Emerging Technologies and Engineering Journal

EngiScience Publisher

Research Article

Averaging-Based Hybrid Ensemble of DenseNet-121 and ResNet-50 for Computer-Aided Brain Tumour Diagnosis: A Step Towards Early Detection

Laiba Hanif [1], Rehah Khan [1,*], Kinza Aziz [1], Irfan Arshad [1], Mamoona Khalid [1]

[1] Photonics and Communications Lab, Electrical Engineering Department, Faculty of Electronics and Electrical Engineering, University of Engineering and Technology, Taxila, 47080, Pakistan

*Corresponding Author: Rehah Khan, E-mail: 21-ee-91@students.uettaxila.edu.pk

Article Info

Article History
Received Nov 01, 2025
Revised Jan 21, 2026
Accepted Jan 27, 2026

Keywords
Brain tumour
Early detection
Deep learning
Ensemble model
Convolutional Neural Networks (CNN)
Hybrid Model

Abstract

One of the deadliest diseases that can affect children and adults is a brain tumour. A brain tumour is the abnormal growth of cells in the brain or near it. Commonly, there are two types of tumours: malignant (cancerous) and benign (non-cancerous). The importance of early detection of brain tumours cannot be overstated, as tumours progress rapidly and decrease survival rates. Due to the complexity of brain tumour structure and classification, manual inspection can lead to false-positive or false-negative results. The complexity of tumour detection can be reduced by using deep learning segmentation and classification models. This paper proposes an AI-driven solution for screening brain tumours using MRI images, emphasising improving the model's sensitivity to reduce missed tumour cases. A deep hybrid ensemble model is developed by combining two CNN models using an averaging ensemble learning technique. The models were trained and evaluated on a publicly available MRI dataset consisting of 4,845 images. Four state-of-the-art deep learning architectures were trained and evaluated, and the best-performing models were combined through the ensemble technique. Our final model achieves 98.49% accuracy and 97.58% sensitivity. Compared with a recent ensemble-based baseline, the proposed approach reduces error rate and false negative cases, improving robustness and clinical reliability.

1. Introduction

The brain is the most important organ of the human body, as it controls the entire nervous system. Irregular, excessive and uncontrolled growth of cells in the brain results in a brain tumour. According to the newer edition of the World Health Organisation classification, brain tumours are generally classified as either cancerous (malignant) or non-cancerous (benign). Gliomas can range from low-grade (slow-growing) to high-grade (malignant) [1]. In contrast, malignant tumours are more dominant and have the ability to spread to other parts of the body. For brain tumour detection and classification, Magnetic resonance imaging (MRI) is the most common modality technique as it shows the brain more clearly than other imaging tests do [2]. However, manual detection and classification of tumours from MRI images can lead to human error. That's why, for finding the exact location of the tumour, a more reliable diagnostic

tool is important. Previous brain tumour diagnostics mostly relied on the radiologist's experience. Manual tumour identification by different radiologists typically takes 2-4 hours. To avoid the future complexity of brain tumour detection and classification, a more reliable tool is needed [3].

Improvements in Medical Imaging Technology and the evolution of Artificial Intelligence over the last two decades have enabled the health care sector to achieve remarkable results in disease analysis and diagnosis [4]. Early detection of cancer using Artificial Intelligence is an emerging field of research. As tumours can appear in different regions of the brain, early detection and classification of brain tumours are challenging. Performing this task with high accuracy is a critical step in diagnosis and treatment to save a patient's life. Using an AI model can help reduce human error and take less time to identify the exact location of the tumour. Machine learning algorithms have achieved significant success in medical imaging analysis. Deep Learning (DL), a subfield of machine learning, which utilises multi-layered neural networks with backpropagation to automatically learn hierarchical data representations, achieving state-of-the-art results across speech, vision, and sequential data domains [5]. However, machine learning techniques are implemented on limited data and require manual feature extraction. In contrast, deep learning (DL) extracts features in a self-learning manner from labelled data, which improves the performance of the models.

Several methods have been proposed by different scholars for brain tumour detection using machine learning. Over the years, researchers have explored a wide variety of machine learning (ML) and deep learning (DL) techniques to address the challenges of brain tumour detection. Early studies primarily relied on traditional ML methods. For example, Zacharaki et al. [4] employed support vector machines (SVMs) with features such as shape, texture, and intensity. While their approach demonstrated the potential of automated classification, its performance was constrained by the limited dataset size, which restricted overall accuracy. Similarly, Abdolmaleki et al. [5] used a shallow neural network trained on thirteen handcrafted features to differentiate between benign and malignant tumours. Although this method achieved moderate accuracy, it highlighted the limitations of relying on small sets of manually extracted features.

As deep learning emerged, more powerful architectures began to replace these conventional methods. Razzak et al. [6] compared different DL models—including CNNs, DNNs, RNNs, and DCNs—with traditional ML algorithms. Their findings demonstrated that deep learning consistently outperformed earlier techniques in both accuracy and processing efficiency. Building on this, Theresa et al. [7] introduced a Deep Recurrent LSTM (DRLSTM) model and showed that it surpassed CNNs and other ML models, particularly in terms of prediction accuracy. In parallel, Ping Li et al. [8] focused on optimising ResNet-based CNNs, achieving enhanced predictive performance through careful pre-processing and hyperparameter tuning.

The trend toward hybrid and ensemble architectures has also gained momentum. Alsubai et al. [9]

combined CNN and LSTM networks using MRI images, demonstrating improved classification accuracy compared to single-model approaches. Zahoor et al. [1] advanced this further by proposing DBFS-EC, an ensemble framework that integrated four CNN models, thereby expanding the feature space and significantly improving MRI classification performance. Similarly, Patil et al. [10] developed an ensemble of deep CNNs that fused SCNN with VGG16thereby improving accuracy, reducing overfitting, and enhancing multiclass classification. More recently, Nongmeikapam et al. [11] compared conventional ML classifiers such as KNN, SVM, and Random Forest with CNN-based methods. Their results reaffirmed that CNN architectures deliver superior predictions and generalisation, particularly for complex medical imaging data.

Taken together, these studies illustrate a clear progression in the field from early reliance on handcrafted features and traditional ML to increasingly sophisticated deep learning and ensemble methods. While most prior work emphasised automated detection, relatively few focused specifically on improving early-detection sensitivity. This gap underscores the motivation for the present study, which aims to design a hybrid ensemble model that reduces false negatives and enhances robustness for timely brain tumour diagnosis.

Table 1. Summary Table of Literature Review.

Previous Study	Model / Method	Key results	Limitations
Zacharaki et al. [4]	SVM with handcrafted features	Moderate accuracy	Limited dataset, manual feature extraction
Abdolmaleki et al. [5]	Shallow Neural Network	Moderate performance	Relies on handcrafted features, low scalability
Razzak et al. [6]	CNN, DNN, RNN, DCN	DL outperformed ML	No focus on early detection sensitivity
Theresa et al. [7]	DRLSTM	Higher accuracy than CNN	High computational complexity
Ping Li et al. [8]	Optimised ResNet-based CNN	Improved accuracy	Single-model dependency
Alsubai et al. [9]	CNN + LSTM Ensemble	Improved classification accuracy	Increased model complexity
Zahoor et al. [1]	DBFS-EC (4 CNN Ensemble)	Expanding the feature space and significantly improving MRI classification performance	Computationally expensive
Patil et al. [10]	VGG16 + SCNN Ensemble	Enhancing accuracy, reducing overfitting, and strengthening multiclass classification	Complex architecture
Nongmeikapam et al. [11]	ML vs CNN comparison	CNN is superior to ML	No emphasis on early detection
Asif et al. [12]	Xception	91.94% accuracy	Small dataset
Borra et al. [13]	U-Net + SVM	High accuracy and specificity	Focused more on segmentation

All the works described above focus on automated brain tumour detection, without emphasis on early detection. This paper aims to increase the final model's sensitivity to reduce false negatives, thereby improving the detection of all positive cases, and to enhance the model's accuracy for correct prediction in all cases.

