

MobDenseNet: A hybrid deep learning model for brain tumor classification using MRI

Meher Afroj ^a, M. Rubaiyat Hossain Mondal ^{a,*} , Md Riad Hassan ^{a,b}, Sworna Akter ^a

^a Institute of Information and Communication Technology, Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh

^b Department of Computer Science and Engineering, Green University of Bangladesh, Dhaka, Bangladesh

ARTICLE INFO

Keywords:
 CNN
 Deep learning
 Hybrid model
 MRI
 Cross-dataset evaluation

ABSTRACT

This paper presents MobDenseNet, an improved deep learning model that assists medical practitioners in diagnosing brain tumors accurately. The proposed MobDenseNet is developed using the concepts of existing deep learning models: MobileNetV1 and DenseNet; the model incorporates hyperparameter fine-tuning and feature fusion ensemble during the feature extraction phase, consolidating layers like batch normalization, dense layers in the classification step to classify brain tumors. The classification is done into multiple classes, including, gliomas, meningiomas, pituitary, and healthy brain. The performance of the proposed model is assessed on two benchmark datasets. The experiments consider 2757 training and 307 testing images for the first dataset of 3064 MRI images, available on Figshare, having classes of glioma, meningioma, and pituitary. The experiment for the second dataset, has 2937 training and 327 testing images with glioma, meningioma, pituitary, and no tumor classes. The model achieves 98.4 % accuracy, 99.9 % AUC, 98.6 % precision, 98.40 % recall, 98.5 % F1-score for the Figshare dataset, and 96.02 % accuracy, 99.4 % AUC, 96.3 % precision, 95.7 % recall and 95.9 % F1-score for the Sartaj Bhuvaji dataset, respectively. The proposed MobDenseNet shows better accuracy than the existing models considered in the research. To demonstrate the effectiveness of the proposed model on diverse and unseen data, cross-dataset evaluations are conducted, where the model is trained using the Figshare dataset and tested using the Sartaj Bhuvaji and two additional datasets. Results indicate that even for the cross-dataset scenario, the proposed model achieves acceptable classification accuracy and outperforms existing models of MobileNetV1 and DenseNet.

1. Introduction

Brain tumors, which pose a severe risk to the body, are caused by a slow and steady growth of cells in the brain and its surrounding tissues. About 300,000 new instances of brain tumors are diagnosed annually, according to The World Health Organization (WHO) [1]. Malignant gliomas and meningiomas are the most serious brain tumors, whereas pituitary tumors are benign in general but can lead to other medical issues. There is still more work to be done in classifying brain tumors. MRI (Magnetic Resonance Imaging) and CT scan (Computed Tomography) are used to define the size, position, and shape of the tissues. However, it takes a lot of time and expertise of the radiologists and physicians to analyze these images to diagnose brain tumors.

There is progress in machine learning (ML), specifically in deep learning (DL), in the context of automated pattern recognition and classification of diseases. Convolutional neural networks (CNNs), with

their multiple layers and high diagnostic accuracy, are now the most successful techniques employed in medical image classification required for disease detection [2–6].

DL models may struggle to get robust and generalizable representations due to a large variation in shape, size, location and intensity of the brain tumor images. Such challenges can be addressed with different strategies, like data augmentation, transfer learning, ensemble methods, and attention mechanisms. Transfer learning can alleviate the problem of limited data and improve the model's overall performance [7,8]. To raise the size and variety of the training data, different data augmentation techniques, such as rotation, flipping, scaling, zooming, light condition etc., can be engaged that also improve the model's ability to handle different types of data in brain tumor characteristics [9,10].

To enhance classification performance, studies have introduced ensemble methods combining multiple individual DL techniques, optimizing their strength while minimizing weaknesses [12,14]. Recent

* Corresponding author.

E-mail address: rubaiyat97@iict.buet.ac.bd (M.R.H. Mondal).

Table 1
Summary of Sections 2.1 and 2.2.

Type	Ref	Method	Dataset	Classes	Accuracy	Remarks
Models with Basic CNN	[14]	2D CNN and convolutional autoencoder	3264 MR images	4	93.4 %	Not very impressive accuracy but less execution time
	[15]	CNN with four different optimizers	MRI images	2, 3, 4	98 %, 92 %, 95 %	Multiclass accuracy can be improved
	[16]	Multi-pathway CNN	3064 TIW-CE MRI images	3	97.3 %	The training procedure is quite expensive
	[17]	23 layers DL model	152 MRI, 3064 MRI images	2, 3	100 %, 97.8 %	To resolve the overfitting issue of small dataset models, complexity is increased
Models with Transfer Learning	[18]	Harris Hawks Optimized CNN	253 MR images	2	98 %	can be tested with a large dataset
	[20]	ResNet50 with SPP	3060 MR images	2	99.02 %	SPP can be computationally expensive.
	[21]	fine-tuned Resnet50 and Inception V3	3459 MR images, 3000 MR images	4, 2	97.6 %, 99.8 %	dataset of four classes was not distributed evenly
	[23]	Pre-trained VGG16, VGG19 and Inception V3 with Aquila Optimizer	3060 MR images	2	97 %, 98.9 %, 98.5 %	Aquila optimizers are susceptible to local optima
	[24]	Xception, ResNet50V2, InceptionResNetV2 and DenseNet201	3064 TIW-CE MRI images	3	98.4 %, 99.6 %, 99.3 %, 98.7 %	Only three classes were considered
	[25]	DenseNet121 and ResNet50V2 with Cycle GAN	3074 MR images and synthetic 23568 MR images	3	95.9 %, 99.4 %, 98.72 %	Synthetic data may have lack of realism

