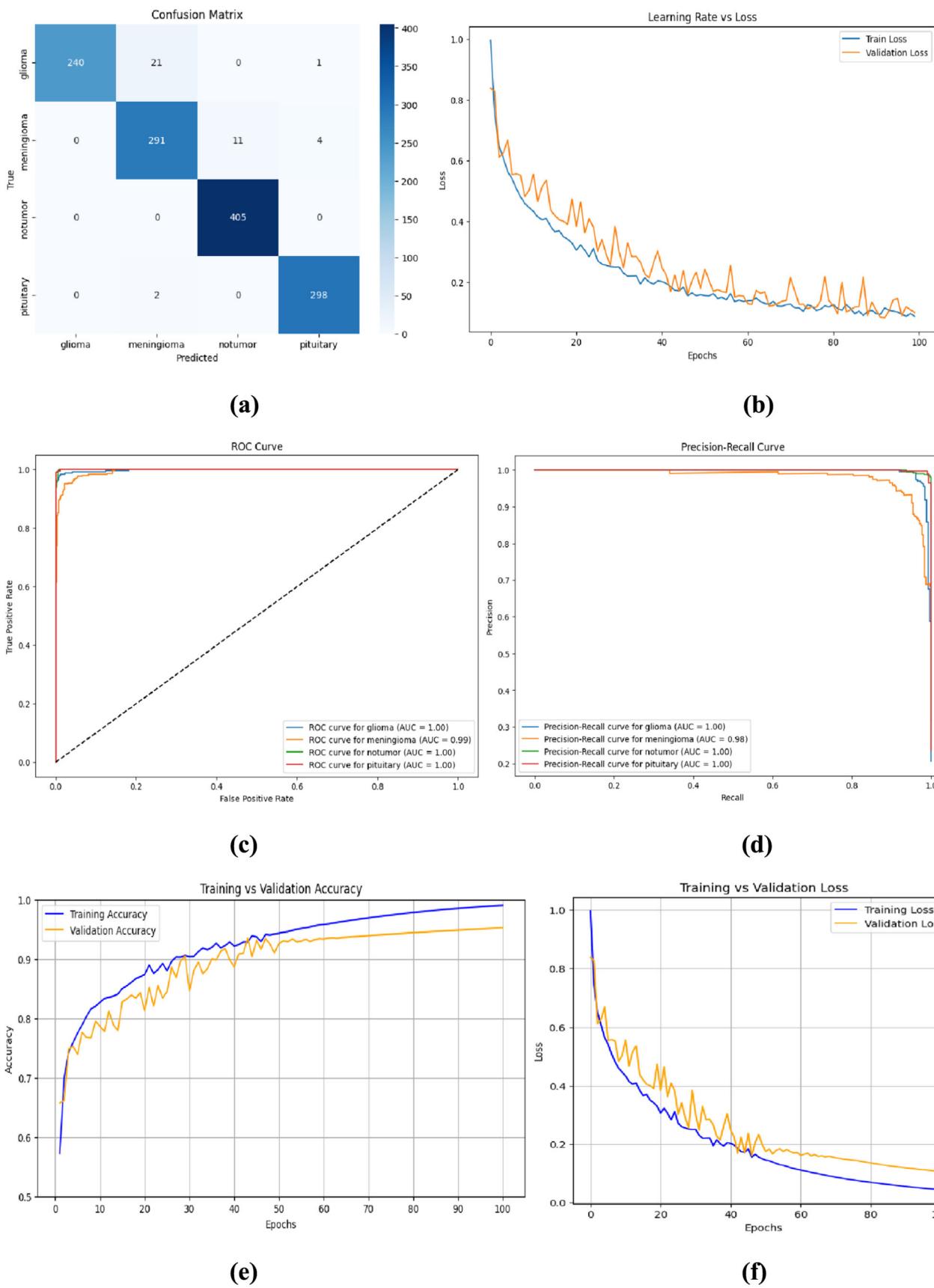
Finally, we trained our model for 100 epochs with early stopping, a reduced learning rate, and checkpoint mechanisms in place, using sigmoid as the fixed activation function. The model achieved a validation accuracy of 97.25% with a validation loss of 0.083, while the training accuracy reached 95.83% with a training loss of 0.1155 as shown in Fig. 10e and f.

Even though ReLU typically outperforms sigmoid in most CNN applications, in this case, sigmoid proved to be more effective. One possible reason is that brain MRI images may have subtle intensity variations, and sigmoid might have acted as a soft gating mechanism, helping the network retain fine-grained details and distinguish tumor types more effectively. Another potential factor is that ReLU suffers from the dying neuron problem, where neurons output zero for negative inputs, leading to vanishing gradients and potentially hindering learning in deeper layers. Similarly, while tanh has a zero-centered output, it still experiences vanishing gradients for large inputs, resulting in slower convergence. In contrast, sigmoid provides smooth gradients, which may have facilitated better convergence in this specific scenario.

Future iterations could include adding batch normalization and dropout layers or even exploring deeper architectures to improve generalization and performance [55].

#### 4.4 Performance comparison

Figure 11a–c illustrates the accuracy for each class (glioma, meningioma, pituitary, and no tumor) across the three models. These graphs highlight the models' proficiency in correctly classifying each tumor type, demonstrating EfficientNetB4's superior accuracy in comparison to the other models.



◀ Fig. 10 Custom 2D CNN performance analysis **a** confusion matrix, **b** learning rate vs loss, **c** ROC curve, **d** precision–recall curve, **e** training vs validation accuracy, **f** training vs validation loss

Table 9 AUC score per class

Class	AUC score
Glioma	0.9983
Meningioma	0.9937
No Tumor	0.9997
Pituitary	0.9998

Table 10 Parameters of custom 2D CNN model

Parameters type	No. of parameters	Total size
Total params	11,334,852	43.26 MB
Trainable params	11,334,852	43.26 MB
Non-trainable params	0	0 KB

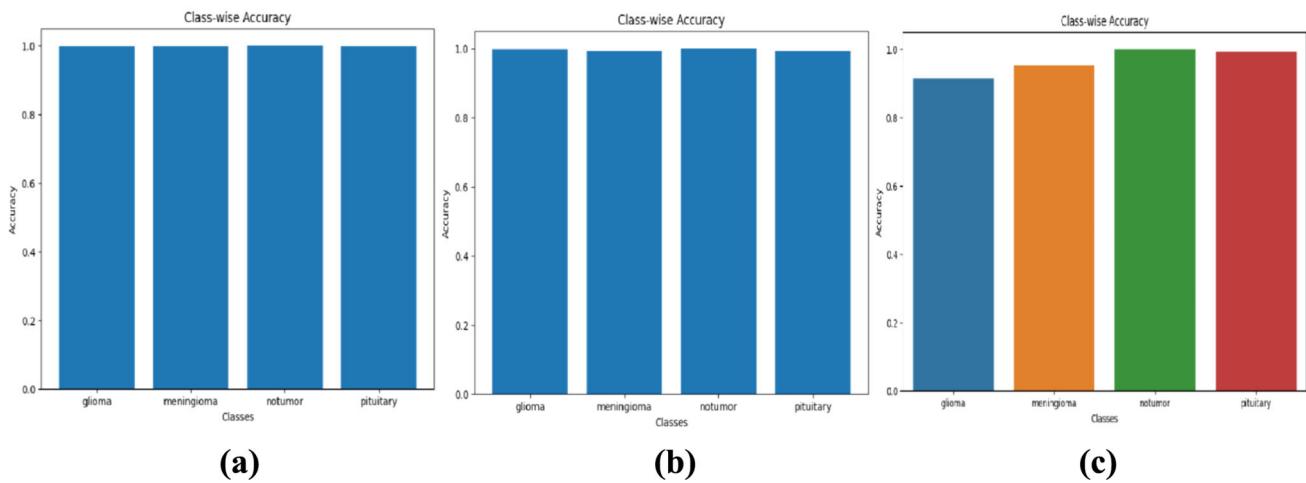


Fig. 11 Class-wise accuracy comparison **a** EfficientNetB4, **b** hybrid model, **c** custom 2D CNN