2. Model Selection

In our proposed methodology, a deep learning approach has been used for the classification of brain tumours. The basic structure and architecture of deep learning are inspired by the neural networks present in the human brain. The significance of deep learning is that it has replaced manual classification and feature extraction with an automated approach [14]. In our work, we have employed the Convolutional Neural Network (CNN), which is primarily designed for object recognition, including image classification, detection, and segmentation. The key benefit of CNNs is that they recognise and extract relevant features automatically, without any human supervision, compared with theipredecessors [15]. Basic Convolutional Neural Network (CNN) architecture is illustrated in Figure 1. To perform brain tumour classification, we have implemented various state-of-the-art CNN architectures, including ResNet-50, DenseNet-121, U-Net, and Xception. This section reviews previous work using the models discussed above for brain tumour detection and the potential reasons why these models should be used for the brain tumour detection task.

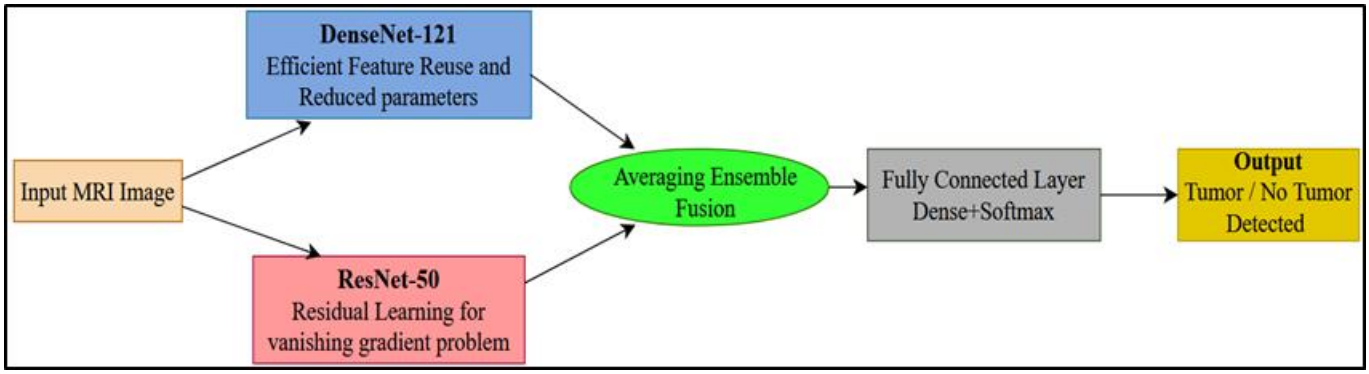


Figure 1. Workflow of the proposed CNN Model Training

2.1. ResNet-50

ResNet-50 mitigates the vanishing gradient problem with its residual skip connections and enables stable training of very deep networks on complex MRI data. This architecture is suitable for medical image classification, where fine-grained intensity variations must be conserved across layers. Shivansh Rathore et al. [16] implemented and compared ResNet-50 with other deep neural networks using a publicly available MRI image dataset. In this research, the image dataset was sourced from Kaggle, comprising both tumour and non-tumour images. The dataset was divided into three other subsets: training, validation, and testing. This dataset includes 1500 tumour and 1500 non-tumour images, with a separate directory containing 60 MRI images designated for testing. The ResNet-50 model achieved an accuracy of 97.17% with the

precision, sensitivity, and F1-score values of 97.32%, 97.17%, and 97.16%, respectively. This result shows strong performance in the confusion matrix; that's why this model was selected for our comparative analysis.

2.2. DenseNet-121

DenseNet-121 was selected for its dense connectivity pattern, which promotes feature reuse and strengthens gradient flow, resulting in improved tumour detection sensitivity. This property is particularly beneficial for early-detection tasks where fine-grained features are essential. Ouisa Nait Belaid et al. [17] used the DenseNet-121 architecture in their study. In their paper, they employed two pre-trained Dense Convolutional Networks (DenseNet121) to extract the features from both the original and cropped MRI images. These extracted features were then combined and passed through a softmax layer for the classification of brain tumours. The proposed method, which concatenates features from the outputs of pre-trained DenseNet-121 models, achieved an accuracy of 98.86% on the test dataset, surpassing the state of the art in brain tumour detection.

2.3. Xception

Xception allows efficient extraction of channel-wise and spatial features and employs depth-wise separable convolutions, which reduce computational complexity. It is capable of modelling complex spatial patterns suitable for learning diverse tumour appearances in MRI Images. Sohaib Asif et al.[12] employed the Xception model for brain tumour detection. On the MRI dataset, it achieved an overall accuracy of 91.94%, with sensitivity, precision, specificity, and F1-score values of 96.55%, 87.50%, 87.88%, and 91.80%, respectively. The MRI images consist of 253 patients, including 155 tumorous and 98 non-tumorous cases. The proposed method outperforms existing approaches, demonstrating its effectiveness in accurately and efficiently classifying brain tumours. Due to its robust performance on the large datasets, we have selected the Xception model as one of the deep learning architectures for our brain tumour detection system.

2.4. U-Net

U-Net is originally designed for segmentation; it was included due to its encoder–decoder architecture with skip connections, which capture both localised structural and global context information. The encoder path effectively learns discriminative features, which makes the U-Net adaptable for classification tasks when segmentation-level feature representation is advantageous. Reddy Borra et al.[13] employed the U-Net architecture, implementing a methodology that includes successive stages such as pre-processing, segmentation and edge detection. The approach demonstrates exceptional performance, achieving 99.67% accuracy, along with high specificity and precision of 99.73% and 98.79%, respectively. The sensitivity also defines strong results at 97.43%. Due to its effectiveness in accurately and automatically detecting and

classifying brain tumours in medical imaging, this model has been selected as one of the comparative architectures in our brain tumour detection analysis.

3. Proposed Methodology

In this paper, a structured method is developed to create a model with improved accuracy and sensitivity, aiding accurate diagnosis of brain tumours. For this purpose, a deep study of various architectures and algorithms was conducted to choose the best available deep learning model. After analysing and comparing the features of different deep learning models, Densenet-121, Resnet-50, U-Net, and the Xception model were selected for training. All the models are trained on the same pre-processed data to have a fair comparison of the model performance. Once models are trained, they are compared based on their performance metrics, and the two best models are selected to create a hybrid model via ensemble learning. An overview of the steps followed through the proposed method is shown in Figure 2.

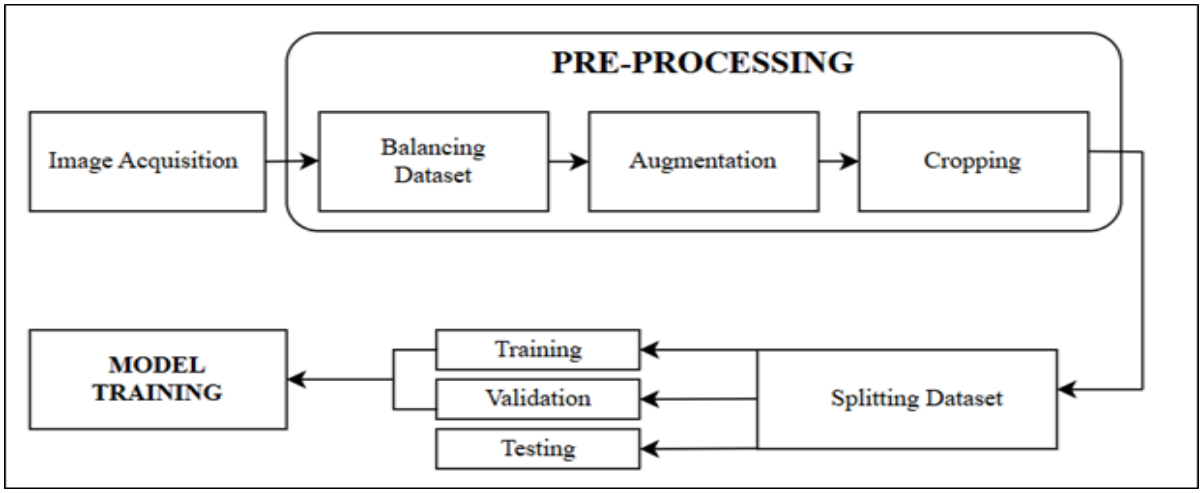


Figure 2. Block Diagram of proposed methodology: from Data Acquisition to Model Training.

3.1. Image Acquisition

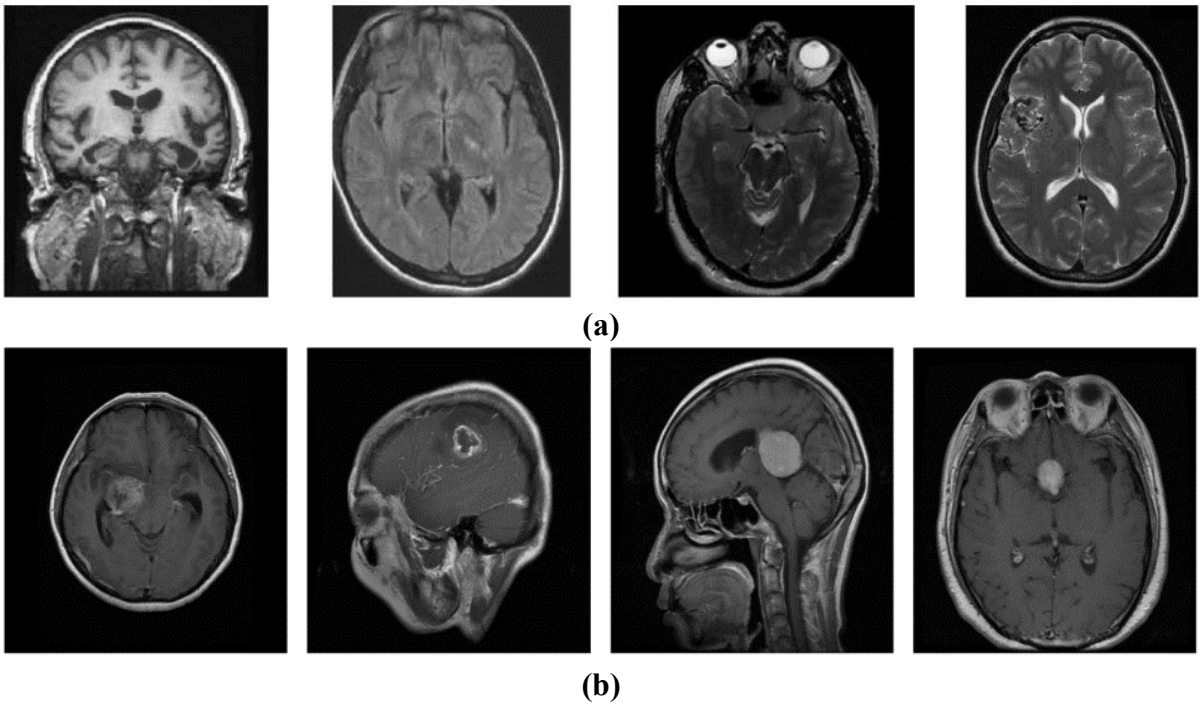
Multiple brain tumour datasets are publicly available for research and training of machine learning or deep learning models[18]. Images available in the datasets are acquired using traditional methods such as Computed Tomography (CT) or Magnetic Resonance Imaging (MRI). For training a brain tumour detection model, open-source datasets can be used from Kaggle or BraTS. For this study, we have utilised a 2D image dataset uploaded by Sartaj on Kaggle.com [19]. The dataset includes two folders named “Training” and “Testing”. Each folder is organised into four subfolders: glioma, meningioma, pituitary tumour, and no tumour, each containing MRI scan images. Table 2 shows the number of images per class in the dataset.

In contrast to Kaggle, the BraTS dataset contains only 3D images of tumorous brain tissue. The BraTS dataset is suitable for tasks primarily focused on segmentation, as it does not provide 3D healthy brain images, which are essential for training tumour detection models. The objective of this study is screening-

level brain tumour detection; the absence of healthy image samples limits the suitability of the proposed method.

Table 2. Class-wise Summary of Original Dataset.

	Tumorous			Healthy
	Glioma	**Meningioma**	**Pituitary**	**No tumor**
1. Training	826	822	827	395
2. Testing	100	115	74	105



(a)

(b)

Figure 3. (a) MRI Images of Healthy Brain (b) MRI Images of Brain Tumour

3.2. Pre-Processing

After data acquisition, it is not ready to be used directly for model training. We need to perform several pre-processing steps and techniques to prepare the dataset for model training. According to our proposed methodology, we have manually created two folders, named tumour and healthy. “Healthy” folder consists of non-tumorous MRI images, whereas “Tumour” folder contains glioma, meningioma and pituitary tumour folder categories. Merging tumour subtypes into a single class increased the dataset's diversity, ensuring that no tumour cases are missed, irrespective of their type or location. This approach would improve the efficiency of binary tumour classification. Pre-processing includes renaming the dataset images, balancing, augmentation, cropping, and splitting the dataset into train, test, and validation folders.

3.2.1. Renaming

Images in the dataset are renamed first to help with file counting and to check the number of files in the tumour and no-tumour folders. We have named the folders as “tumorous” and “healthy”.

3.2.2. Balancing dataset

After combining all the tumorous image files into one folder, the total count of tumorous images in the training folder is 2475, and the number of non-tumorous or healthy images in the training folder is 395. The dataset is clearly imbalanced. The graph of the unbalanced dataset is shown in Figure 4a. Data Augmentation helped to increase the volume of the training dataset. Since all types of tumorous images were combined into a single folder, this dataset is already diverse, with varying sizes, shapes, and locations. To prevent the model from becoming biased toward predicting “tumour” only, augmentation is applied only to the healthy images. For this purpose, the ImageDataGenerator from theensorFlow library is used [20]. This library generates multiple versions of each image by rotating, flipping, shifting or changing the brightness of the image [21]. Images generated by ImageDataGenerator increase dataset diversity and help balance the dataset. The graph of the balanced dataset is shown in Figure 4b. A balanced dataset increases the model's generalizability and, as a result, helps overcomeverfitting during model training [20].