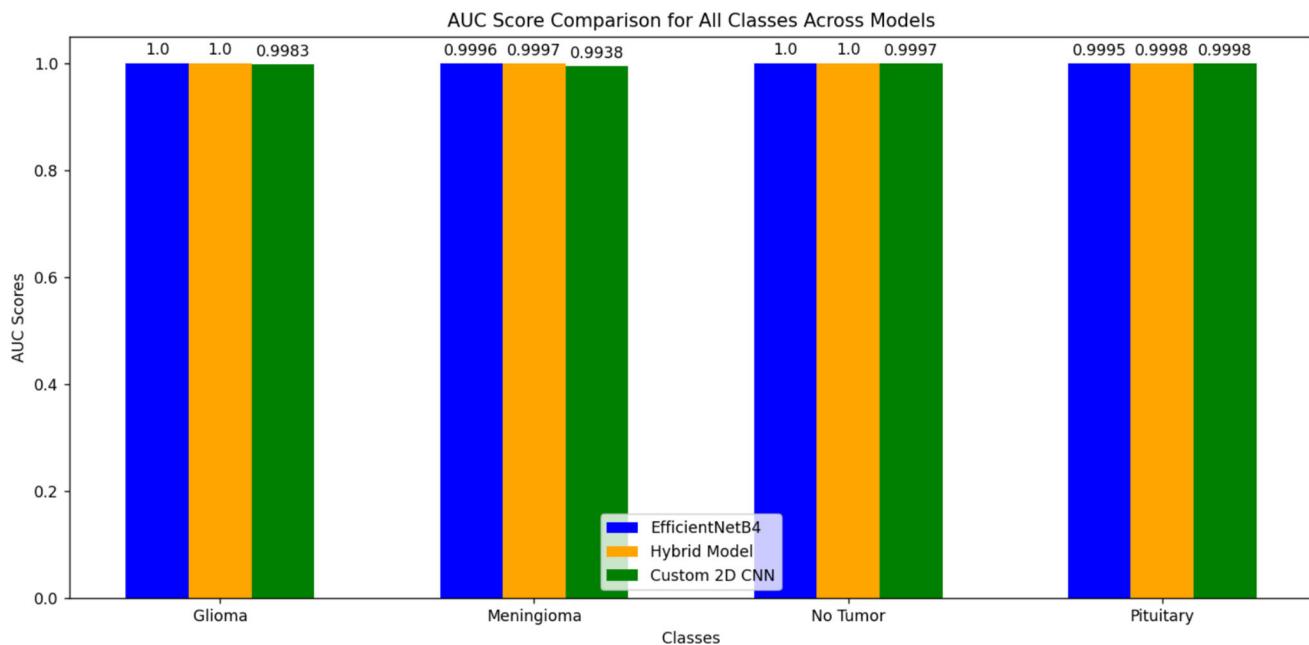
Figure 12 shows AUC scores for each class of brain tumor. This metric provides insight into the models' overall ability to distinguish between different classes, where both EfficientNetB4 and the hybrid model achieved near-perfect AUC scores, indicating excellent classification performance.

In Fig. 13, we compare the training accuracy across models. EfficientNetB4 quickly achieves near-perfect accuracy, followed by the hybrid model with slightly lower but still high accuracy. The custom 2D CNN shows a gradual improvement, achieving competitive accuracy over time.

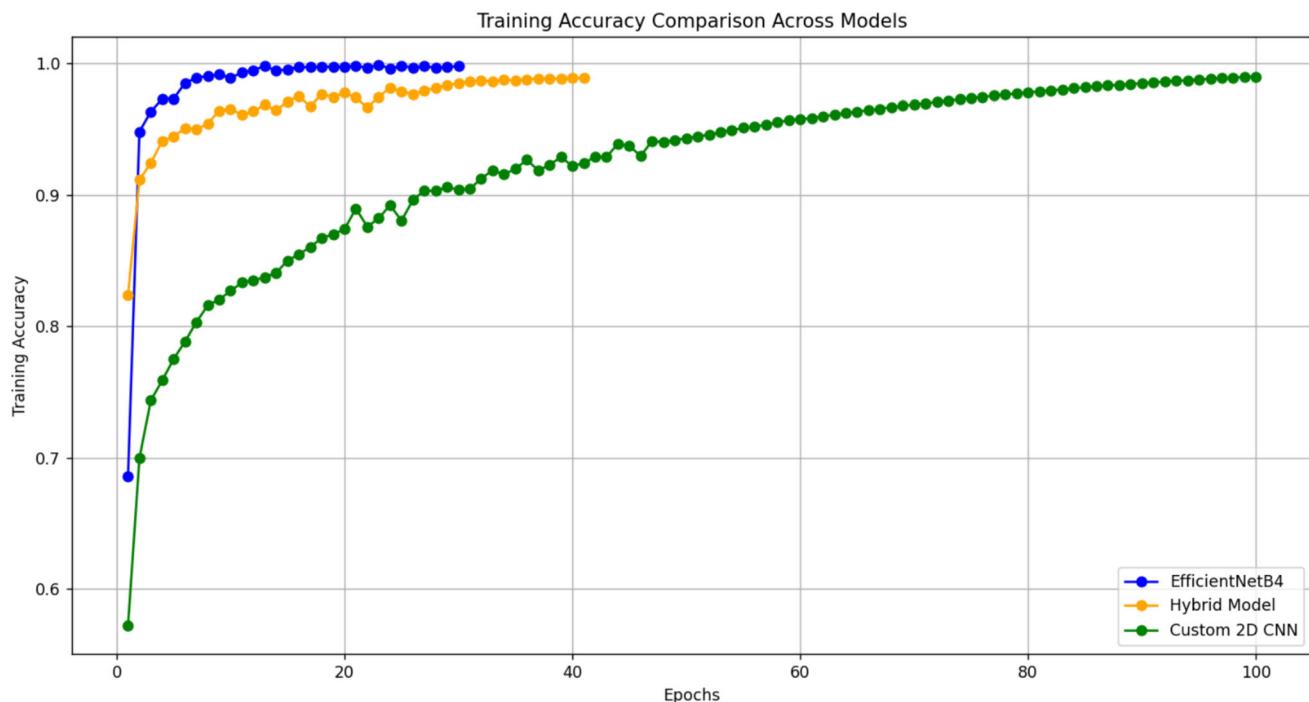
Similarly, in Fig. 14 we compare the validation accuracy across models. The hybrid model shows the highest and most stable validation accuracy, followed by EfficientNetB4, which stabilizes around 0.95, and the custom 2D CNN, which shows steady improvement, stabilizing around 0.9 after 40 epochs.

In Fig. 15, we compare the training loss of each model across the training epochs. EfficientNetB4 and the hybrid model demonstrate a faster convergence rate with a steady decrease in training loss, while the custom 2D CNN exhibits more fluctuations.

In Fig. 16, we compare the validation loss for the three models. Lower validation loss indicates better generalization to new, unseen data. EfficientNetB4 and the hybrid model consistently maintain low validation loss, underscoring their robustness in real-world scenarios.



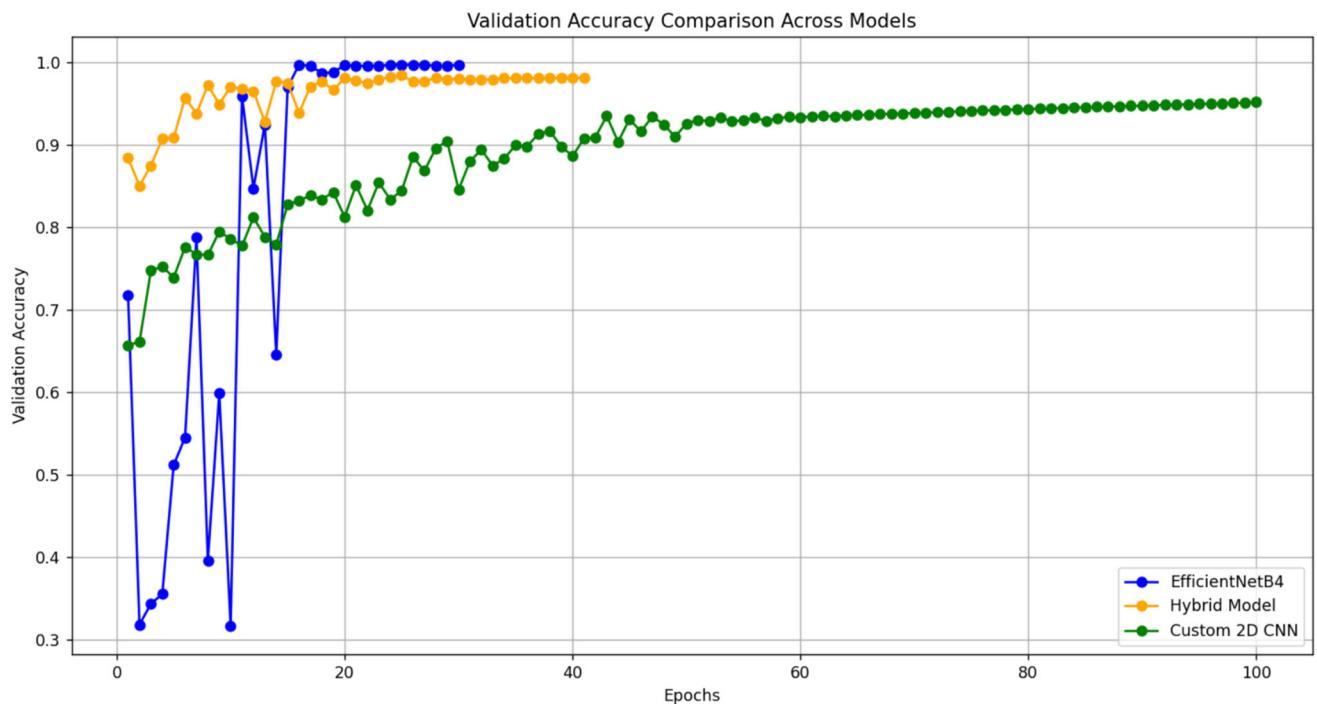
**Fig. 12** Model-based comparison on AUC scores for each class of brain tumor