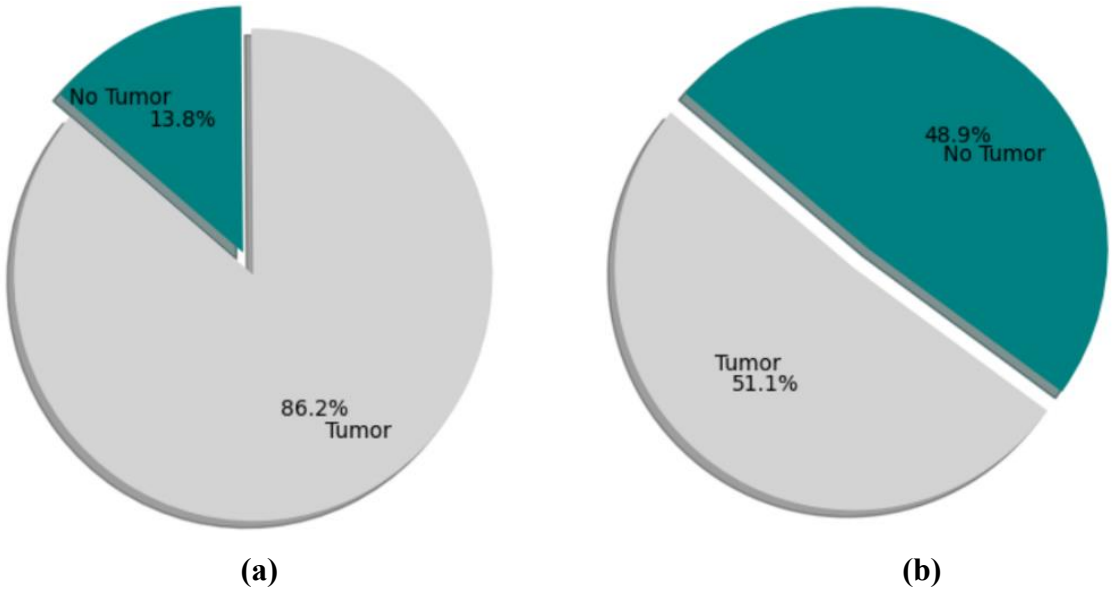


Figure 4. (a) Graph Representing Unbalanced Dataset **(b)** Graph Representing Balanced Dataset

3.2.3 Augmentation

To achieve a balanced dataset, we have applied data augmentation only to healthy images, as they are fewer in number than tumorous images. ImageDataGenerator provides multiple transformation techniques for images. For our purposes, we will generate five augmented images from each healthy image in the training dataset. Augmentation parameters selected for our proposed methodology include adjusting the rotation and brightness ranges, flipping horizontally and vertically, and adjusting the fill mode, as shown in Table 3.

After data augmentation, the balance between the total number of tumorous and healthy images is achieved. The graph of the balanced dataset is shown in Figure 4b.

Table 3. Augmentation Parameters.

Augmentation Parameters	Value
Rotation Range	$\theta = [-15^o, 15^o]$
Brightness range	(0.3, 1.0)
Horizontal Flip	True
Vertical Flip	True
Fill Mode	"Nearest"

3.2.4. Cropping

Cropping is another important step in the pre-processing of the dataset as it helps reduce the complexity of the data [22]. Cropping in our proposed method focuses on extracting the brain region or cropping out the black regions present in the MRI images, which are of no use during model training. Several Image processing techniques are used to extract the Region of Interest (ROI). The first MRI image is converted to grayscale and then to binary, as it is easier to extract the ROI from binary images. To remove noise from MRI images, *GaussianBlur* is used. This function reduces the noise component in the image, making it appear clearer and more visible, suitable for efficient model training. Mathematically, the function of Gaussian Blurring can be represented as shown in Equation (1).

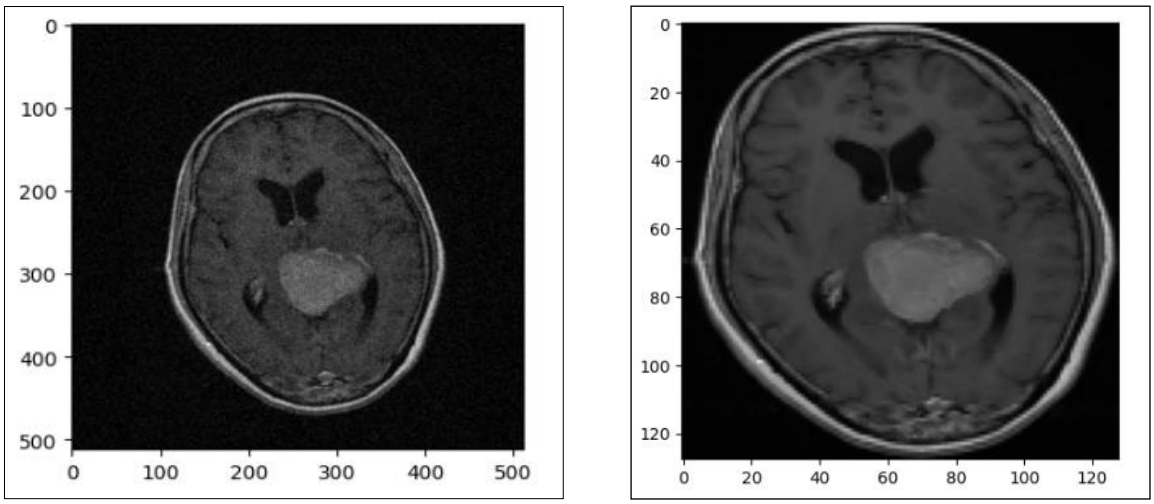
$$I'(x,y) = \sum_{i=-k}^{k} \sum_{j=-k}^{k} I(x+i, y+j).G(i,j) \quad (1)$$

If a Gaussian blur is used alone, it will be insufficient for effective image pre-processing to remove noise components. Therefore, our approach incorporates a two-step morphological operation: erosion followed by dilation to clean the images thoroughly. Erosion helps effectively remove small, isolated noise elements that persist despite applying a Gaussian blur. Next, dilation is applied to restore the size of significant regions, reconnect any broken parts, and fill small gaps, ensuring a continuous representation of the brain region.

A predefined function for finding contours is used to detect the boundary of an object in a binary image. After estimating the object boundary, the max function is used to return the largest contour amongst all, as the largest contour is assumed to be the brain. The area is defined in Equation (2) below.

$$Area = \frac{1}{2}\left|\sum_{i=1}^{n}(x_i y_{i+1} - x_{i+1} y_i)\right| \quad (2)$$

Since the images in the dataset were not of equal size, and for efficient model training, all images must be of the same height, width, and channel count. We resized all images to (128 × 128 × 3) by cropping. The image size is selected to preserve essential features necessary for efficient model training. After cropping, all images are saved in a folder; this is the final training dataset, ready for model training.



(a) **(b)**

Figure 5. (a) Image before pre-processing and cropping. **(b)** Image after pre-processing and cropping.

3.3. Splitting the Dataset

It is crucial to split processed data into multiple datasets, i.e. training, testing and validation, to determine the ability of every deep learning model. Data splitting is performed after augmentation when the dataset is balanced. We split the dataset into three categories: training, testing, and validation. At different stages, each dataset serves different purposes for model development. The training dataset is the largest of the three datasets, as it contains data on which the model trains. During the training process, the model learns different patterns, optimises its parameters, and develops relationships with the training data to perform specific tasks. The validation dataset is used to evaluate the model during training. It tests the model's execution and monitors the performance metrics. This helps select the best model from all the trained models by evaluating model performance at each epoch. The testing dataset is used to evaluate the final model after the training process. It determines how accurate our model is by analysing true positives and true negatives [23].

For our model training, we have used *the train_test_split library from scikit-learn.* This library helps split matrices or images into random training and test datasets[24]. The ratios for the training, validation, and test datasets are 70%, 15%, and 15%, respectively. The 70% training dataset consists of 3390 images, of which 1658 are healthy, and 1732 are tumorous. The 15% validation dataset consists of 727 images, including 356 healthy brain images and 371 tumorous brain images. The testing dataset contains 728 images, including 356 healthy and 372 tumorous images.

Table 4. Image count of the dataset after splitting

	Training Dataset (70%)		**Testing Dataset (15%)**		**Validation Dataset (15%)**	
	Tumorous	**Healthy**	**Tumorous**	**Healthy**	**Tumorous**	**Healthy**
No of Images	1732	1658	372	356	371	356

3.4. Model Training

3.4.1. Experimental Setup

To implement the proposed brain tumour detection model, all models were trained in the Jupyter Notebook environment using Python 3.12. Model Training utilised Keras API with TensorFlow as the backend. The model was trained across three hardware systems, enabling selection of the best-performing models for constructing the final ensemble. This multi-system approach also facilitates analysis of the correlation between model performance and hardware specifications.