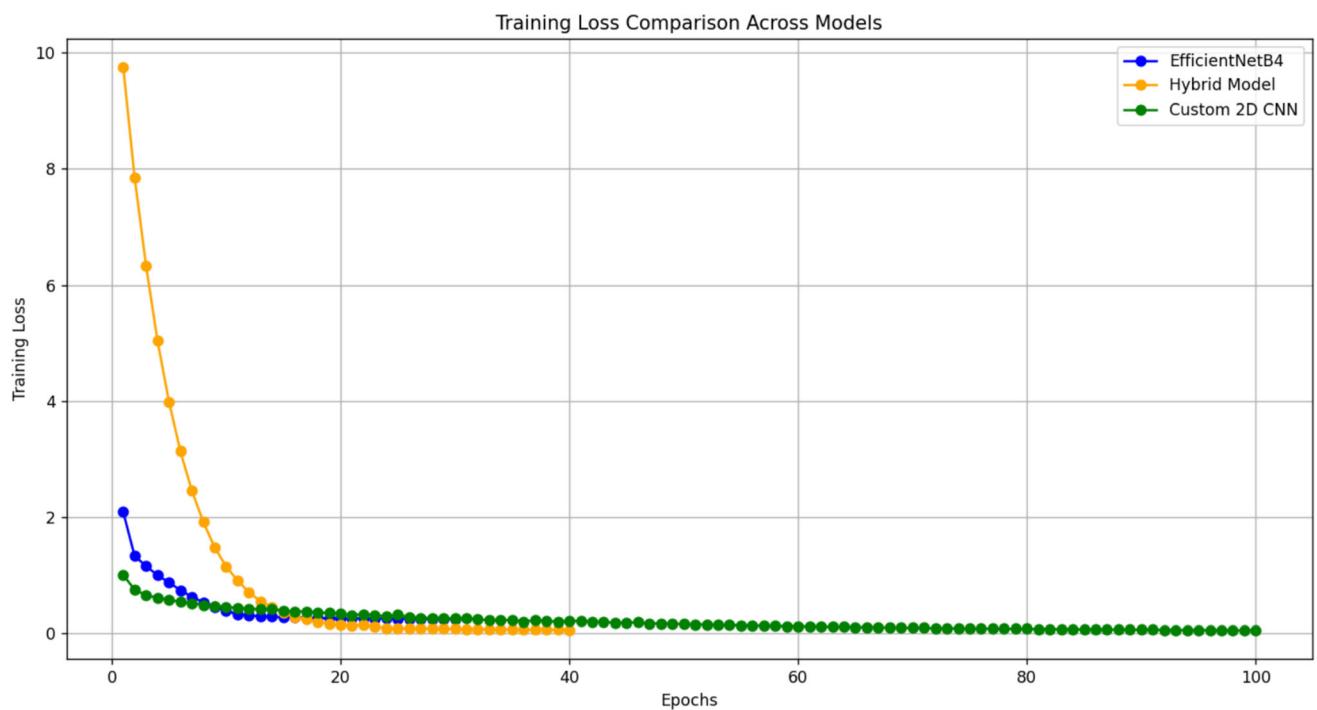
**Fig. 13** Model-based comparison on training accuracy across epochs

In Fig. 17, we present a heatmap of the confusion matrices, providing a visual comparison of misclassifications. EfficientNetB4 shows the least misclassifications, emphasizing its reliability in accurate predictions, followed closely by the hybrid model.

Figure 18 shows the ROC curves for the three models, illustrating their performance. EfficientNetB4 and the hybrid model exhibit near-perfect classification with ROC curves close to the top left reflecting high sensitivity and specificity in classification, while Custom 2D CNN shows strong yet slightly lower performance.