3.4.2. Training

All models are trained using the same methodology to ensure a fair and consistent comparison. Models are compiled by defining the *input size* (128, 128, 3) and *number of classes = 2*. "Adam" optimiser has been used to optimise our trained model. Adam optimiser is computationally efficient, uses less memory, and is particularly effective for problems involving a large number of parameters[25]. The methodology followed in this paper for model training uses callbacks during training. *The callback monitor "val_loss" helps save the best model based on its validation loss*. *The patience value is set to 3 to automatically reduce the model's learning rate if the validation loss is not improving for 3* consecutive epochs. *min_lr* is set to *0.000001; this ensures that the learning rate does not fall below this value and helps fine-tune* the final saved model. In ModelCheckpoint, save_best_only is enabled to save the best models only if their validation loss has improved over previous epochs.

Each model is trained on the same lines of code, and the best-performing model is selected for the final averaging technique.

3.4.3. Comparative Analysis of Trained Model

After training all four architectures on three different PCs, the best parameter values are shown in Table 5 below. This comparative analysis, based on selected performance metrics (accuracy and sensitivity), helps select the best architecture among the trained models for constructing the final hybrid ensemble model.

Table 5. Comparison Table of Trained Models

Model	Accuracy	Sensitivity	Specificity	F1-Score
ResNet-50	95.88%	93.55%	98.31%	94.70%
DenseNet-121	98.49%	97.31%	99.72%	97.90%
U-Net	96.25%	94.35%	97.75%	95.18%
Xception	97.25%	95.70%	98.88%	96.47%

A bar graph showing the accuracy and sensitivity of the trained models is shown in Figure 6 to observe the performance of different architectures easily.

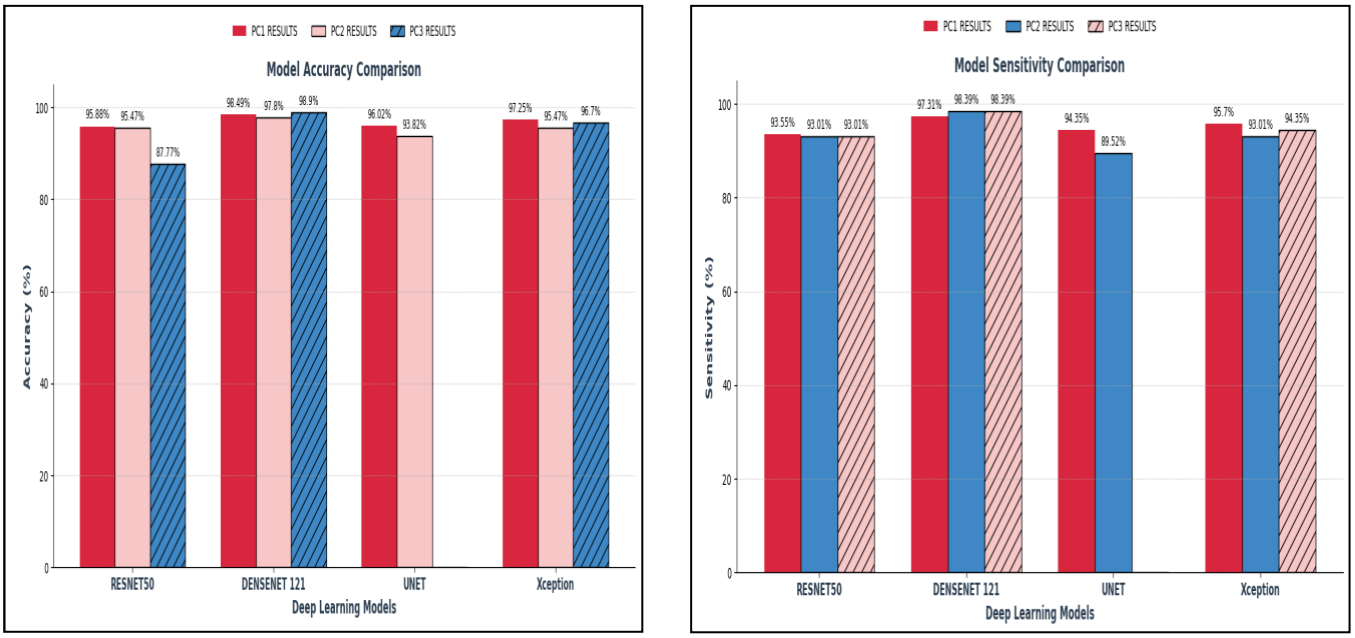


Figure 6. **(a)** Graph comparing the accuracy of Models. **(b)** Graph Comparing Sensitivity of Models.

3.5. Ensemble Learning

Ensemble learning combines weak learners. This technique makes the final model stronger than the individual models. Ensemble learning consists of three major types: Bagging, Boosting and Stacking. The methodology defined in this paper uses models that are already very deep and strong learners. Due to the complexity of the chosen model's architecture, applying stacking or voting may add trainable parameters, increasing computational cost and the risk of overfitting. Therefore, a simple averaging or bagging method is used. Ensemble learning increases the diversity of models and helps reduce the bias of individual base models, enabling the final model to inherit the strengths of each base model. In this technique, models are trained on a training dataset, and their outputs are averaged using the averaging function[26]. Averaging function: train individual models on the training dataset and make predictions on the testing dataset. The predictions from each model are averaged using an averaging function to provide the final decision.

To create our proposed hybrid model, we have selected the models with the best testing performance. The parameters of primary focus taken for analysis are Accuracy and Sensitivity. From Table 5, we can observe that DenseNet-121, ResNet-50, and Xception give the best results on the test dataset. These models have been selected to create the final ensemble model. DenseNet-121 is ensembled once with both models, because the trained DenseNet-121 models are more sensitive than the other three models[27]. Since the parameter of focus is increased sensitivity, which measures the proportion of actual true positive cases, this helps detect tumour cases that might be missed by manual inspection. Beyond improved sensitivity of individual models, the proposed ensemble technique mitigates their limitations and produces more stable predictions by combining the strengths of different architectures. This enables the final model

to detect subtle patterns that a single network might miss. A block diagram of ensemble learning is shown in Figure 7.

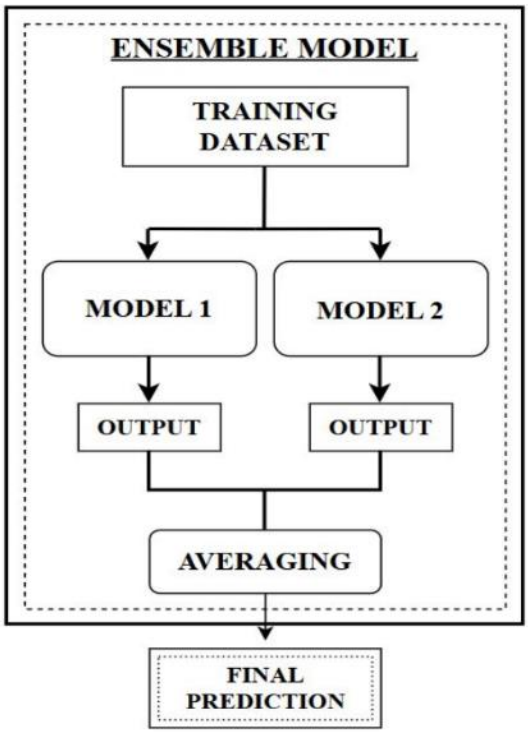


Figure 7. Block Diagram of Ensemble Learning.