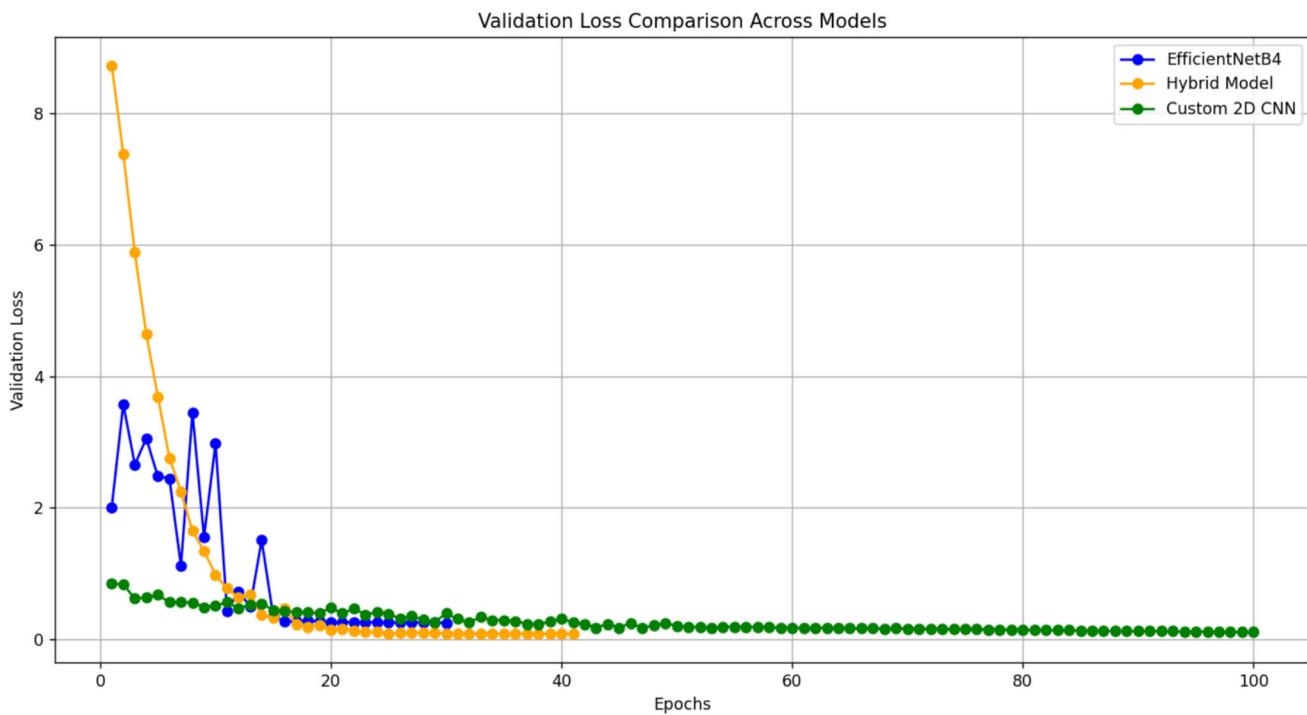


**Fig. 14** Model-based comparison on validation accuracy across epochs

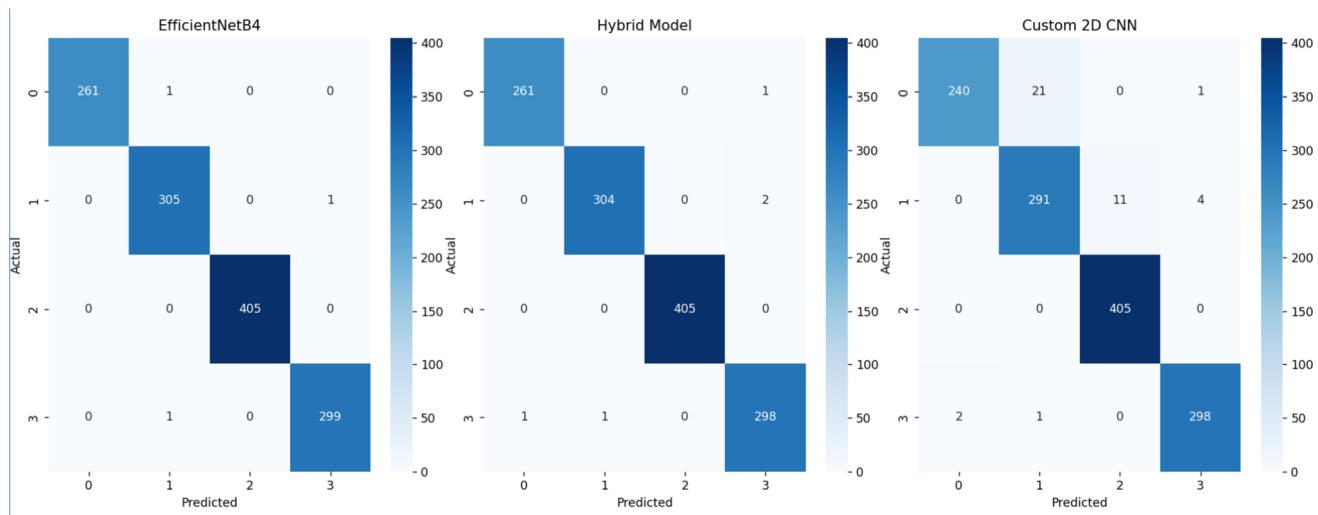


**Fig. 15** Model-based comparison on training loss

Figure 19 provides a comparison of precision, recall, and F1-scores across different classes for the three models. EfficientNetB4 consistently achieves the highest scores, followed by the hybrid model, while the custom 2D CNN shows competitive yet slightly lower performance.



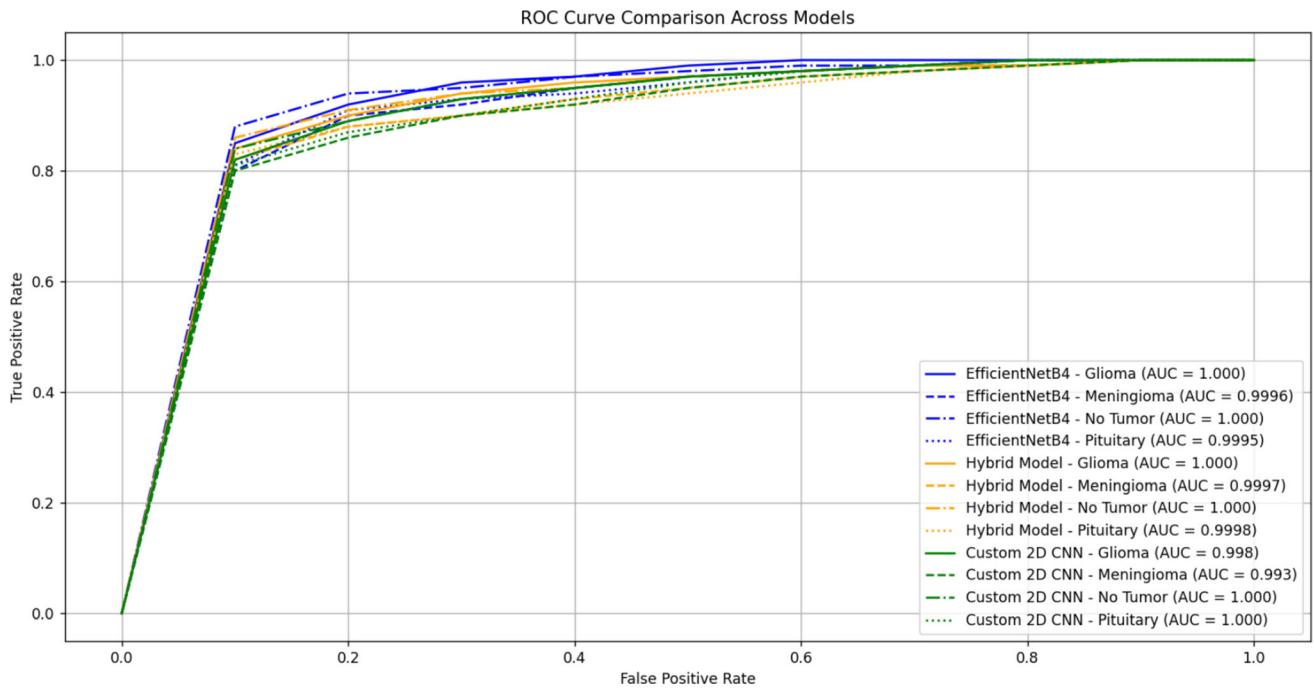
**Fig. 16** Model-based comparison on validation loss



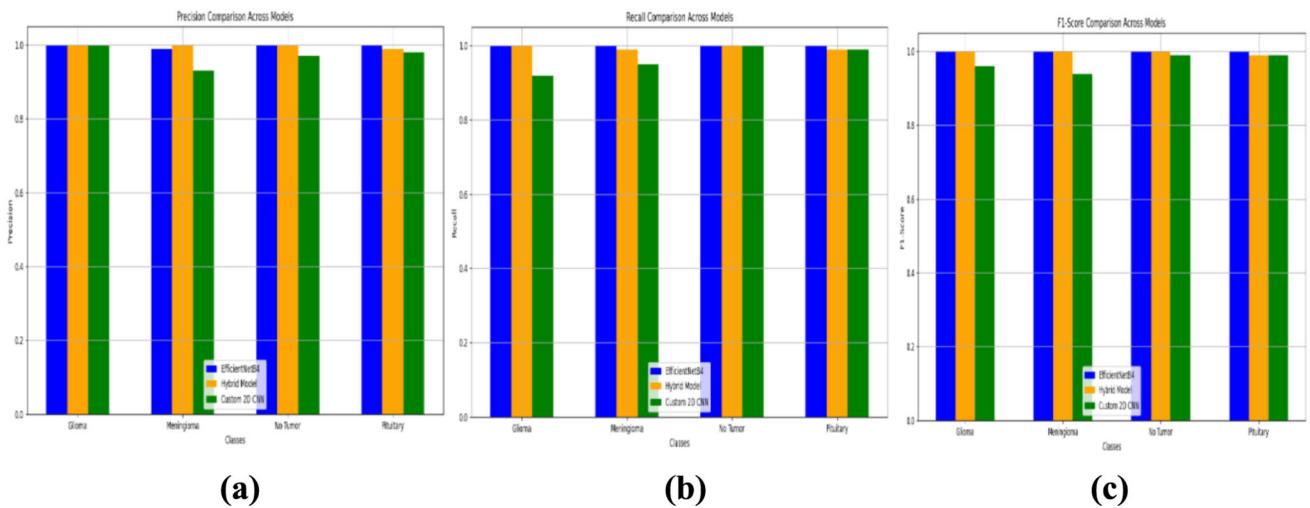
**Fig. 17** Heatmap for model-based comparison on confusion matrix

Figure 20 presents the heatmaps for precision, recall, and F1-score across all three models: (a) EfficientNetB4, (b) hybrid model, and (c) custom 2D CNN. These heatmaps visually represent the classification performance for each tumor category, highlighting how well each model balances precision and recall.

In Fig. 21, we compare the training times across epochs for all models. EfficientNetB4 demonstrates the shortest training times, while the custom 2D CNN has the longest due to a higher number of epochs. The hybrid model falls in between.