4. Results and Discussion

In this section, the final results of the proposed hybrid model are discussed. The final ensemble model is tested using 15% of the testing dataset, initially separated from the total dataset. The best-performing combination of ResNet + DenseNet-121 and DenseNet-121 + Xception is compared.

4.1. Performance Analysis

The simplest and most common way to analyse any model's performance is to validate and compare it with different statistical metrics. In this study, our model's performance is evaluated using accuracy, sensitivity, precision, and F1-score, which are commonly used for brain tumour detection and classification [7]. Typically, accuracy is considered the most important parameter for evaluating a model's detection efficiency. Secondly, a high sensitivity score indicates a lower risk of overlooking or falsely detecting tumours, which is vital for brain tumour detection. In our analysis, the key parameters we have emphasised are accuracy and sensitivity. It helps to examine the model's ability to identify true positives and negatives accurately. The following formula can calculate accuracy and sensitivity:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{9}$$

$$Sensitivity = \frac{TP}{TP + FN}\ , \tag{10}$$

where, TP = True Positive, TN = True Negative, FP = False Positive and FN = False Negative. Similarly, other parameters such as precision and f1-score are widely used for testing and validating the model's performance, which are computed as follows:

$$Precision = \frac{TP}{TP + FP} \tag{11}$$

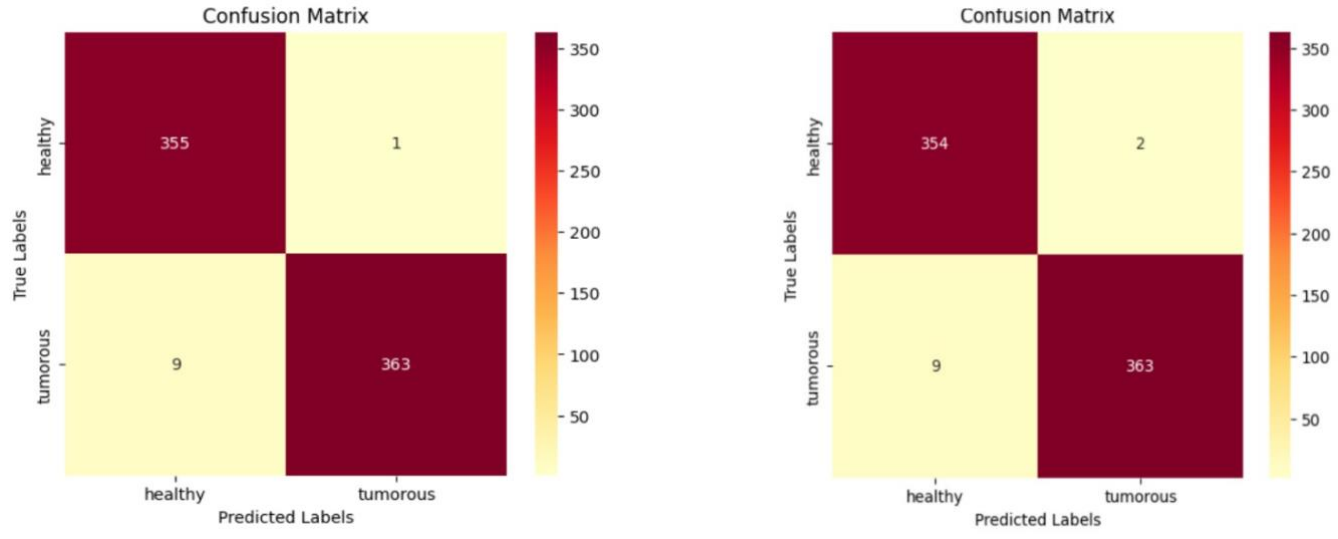
$$Specificity = \frac{TN}{TN + FP} \tag{12}$$

$$F1_{score} = 2 \times \frac{Precision \times Sensitivity}{Precison + Sensitivty} \tag{13}$$

Table 6. Comparison of the Proposed Model with Other Architecture Models.

References	Dataset Size (No. of images)	Model	Accuracy (%)	Sensitivity (%)	Specificity (%)	F1-Score (%)
Sawle et al. [28]	7860	VGG-16 ResNet-50	70 97	xx	xx	xx
Sinha et al. [18]	2530	Modified CNN	98	xx	xx	xx
Saeedi et al. [29]	3264	2D CNN	96.47	xx	xx	xx
Patil et al. [10]	3064	(VGG-16 + SCNN)	97.77	96.66	98.33	97.47
Proposed Model	4845	Xception + DenseNet-121	98.63	97.58	99.72	98.10
		ResNet-50 + DenseNet-121	98.49	97.58	99.44	98.03

The confusion matrix of Xception + DenseNet-121 and DenseNet-121 + ResNet is shown in Figure 8 (a) and (b), respectively.



(a) **(b)**

Figure 8. (a) Confusion Matrix of DenseNet-121 + Xception Hybrid Model**. (b)** Confusion Matrix of DenseNet-121 + ResNet-50 Hybrid Model.

The proposed combinations of DenseNet-121 + Xception and ResNet + DenseNet-121 achieved almost the same results, with minor differences. The sensitivities of both models are equal, and since sensitivity is one of the key parameters of our analysis, we can choose either the Xception or the ResNet combination. Xception's architecture is more complex than ResNet's. It takes longer, with more computational and training time and resources. On the other hand, ResNet has been widely implemented for clinical purposes and for the tumour detection task. Based on the conducted literature review and experimental evaluation, the performance metrics of the ResNet-50 and DenseNet-121 ensemble match well with the requirements of the proposed research and are considered the final ensemble model for brain tumour detection.

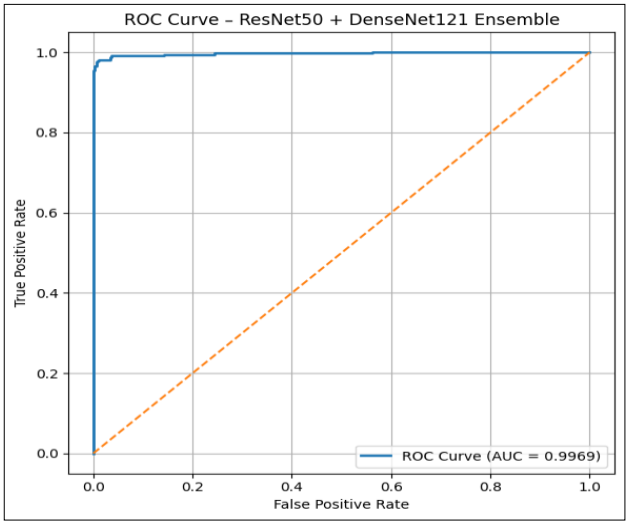


Figure 9. ROC Curve for ResNet-50 + DenseNet-121 Ensemble Model.