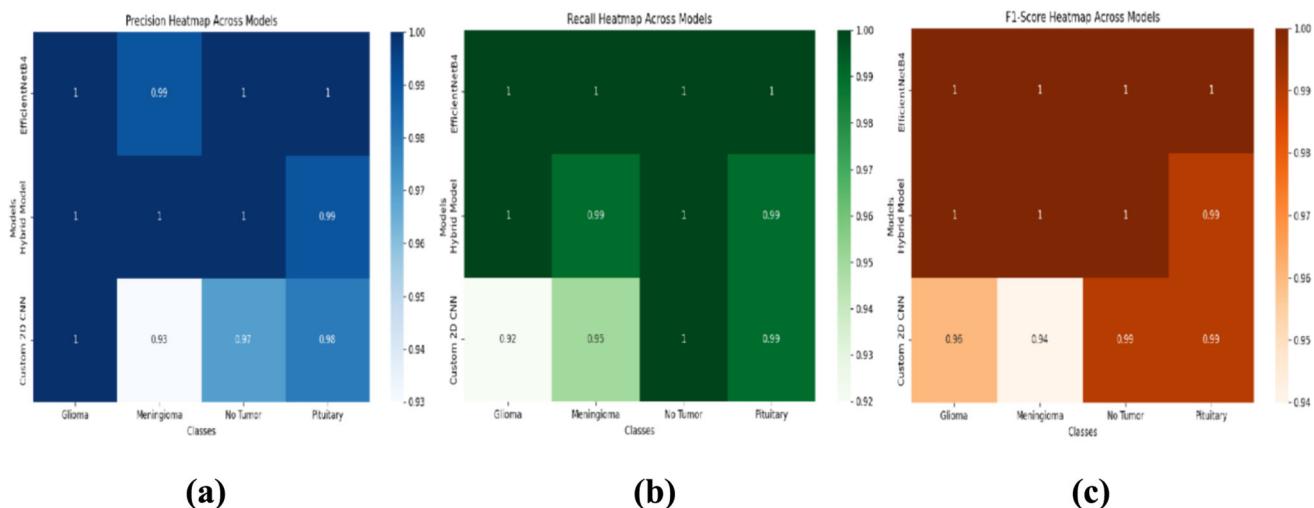
**Fig. 18** ROC comparison on models



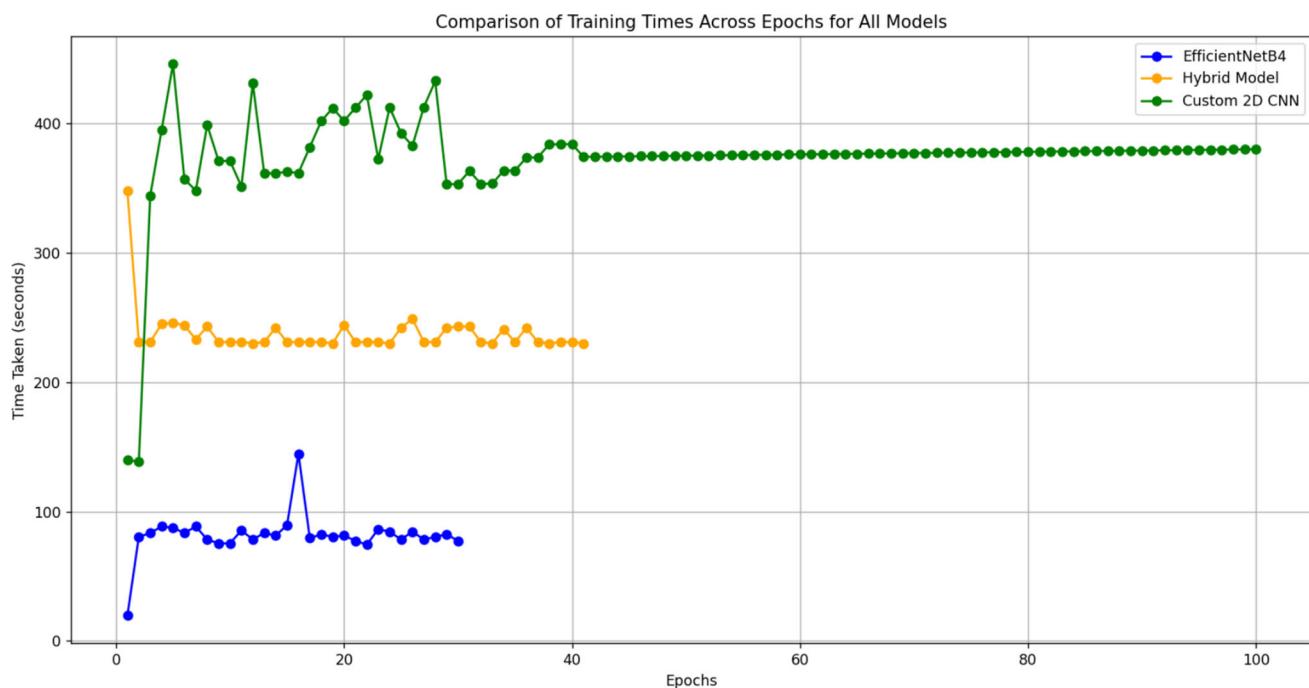
**Fig. 19** Performance metrics comparison: **a** precision, **b** recall **c** F1-score

#### 4.5 Comparison with other machine learning models

Algorithms	Accuracy	Precision	Recall	F1-Score
<b>EfficientNetB4</b>	<b>99.76%</b>	<b>100%</b>	<b>100%</b>	<b>100%</b>
<b>Vision Transformer + EfficientNetB4 (Hybrid Model)</b>	<b>99.6%</b>	<b>100%</b>	<b>100%</b>	<b>100%</b>
<b>2D Convolutional Neural Network (CNN)</b>	<b>97.25%</b>	<b>97%</b>	<b>97%</b>	<b>97%</b>
Support Vector Machines (SVMs) [56]	98.18%	95.50%	97.95%	97.95%
K-means Clustering (KNN) [57]	90%	85%	97%	91%
Random Forests (RF) [58]	83%	85%	93%	88%
Fuzzy C-means (FCM) [59]	98.56%	99.14%	99.25%	99.19%
Multilayer Perceptron (MLP) [60]	89%	99%	90%	94%



**Fig. 20** Heatmaps for precision, recall, and F1-score across all three models: **a** EfficientNetB4, **b** hybrid model, **c** custom 2D CNN



**Fig. 21** Comparison of training times across epochs for all models

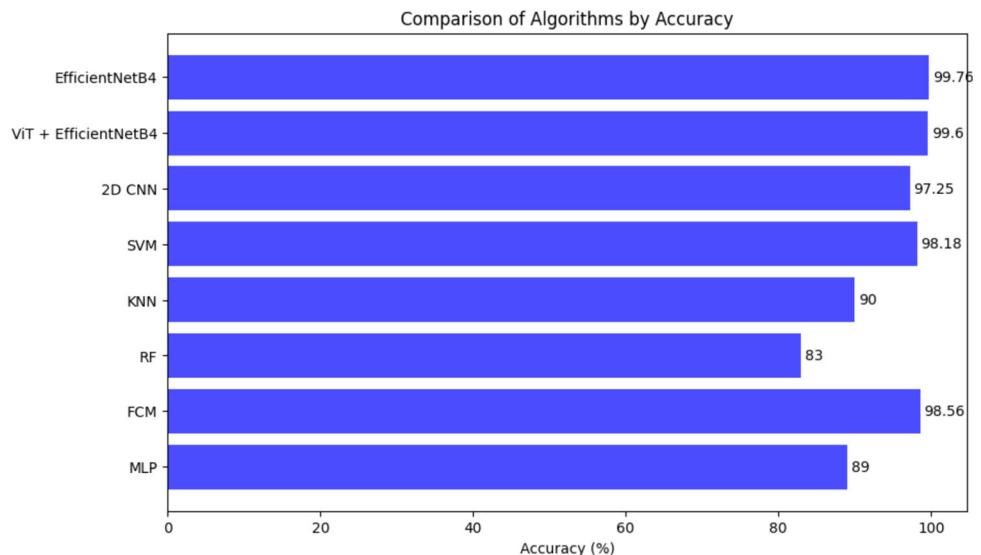
Figure 22 illustrates the accuracy performance of various algorithms, highlighting the EfficientNetB4 and the hybrid model of ViT + EfficientNetB4 as top performers.

Figure 23 demonstrates the precision rates achieved by different algorithms, showing that EfficientNetB4 and the hybrid model consistently achieve 100% precision.

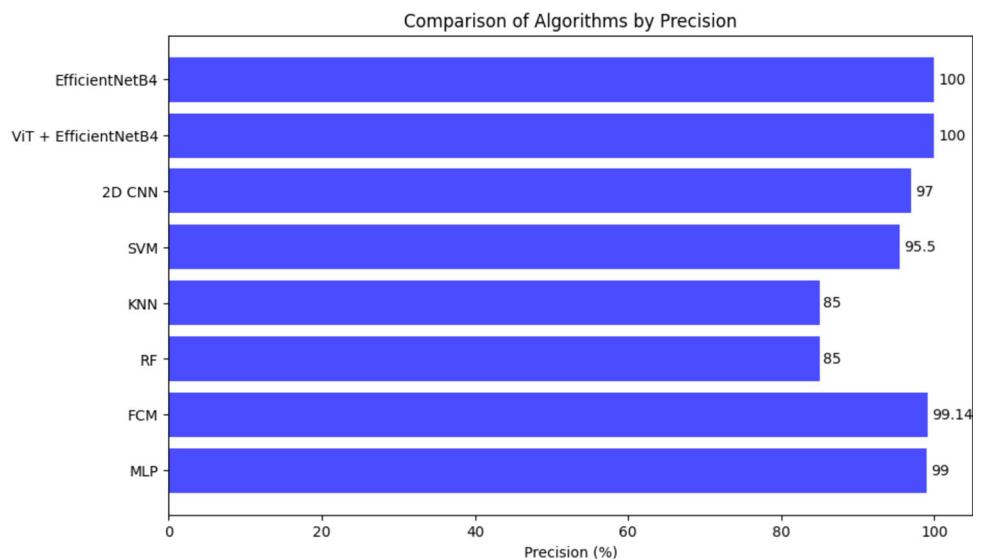
Figure 24 presents a comparison of recall scores among the algorithms, indicating that both EfficientNetB4 and the hybrid model maintain perfect recall.

Figure 25 compares the F1-scores of all tested algorithms, underscoring the superior performance of EfficientNetB4 and the ViT + EfficientNetB4 hybrid model with 100% F1-scores.

**Fig. 22** Comparison of accuracy across algorithms



**Fig. 23** Comparison of precision across algorithms

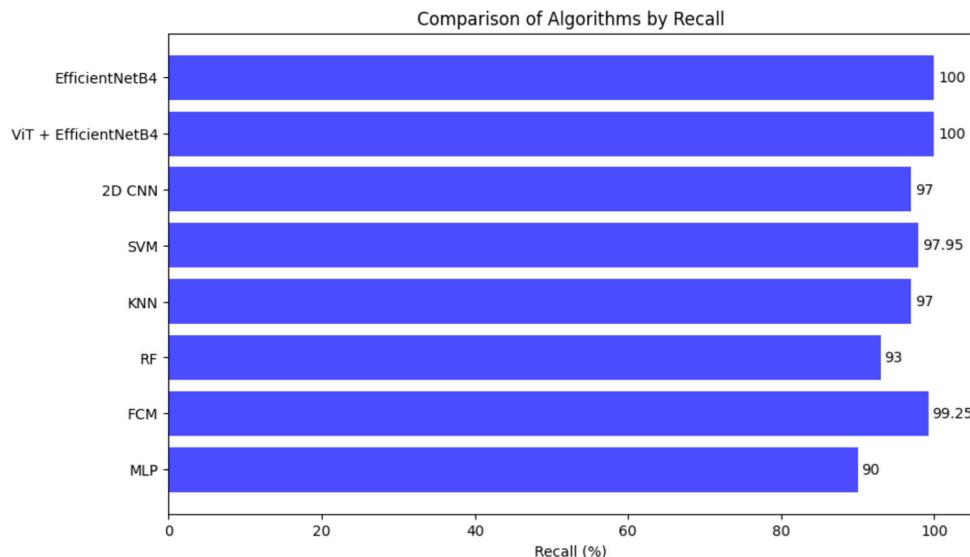


## 5 Limitations and future work

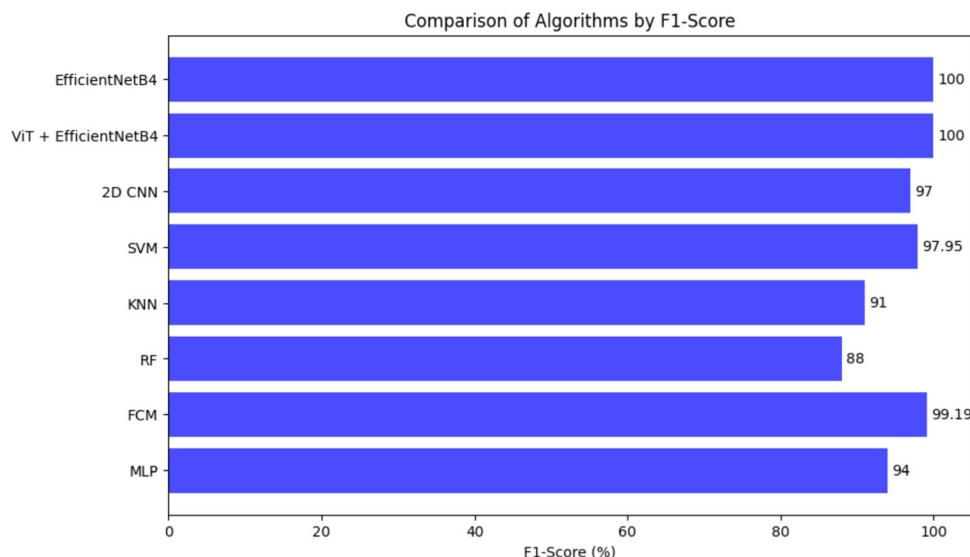
In the proposed future work, we aim to enhance our model's performance and address potential challenges in brain tumor detection and classification. Our current models, such as EfficientNetB4, a hybrid Vision Transformer + EfficientNetB4, and a custom CNN model, yield promising results. However, several improvements can increase their reliability.

The ViT + EfficientNetB4 hybrid model for brain tumor classification presents several challenges and areas for improvement. These include high computational complexity due to the large parameter count (108 M), long training times as shown in Fig. 21, and extensive hardware requirements, which may not be feasible in resource-constrained settings. Additionally, the model's performance may be hindered by limited dataset sizes, domain shift issues across different MRI scanners, and the risk of overfitting. The lack of model interpretability makes it harder for medical practitioners to trust predictions, and the model's robustness to noise, motion artifacts, and distortions in MRI images is a concern. Furthermore, large hybrid models may face challenges in real-time clinical deployment due to slower inference speeds and high energy consumption. To address these issues, future work could focus on optimizing model size through pruning or distillation, improving generalization via data

**Fig. 24** Comparison of recall across algorithms



**Fig. 25** Comparison of F1-score across algorithms



augmentation and domain adaptation, enhancing interpretability with attention visualization, and exploring alternative hybrid architectures like CNN + Swin Transformer for better spatial feature learning [61].

Meanwhile, our other proposed models, like the EfficientNetB4 model, which is more lightweight than transformer-based models, still require substantial GPU memory, making it less suitable for resource-constrained environments such as mobile or edge computing applications. Additionally, its feature extraction capability may be less effective for highly heterogeneous medical datasets, potentially limiting generalization across different MRI scanners and institutions. Similarly, the custom 2D CNN model, despite being specifically designed for this task, exhibited lower accuracy (97.25%) compared to EfficientNetB4 and the hybrid model, suggesting that its shallower architecture might not capture complex tumor features as effectively. The custom 2D CNN also suffered from overfitting, requiring strong regularization techniques, and its training time was significantly longer as clearly shown in Fig. 21 due to a less optimized structure.

Future research will integrate multiple imaging modalities, like CT or PET scans, alongside MRI, to enhance tumor characterization. While our models have performed well on the Kaggle dataset, they have yet to be tested in real-world scenarios. Addressing challenges such as class imbalance, where some tumors appear less frequently in the data, will be crucial. We plan to incorporate techniques like generative adversarial networks (GANs) [62] and advanced data augmentation to create additional examples of rarer tumor types, enhancing the model's ability to recognize them.

Since transparency is essential in medical diagnostics, we plan to use explainable AI tools like Grad-CAM and SHAP. These tools will help build trust in the model's decisions. To make a real difference, our models could be integrated into clinical processes as tools to support doctors in decision-making. Future steps will focus on creating easy-to-use interfaces, enabling real-time processing, and testing the models' usefulness in healthcare settings through pilot studies.

Additionally, we see potential in using MRI sequences to study how tumors change over time, which could allow for predictive insights into tumor growth. This could be achieved by applying techniques such as recurrent neural networks (RNNs) or temporal convolutional networks (TCNs) [63]. These improvements aim to develop a more adaptable, understandable, and effective model, providing healthcare professionals with practical insights that support early diagnosis and more informed treatment planning.

We also plan to integrate multi-modal data, such as genetic and molecular markers, with MRI scans to gain a deeper understanding of each tumor's characteristics, enhancing both classification accuracy and progression predictions [64]. This holistic approach would support personalized treatment planning by providing clinicians with richer context around each case.

To ensure robustness, we plan to validate our models on diverse, multi-institutional datasets like BraTS and explore domain adaptation techniques. These improvements will enhance model reliability and applicability in real-world medical diagnostics.

Another promising direction is developing adaptive models that update themselves with new data, ensuring the system stays current with evolving diagnostic standards and advances in imaging. Such adaptability would improve the model's robustness across different imaging technologies and clinical practices, making it versatile for use in diverse healthcare settings. We also plan to reduce the model's memory and processing demands so it can operate effectively on standard hospital equipment, eliminating the need for specialized hardware.

## 6 Conclusion

We considered several deep learning models in this study in the context of classifying MRI brain images into four categories: glioma, meningioma, pituitary, and no tumor. Our experiments involved the hybrid Vision Transformer + EfficientNetB4 model, custom 2D CNN, and EfficientNetB4. Each of the deep learning models showed great potential, and the best model was the EfficientNetB4 with the highest accuracy and a balanced performance by the hybrid model on all metrics.