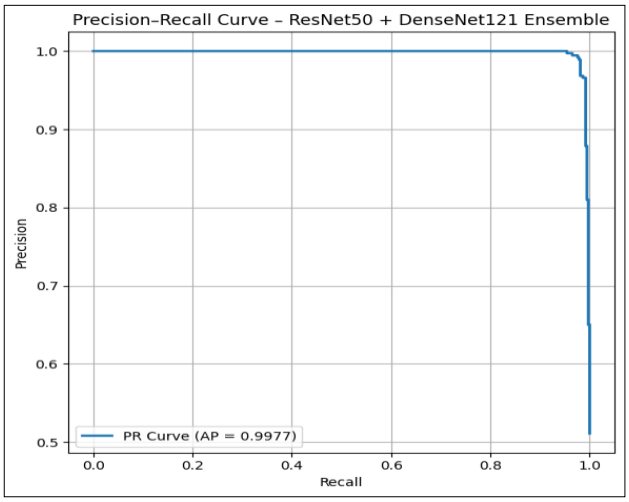
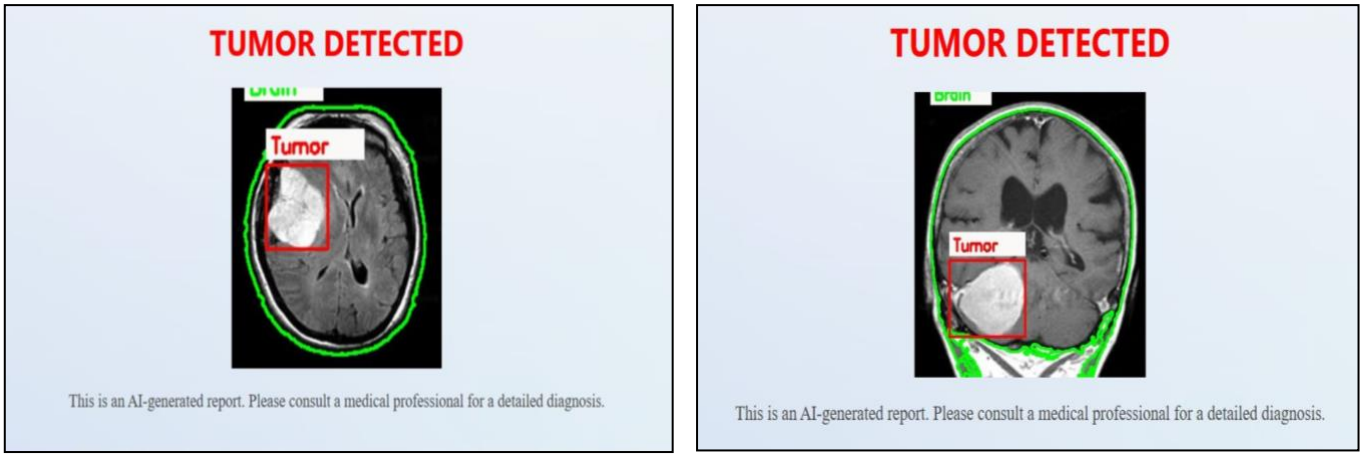


Figure 10. Precision Recall Curve for ResNet-50 + Densenet-121 Ensemble Model.

The final selected ensemble model of Densenet-121 and Resnet-50 achieved an ROC-AUC of 0.9969 and an average precision (AP) of 0.9977 on the test dataset. The precision-recall curve demonstrates consistent high precision across most recall values, indicating strong robustness in detecting tumour cases while minimising false positives.

4.2 Model's Prediction

The proposed hybrid ensemble model applies pre-processing to the uploaded MRI image, removing the area outside the brain using the cropping technique described in the pre-processing section. Then the image is passed to the model, which predicts whether it contains a "tumour" or "no tumour". If no tumour is detected, the result is displayed as "No Tumor Detected"; if a tumour is detected, the model applies post-processing techniques to identify the area where the tumour is located. The system visually highlights the tumour region by detecting pixel-intensity variations within the brain. The largest contour in the image is highlighted in green and labelled 'Brain'. The tumour region is enclosed within a red bounding box labelled 'Tumour'. Figures 11 and 12 show the predictions of the final model.



(a) **(b)**

Figure 11. (a) Model prediction on tumorous test images. **(b)** Model prediction on another tumorous test image.

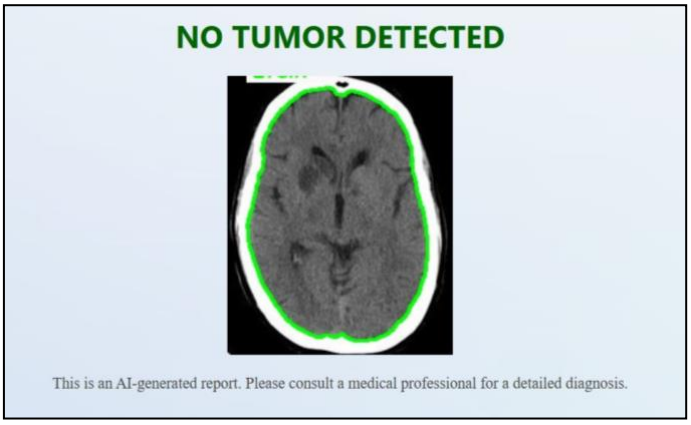


Figure 12. Model prediction on a healthy test image.

5. Challenges and Limitations

During the research and deployment phase of the proposed model, several challenges were encountered. The best model trained on the image dataset requires a system with a powerful CPU with 24 cores, 32 threads, and a clock speed above 5GHz. GPU Video RAM should be greater than 8GB. For

Memory specs, it is advisable to have at least twice as much memory as VRAM in the laptop GPU[30]. Another challenge relates to the lack of publicly available three-dimensional (3D) healthy brain MRI datasets. Models developed for medical purposes are effectively trained when the provided dataset is 3D.

There are many sources online that provide 3D datasets of tumorous images; however, 3D MRI images of the healthy brain are not publicly available. Furthermore, a brain tumour dataset for different tumour types is publicly available online. However, datasets dedicated to different stages of brain tumours are not yet publicly available. For the development of an AI model that can detect tumours at an early stage, training and validation on an early-stage dataset are crucial. The availability of a stage 1 or stage 2-labelled dataset would fill a major gap in the development of AI models for early-stage tumour detection in the future. Additionally, the proposed model was trained and evaluated using a single publicly available MRI dataset, which may limit its generality. The absence of external validation may limit the application of the model across diverse medical cases.

Future work will focus on testing and validating the model using a dataset with greater diversity and on extending the work to multiclass tumour classification.

6. Conclusion

Early detection of brain tumours is an important step in timely patient treatment. Automated detection of tumours using AI models helps in this regard. Models with high accuracy and sensitivity can detect fine and subtle features in brain tumour MRI images. This paper proposed a methodology for creating a hybrid model using two deep convolutional neural networks. The two models are combined using an averaging technique in ensemble learning. Several deep learning models were implemented to select the best model for the final ensemble. ResNet-50 and DenseNet-121 were found to be the best-performing architectures in terms of evaluation metrics and as cost-effective solutions. The hybrid model combines the advantages of skip connections in ResNet-50, which helped train deep neural networks without the vanishing gradient problem, and the dense connectivity of DenseNet-121, which enhanced gradient flow during training. Our proposed model achieved an accuracy of up to 98.49% and sensitivity up to 97.58%. However, the final model has not been validated on external or early-stage datasets; this may affect its generalisation to diverse populations or imaging conditions. The proposed work in this paper and the model's performance highlight its potential as a supportive tool that could contribute to future research focused on early brain detection.

Data Availability Statement: The dataset used in this study is publicly available from Kaggle: Brain Tumor Classification (MRI) Dataset.

Funding: The authors declare that no funding was received for this research.

Declaration of Competing Interest: The authors declare they have no known competing interests.

Research Involving Human/Animal Participants: This study used a publicly available, anonymized MRI dataset and did not involve human participants or animals; therefore, ethical approval and informed consent were not required.

References

[1] M. Mumtaz Zahoor, S. A. Qureshi, S. Hussain Khan, and A. Khan, "A New Deep Hybrid Boosted and Ensemble Learning-based Brain Tumor Analysis using MRI," in "arXiv e-prints," 2022.

[2] M. Clinic. "Brain tumor – Diagnosis and treatment." https://www.mayoclinic.org/diseases-conditions/brain-tumor/diagnosis-treatment/ (accessed.

[3] S. Chhabda, O. Carney, F. D'Arco, T. S. Jacques, and K. Mankad, "The 2016 World Health Organization classification of tumours of the central nervous system: what the paediatric neuroradiologist needs to know," *Quantitative Imaging in Medicine and Surgery,* vol. 6, no. 5, pp. 486-489, 2016, doi: https://doi.org/10.21037/qims.2016.10.01.