Very efficient and very high accuracy EfficientNetB4 with far fewer parameters scaling well hybrid model took the best from both worlds—Vision Transformers and EfficientNetB4; in turn, achieving good generalization with great performance. The base line offered by the custom 2D CNN showed promising robustness yet has a lot of leeway for further improvement.

We had faced the following challenges and problems in our research: overfitting and class imbalance. We reduced some cases of overfitting and improved the generalization of our models, aided by using more developed data augmentation and fine-tuning of hyperparameters. More successful techniques were reducing the learning rate and early stopping.

Future work will involve further integration of other imaging modalities like CT or PET scans with MRI for improved characterization of tumors. GANs and other forms of advanced data augmentation to handle class imbalance will continue to be worked on. Explainable AI tools are also something we would want to work on for us to gain confidence in our models and further integrate them into clinical workflows, thus ensuring real-time processing and enabling doctors to make the proper decisions.

We have shown the possibility that deep learning models can be successful in brain tumor classification tasks, providing useful insights as well as a foundation for future research. Using new techniques and overcoming the associated challenges, our models would have a huge impact in clinical settings in terms of early diagnosis and treatment planning, thus improving patient outcome.

**Data availability** Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

## Declarations

**Conflict of interest** The authors declare no conflict of interest.

## References

1. Cleveland Clinic, Brain Cancer (Brain Tumor), <https://my.clevelandclinic.org/health/diseases/6149-brain-cancer-brain-tumor>
2. Brain Tumor Mayo Clinic, By Mayo Clinic Staff, <https://www.mayoclinic.org/diseases-conditions/brain-tumor/symp-toms-causes/syc-20350084>
3. Aaron Cohen-Gadol MD (2024) 72 Must-Know Brain Tumor Statistics, <https://www.aaroncohen-gadol.com/en/patients/brain-tumor/types/statistics>
4. Peiyi Gao, MD Wei Shan, MD, PhD, Yue Guo, MA et al (2022) Development and validation of a deep learning model for brain tumor diagnosis and classification using magnetic resonance imaging. JAMA Network Open <https://doi.org/10.1001/jamanetworkopen.2022.25608>
5. Abdusalomov AB, Mukhiddinov M, Whangbo TK (2024) Brain tumor detection based on deep learning approaches and magnetic resonance imaging. Cancers. <https://doi.org/10.3390/cancers15164172>
6. Mathivanan SK, Sonaimuthu S, Murugesan S, Rajadurai H, Shivalahare BD, Shah MA (2024) Employing deep learning and transfer learning for accurate brain tumor detection. Sci Rep. <https://doi.org/10.1038/s41598-024-57970-7>
7. Nishanth Rao K, Khalafb OI, Krishnasree V, Kumarc AS, Alsekaitd DM, Siva Priyankae S, Alattasf AS, AbdElminaamg DS (2024) An efficient brain tumor detection and classification using pre-trained convolutional neural network models. Heliyon. <https://doi.org/10.1016/j.heliyon.2024.e36773>
8. Balamurugan AG, Srinivasan S, Preethi D, Monica P, Mathivanan SK, Shah MA (2024) Robust brain tumor classification by fusion of deep learning and channel-wise attention mode approach. BMC Med Imaging. <https://doi.org/10.1186/s12880-024-01323-3>
9. Manilal PR, Shah DJ (2024) Early detection of brain tumors: a comprehensive study on MRI-based diagnosis using a combination of convolutional deep learning and machine learning techniques <https://ijisae.org/index.php/IJISAE/article/view/6380>
10. Rasool Reddy K, Rajesh KNVPS, Ravi RV (2024) BrainCDNet: a concatenated deep neural network for the detection of brain tumors from MRI images. Front Hum Neurosci. <https://doi.org/10.3389/fnhum.2024.1405586>
11. Ullah N, Javed A, Alhazmi A, Hasnain SM, Tahir A, Ashraf R (2024) TumorDetNet: a unified deep learning model for brain tumor detection and classification. PLOS One. <https://doi.org/10.1371/journal.pone.0291200>
12. Jain S, Sachdeva B (2024) A systematic review on brain tumor detection using deep learning. AIP Conf Proc 10(1063/5):0221113
13. Rasool N, IqbalBhat J (2024) Brain tumour detection using machine and deep learning: a systematic review. Multimed Tools Appl. <https://doi.org/10.1007/s11042-024-19333-2>
14. Saeedi S, Rezayi S, Keshavarz H, Niakan Kalhori S R (2023) MRI-based brain tumor detection using convolutional deep learning methods and chosen machine learning techniques. BMC Med Inform Decis Mak. <https://doi.org/10.1186/s12911-023-02114-6>
15. Buchade AC, Prasad Kantipudi MVV (2023) Recent trends on brain tumor detection using hybrid deep learning methods. Revue d'Intell Artif. <https://doi.org/10.18280/ria.380111>
16. Khan Md KH, Guo W, Liu J, Dong F, Li Z, Patterson TA, Hong H (2024) Machine learning and deep learning for brain tumor MRI image segmentation. Exp Biol Med. <https://doi.org/10.1177/15353702231214259>
17. Ghauri MS, Wang J-Y, Reddy AJ, Shabbir T, Tabaie E, Siddiqi J (2024) Brain tumor recognition using artificial intelligence neural-networks (BRAIN): a cost-effective clean-energy platform. Neuroglia. <https://doi.org/10.3390/neuroglia5020008>
18. Malakouti SM, Menhaj MB, Suratgar AA (2024) Machine learning and transfer learning techniques for accurate brain tumor classification. Clin eHealth. <https://doi.org/10.1016/j.ceh.2024.08.001>
19. Bohaju J Brain Tumor Kaggle Dataset, Available Online: <https://www.kaggle.com/datasets/jakeshbohaju/brain-tumor>
20. Amin J, Sharif M, Haldorai A, Yasmin M, Nayak RS (2021) Brain tumor detection and classification using machine learning: a comprehensive survey. Complex Intell Syst. <https://doi.org/10.1007/s40747-021-00563-y>
21. Keerthana A, Kavin Kumar B, Akshaya KS and Kamalraj S Ph.D (2024) Brain tumour detection using machine learning algorithm. Phys: Conf Ser <https://doi.org/10.1088/1742-6596/1937/1/012008>