[4] E. I. Zacharaki *et al.*, "Classification of brain tumor type and grade using MRI texture and shape in a machine learning scheme," *Magnetic Resonance in Medicine,* vol. 62, no. 6, pp. 1609-1618, 2009, doi: https://doi.org/10.1002/mrm.22147.

[5] P. Abdolmaleki, F. Mihara, K. Masuda, and L. D. Buadu, "Neural networks analysis of astrocytic gliomas from MRI appearances," *Cancer Letters,* vol. 118, no. 1, pp. 69-78, 1997/09/16/ 1997, doi: https://doi.org/10.1016/S0304-3835(97)00233-4.

[6] M. I. Razzak, S. Naz, and A. Zaib, "Deep Learning for Medical Image Processing: Overview, Challenges and the Future," in *Classification in BioApps: Automation of Decision Making*, N. Dey, A. S. Ashour, and S. Borra Eds. Cham: Springer International Publishing, 2018, pp. 323-350. 10.1007/978-3-319-65981-7_12

[7] E. Aarthi *et al.*, "Detection and classification of MRI brain tumors using S3-DRLSTM based deep learning model," *International Journal of Electrical and Electronics Research,* vol. 10, no. 3, pp. 597 - 603, 2022, doi: https://doi.org/10.37391/IJEER.100331.

[8] A. u. Haq, J. P. Li, S. Khan, M. A. Alshara, R. M. Alotaibi, and C. Mawuli, "DACBT: deep learning approach for classification of brain tumors using MRI data in IoT healthcare environment," *Scientific Reports,* vol. 12, no. 1, p. 15331, 2022, doi: https://doi.org/10.1038/s41598-022-19465-1.

[9] S. Alsubai, H. U. Khan, A. Alqahtani, M. Sha, S. Abbas, and U. G. Mohammad, "Ensemble deep learning for brain tumor detection," (in English), *Frontiers in Computational Neuroscience,* Original Research vol. 16, pp. 1 - 14, 2022, doi: https://doi.org/10.3389/fncom.2022.1005617.

[10] S. Patil and D. Kirange, "Ensemble of Deep Learning Models for Brain Tumor Detection," *Procedia Computer Science,* vol. 218, pp. 2468-2479, 2023, doi: https://doi.org/10.1016/j.procs.2023.01.222.

[11] N. T. Singh, P. Kaur, A. Chaudhary, and S. Singla, "Detection of Brain Tumors Through the Application of Deep Learning and Machine Learning Models," in *2023 IEEE 8th International Conference for Convergence in Technology (I2CT)*, 2023, pp. 1-6, doi: https://doi.org/10.1109/I2CT57861.2023.10126474.

[12] S. Asif, W. Yi, Q. U. Ain, J. Hou, T. Yi, and J. Si, "Improving Effectiveness of Different Deep Transfer Learning-Based Models for Detecting Brain Tumors From MR Images," *IEEE Access,* vol. 10, pp. 34716-34730, 2022, doi: https://doi.org/10.1109/ACCESS.2022.3153306.

[13] S. R. Borra, M. K. Priya, M. Taruni, K. S. Rao, and M. S. Reddy, "Automatic Brain Tumor Detection and Classification Using UNET and Optimized Support Vector Machine," *SN Computer Science,* vol. 5, no. 5, p. 540, 2024, doi: https://doi.org/10.1007/s42979-024-02881-7.

[14] N. Ghaffar Nia, E. Kaplanoglu, and A. Nasab, "Evaluation of artificial intelligence techniques in disease diagnosis and prediction," *Discover Artificial Intelligence,* vol. 3, no. 5, pp. 1 - 14, 2023, doi: https://doi.org/10.1007/s44163-023-00049-5.

[15] L. Alzubaidi *et al.*, "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," *Journal of Big Data,* vol. 8, no. 53, pp. 1 - 74, 2021, doi: https://doi.org/10.1186/s40537-021-00444-8.

[16] S. Rathore, T. Rana, U. Mittal, T. Gupta, N. Malik, and A. S. M. S, "Brain Tumor Detection by using ResNet-50 and Image Processing Tools," in *2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, 12-13 May 2023 2023, pp. 480-483, doi: https://doi.org/10.1109/ICACITE57410.2023.10183081.

[17] O. N. Belaid, M. Loudini, and A. Nakib, "Brain tumor classification using DenseNet and U-net convolutional neural networks," in *2024 8th International Conference on Image and Signal Processing and their Applications (ISPA)*, 21-22 April 2024 2024, pp. 1-6, doi: https://doi.org/10.1109/ISPA59904.2024.10536704.

[18] A. Sinha, R. P. A, M. Suresh, N. M. R, D. A, and A. G. Singerji, "Brain Tumour Detection Using Deep Learning," in *2021 Seventh International conference on Bio Signals, Images, and Instrumentation (ICBSII)*, 25-27 March 2021 2021, pp. 1-5, doi: https://doi.org/10.1109/ICBSII51839.2021.9445185.

[19] Sartaj. "Brain Tumor Classification (MRI) Dataset." https://www.kaggle.com/datasets/sartajbhuvaji/brain-tumor-classification-mri (accessed.

[20] A. B. Abdusalomov, M. Mukhiddinov, and T. K. Whangbo, "Brain Tumor Detection Based on Deep Learning Approaches

and Magnetic Resonance Imaging," *Cancers,* vol. 15, no. 16, p. 4172, 2023.
[21] CloudThat. "Image Data Augmentation Using Keras API ImageDataGenerator." https://www.cloudthat.com/resources/blog/image-data-augmentation-using-keras-api-imagedatagenerator (accessed.
[22] E. Gomede. "The Role and Importance of Feature Extraction in Machine Learning." https://medium.com/the-modern-scientist/the-role-and-importance-of-feature-extraction-in-machine-learning-0bb7ebc3b289 (accessed.
[23] K. R. ÇETİN. "The Importance of Splitting Datasets into Training, Validation, and Test Sets." Medium. https://ruveydakardelcetin.medium.com/the-importance-of-splitting-datasets-into-training-validation-and-test-sets-417caaeae91d (accessed.
[24] s.-l. developers. "train_test_split — scikit-learn documentation." https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.train_test_split.html (accessed.
[25] Keras. "Adam optimizer — Keras API." https://keras.io/api/optimizers/adam/ (accessed.
[26] GeeksforGeeks. "Ensemble Learning — A Comprehensive Guide." https://www.geeksforgeeks.org/a-comprehensive-guide-to-ensemble-learning/ (accessed.
[27] M. S. Boudrioua, "COVID-19 detection from chest X-ray images using CNNs models: Further evidence from Deep Transfer Learning," *The University of Louisville Journal of Respiratory Infections,* vol. 4, no. 1, p. 53, 2020, doi: https://doi.org/10.18297/jri/vol4/iss1/53.
[28] A. Sawle, S. Bhosale, S. Abhang, P. Ghagare, and D. Argade, "Brain Tumor Detection Using Deep Learning," *International Journal for Research in Applied Science & Engineering Technology (IJRASET),* vol. 12, no. 3, pp. 2402-2408, 2024, doi: https://doi.org/10.22214/ijraset.2024.59360.
[29] S. Saeedi, S. Rezayi, H. Keshavarz, and S. R. Niakan Kalhori, "MRI-based brain tumor detection using convolutional deep learning methods and chosen machine learning techniques," *BMC Medical Informatics and Decision Making,* vol. 23, no. 16, pp. 1 - 17, 2023, doi: https://doi.org/10.1186/s12911-023-02114-6.
[30] ASUS. "How to Choose a Computer for AI and Machine Learning Work?" https://www.asus.com/content/how-to-choose-a-computer-for-ai-and-machine-learning-work/ (accessed.