22. Sharma M, Sharma P, Mittal R, Gupta K (2024) Brain tumour detection using machine learning. *J Electron Inform.* <https://doi.org/10.36548/jei.2021.4.005>
23. Hossain R, Ibrahim RB, Hashim HB (2024) Automated brain tumor detection using machine learning: a bibliometric review. *World Neurosurg.* <https://doi.org/10.1016/j.wneu.2023.03.115>
24. Aarthilakshmi M, Meenakshi S, Poorna Pushkala A, Rama@Ramalakshmi V, Prakash NB (2024) Brain tumor detection using machine learning. <https://www.ijstr.org/final-print/apr2020/Brain-Tumor-Detection-Using-Machine-Learning.pdf>
25. Asma Parveen A, Dr. Kamalakkannan T (2024) A research on brain tumor detection using machine learning techniques and deep learning approach. <https://www.jetir.org/view?paper=JETIR2309630>
26. BIDS Website Brain Tumor MRI, PET, MEG, EEG and iEEG Public Datasets Available Online: <https://openneuro.org/>
27. Siddhant Barshile, Subham Patil, Hemant Bhujbal, Prof. Nalavade Mam KC (2024), brain cancer detection using machine learning” <https://ijcrt.org/papers/IJCRT2203295.pdf>
28. Shekar AR Department of information technology, Jntuh Ucesth, Dr. G. Venkata Rami Reddy, Professor of IT, Jntuh Ucesth (2024) Brain tumour detection using machine learning <https://www.ijarst.in/public/uploads/paper/920741694703997.pdf>
29. Mahfuz Ahmed Md, Maruf Hossain Md, Rakibul Islam Md, Shahin Ali Md, Nafi AAI N, Faisal Ahmed Md, Ahmed KM, Sipon Miah Md, Mahbubur Rahman Md, Niu M, Khairul Islam Md (2024) Brain tumor detection and classification in MRI using hybrid ViT and GRU model with explainable AI in Southern Bangladesh. *Sci Rep.* <https://doi.org/10.1038/s41598-024-71893-3>
30. Priya A, Vasudevan V (2024) Brain tumor classification and detection via hybrid alexnet-gru based on deep learning. *Biomed Signal Process Control.* <https://doi.org/10.1016/j.bspc.2023.105716>
31. Vimala BB, Srinivasan S, Mathivanan SK, Mahalakshmi A, Jayagopal P, Dalu GT (2023) Detection and classification of brain tumor using hybrid deep learning models. *Sci Rep.* <https://doi.org/10.1038/s41598-023-50505-6>
32. Dixon J, Akinniyi O, Abdelhamid A, Saleh GA, Rahman Md M, Khalifa F (2024) A hybrid learning-architecture for improved brain tumor recognition. *Algorithms.* <https://doi.org/10.3390/a17060221>
33. Brain Tumor Figshare Dataset Available Online: [https://figshare.com/articles/dataset/brain\\_tumor\\_dataset/1512427](https://figshare.com/articles/dataset/brain_tumor_dataset/1512427)
34. Chaki J, Wozniak M Brain tumor MRI dataset. Available online: <https://ieee-dataport.org/documents/brain-tumor-mri-dataset>
35. Rohini A, Praveen C, Mathivanan SK, Muthukumaran V, Mallik S, Alqahtani MS, Al-Rasheed Amal, Soufiene BO (2024) Multimodal hybrid convolutional neural network-based brain tumor grade classification. *BMC Bioinform.* <https://doi.org/10.1186/s12859-023-05518-3>
36. Mohsen S, Ali AM, El-Rabaie El-SM, ElKaseer A, Scholz SG, Hassan AMA (2024) Brain tumor classification using hybrid single image super-resolution technique with ResNext101\_32× 8d and VGG19 pre-trained models. *IEEE Access.* <https://doi.org/10.1109/ACCESS.2023.3281529>
37. Balamurugan T, Gnanamanoharan E (2024) Brain tumor segmentation and classification using hybrid deep CNN with LuNet classifier. *Neural Comput Appl.* <https://doi.org/10.21203/rs.3.rs-1599383/v1>
38. Amran GA, Alsharam MS, Abdullah OA, Blajam AA, Hasan MY, Alfaifi MH, Amran AG, Eldin SM (2022) Brain tumor classification and detection using hybrid deep tumor network. *Electronics.* <https://doi.org/10.3390/electronics11213457>
39. Hamada A (2020) BrH35: brain tumor detection dataset, Available Online: <https://www.kaggle.com/datasets/ahmedhamada0/brain-tumor-detection>
40. Senan EM, Jadhav M E, Rassem TH, Aljaloud AS, Mohammed BA, Al-Mekhlafi Z G (2022) Early diagnosis of brain tumour mri images using hybrid techniques between deep and machine learning. *Comput Math Methods Med.* <https://doi.org/10.1155/2022/8330833>
41. AlTahhan FE, Khouqeer GA, Saadi S, Elgarayhi A, Sallah M (2023) Refined automatic brain tumor classification using hybrid convolutional neural networks for MRI scans. *Diagnostics.* <https://doi.org/10.3390/diagnostics13050864>
42. Mahfuz Ahmed Md, Maruf Hossain Md, Rakibul Islam Md, Shahin Ali Md, Nafi AAI N, Faisal Ahmed Md, Ahmed KM, Sipon Miah Md, Mahbubur Rahman Md, Niu M, Khairul Islam Md (2024) Brain tumor detection and classification in MRI using hybrid ViT and GRU model with explainable AI in Southern Bangladesh. *Sci Rep.* <https://doi.org/10.1038/s41598-024-71893-3>
43. Nickparvar M (2021) Brain tumor MRI dataset Kaggle source dataset, Available Online: <https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset>
44. Tan, Mingxing, and Quoc Le (2019) Efficientnet: rethinking model scaling for convolutional neural networks. In international conference on machine learning (pp. 6105–6114). PMLR <https://doi.org/10.48550/arXiv.1905.11946>
45. Khaliki MZ, Basarslan MS (2024) Brain tumor detection from images and comparison with transfer learning methods and 3-layer CNN. *Sci Rep.* <https://doi.org/10.1038/s41598-024-52823-9>
46. Tripathy S, Singh R, Ray M (2023) Automation of brain tumor identification using efficientnet on magnetic resonance images. *Procedia Comput Sci.* <https://doi.org/10.1016/j.procs.2023.01.133>
47. Jia D et al (2009) Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition (pp. 248–255). IEEE. <https://doi.org/10.1109/CVPR.2009.5206848>
48. <https://huggingface.co/google/vit-large-patch16-224>

49. Huo Y, Jin K, Cai J, Xiong H (2023) Vision transformer (ViT)-based applications in image classification. <https://doi.org/10.1109/BigDataSecurity-HPSC-IDS58521.2023.00033>
50. Zhagalasbayev A, Aiteni T, Khaimuldin N (2024) Using the CNN architecture based on the EfficientNetB4 model to efficiently detect Deepfake images. [https://www.researchgate.net/publication/378157870\\_Using\\_the\\_CNN\\_architecture\\_based\\_on\\_the\\_EfficientNetB4\\_model\\_to\\_efficiently\\_detect\\_Deepfake\\_images](https://www.researchgate.net/publication/378157870_Using_the_CNN_architecture_based_on_the_EfficientNetB4_model_to_efficiently_detect_Deepfake_images)
51. Hossain T, Shishir FS, Ashraf M, MD Abdullah Al Nasim, Shah FM (2019) Brain tumor detection using convolutional neural network. In: 2019 1st international conference on advances in science, engineering and robotics technology (ICASERT) (pp. 1–6). IEEE. <https://doi.org/10.1109/ICASERT.2019.8934561>
52. Cherian JM, Kumar R (2023) Fundamentals of machine learning. In: Badar M (ed) A guide to applied machine learning for biologists. Springer, Cham
53. Goutham V, Sameerunnisa A, Babu S and Prakash TB (2022) Brain tumor classification using EfficientNet-B0 model. In: 2022 2nd international conference on advance computing and innovative technologies in engineering (ICACITE). IEE <https://doi.org/10.1109/ICACITE53722.2022.9823526>
54. Ananth AD, Palanisamy C (2021) Extended and optimized deep convolutional neural network-based lung tumor identification in big data. Int J Imaging Syst Tech. <https://doi.org/10.1002/ima.22667>
55. Wadhhah A et al (2021) Deep CNN for brain tumor classification. Neural Process Lett. <https://doi.org/10.1007/s11063-020-10398-2>
56. Mishra L, Verma S, Varma S (2022) Hybrid model using feature extraction and non-linear SVM for brain tumor classification. <https://doi.org/10.48550/arXiv.2212.02794>
57. Singh M, Bhange PU, Touhang J (2023) Brain tumor detection using machine learning methods-KNN, SVM, DT, RF. Int J Chem Eng Process. <https://doi.org/10.37628/jcep.v9i2.1427>
58. Kumar A, Jha AK, Agarwal JP, Yadav M, Badhe S, Sahay A, Epari S, Sahu A, Bhattacharya K, Chatterjee A, Ganeshan B, Rangarajan V, Moyiadi A, Gupta T, Goda JS (2023) Machine-Learning-based radiomics for classifying glioma grade using multiparametric MRI. J Personalized Med. <https://doi.org/10.3390/jpm13060920>
59. Bhimavarapu U, Chintalapudi N, Battineni G (2024) Brain tumor detection and categorization with segmentation of MRI images using improved fuzzy C-means clustering algorithm. Bioengineering. <https://doi.org/10.3390/bioengineering11030266>
60. Venkata Sai Krishna Chaitanya Kandula Automated brain tumor classifier with deep learning, <https://scholarworks.lib.csusb.edu/cgi/viewcontent.cgi?article=3047&context=etd>
61. Haonan Y, Yang D (2023) CSwin-PNet: a CNN-swin transformer combined pyramid network for breast lesion segmentation in ultrasound images. Expert Syst Appl 2135:2151
62. Neelima G, Chigurukota DR, Maram B, Girirajan B (2022) Optimal DeepMRSeg based tumor segmentation with GAN for brain tumor classification. Biomed Signal Process Control. <https://doi.org/10.1016/j.bspc.2022.103537>
63. Al AS, Muhammad FD, Dawsari HAI (2024) CNN-TCN: deep hybrid model based on custom CNN with temporal CNN to recognize sign language. J Disability Res 51:135
64. Behrad F, Abadeh MS (2022) An overview of deep learning methods for multimodal medical data mining. Expert Syst Appl. <https://doi.org/10.1016/j.eswa.2022.117006>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

## Authors and Affiliations

**S. Selva Karthik<sup>1</sup> · V. Achuthan<sup>1</sup> · A. Dennis Ananth<sup>1</sup> · R. Venkatesan<sup>1</sup> · M. Rajakumaran<sup>1</sup> · S. Markkandeyan<sup>1</sup>**

 M. Rajakumaran  
rajakumaran@cse.sastra.edu

<sup>1</sup> School of Computing, SASTRA Deemed University, Thanjavur, Tamilnadu, India