approaches also combine MRI with other imaging modalities like CT, PET or clinical data to improve diagnostic accuracy [11,12]. Generative adversarial networks (GANs) are also used by researchers for data augmentation and the production of synthetic data to increase the number of images in the dataset [13].

Despite these advancements, reliable classification remains challenging due to variations in imaging features like rotation, zooming, tiling, and other unusual image orientations. Moreover, overfitting and class imbalance can have an impact on DL models, therefore the quantity and quality of training data are vital [9,15]. Datasets are frequently acquired utilizing varied protocols, devices, settings, and patient groups, resulting in variances in data quality and distribution. As a result, DL models frequently fail to generalize well when tested on datasets other than those used for training. Therefore, cross-dataset evaluation can result in more trustworthy and generalizable DL models in health informatics. To ensure model robustness and generalizability on unseen, diverse datasets, cross-dataset evaluations are critical. This research focuses on the development of a hybrid DL model and then applies the model for cross-dataset evaluation.

The research contributions are listed below.

- (1) A novel hybrid architecture named MobDenseNet is introduced by using the concepts of DenseNet121 and MobileNetV1 framework with some batch normalization layers, dense layers, and hyperparameter optimization to distinguish brain tumors using brain MRI images of two benchmark datasets, one with three classes and another with four classes.
- (2) The performance of MobDenseNet model is evaluated and compared with some state-of-art DL models reported in the literature review in terms of accuracy, AUC, loss, precision, recall and F1-score.
- (3) Cross-dataset evaluations are carried out to illustrate the efficacy of the proposed model. We use the proposed MobDenseNet model to train a dataset and use the images from another dataset for testing purposes.

The discussions are organized in the following order. A review of the literature is presented in Section 2, materials and techniques are in Section 3, an overview of the model architecture is described in Section 4, and experimental findings are in Section 5. The discussion is in Section 6, the limitation is in Section 7, and the conclusion is in Section 8.

2. Related works

Given the present success of DL networks in the classification of medical images, including various forms of brain tumors, numerous

studies have been carried out [3–13]. The current research is discussed in four areas: 1) DL models with basic CNN [14–19], 2) DL models with transfer learning [20–27], 3) models with combination of ML and DL [28–33], 4) hybrid DL models [34–39].

2.1. DL models with basic CNN

Saeedi et al. [14] investigated 3264 MRI brain images across four classes using 2D CNN and a convolutional auto-encoder. The testing accuracy of the proposed 2D CNN is 93.44 % which outperformed the other six ML algorithms (K-NN, RF, SVM, LR, SGD, MLP) with faster execution time. One study [15] simulated a CNN to perform both binary and multiclass classification investigating four different optimizers. Adam optimizer achieved the highest binary accuracy of 98 % where Nesterov momentum performed the best with 92 % accuracy for three classes and 95 % accuracy for four classes. A multi-pathway CNN architecture was proposed by Francisco et al. [16] for automatic brain tumor segmentation and classification of three classes on 3064 MRI images and achieved 97.3 % accuracy. Shaikat et al. [17] applied 23 convolutional layer DL model on both binary and multiclass dataset achieving 100 % and 97.8 % classification accuracies, respectively. Sarah et al. [18] developed a network of CNN named Harris Hawks Optimized Convolutional Neural Network to identify tumor region of a dataset consisting of 253 MR images having two classes, obtaining a classification accuracy of 98 %. Xu et al. [19] used an image processing technique named active contouring with texture analysis to outline and detect the suspicious tumor region on a dataset of over 7000 MRI images of four classes. The technique performed well with 92 % AUC compared to the R-CNN technique.

2.2. Deep models with transfer learning

Karrar et al. [20] utilized fine-tuned ResNet50 with Spatial Pyramid Pooling (SPP) layer to handle different sized images and capture multi-scale images on a Kaggle dataset of two class (tumor, no-tumor), achieving an accuracy of 99.02 % outperforming other DL models (VGG16, AlexNet, GoogleNet). Marco et al. in one study [21] investigated 7023 MRI images of four classes using seven different CNN models such that Generic CNN, InceptionV3, ResNet50, InceptionResNetV2, MobileNetV2 Xception, and EfficientNetB0 where the accuracy varied from 81 % to 97 %. Sayed et al. [22] proposed two fine-tuned methods, ResNet50 and Inception V3, for binary and multiclass classification with different hyperparameter optimizations and found that pairing of ResNet50 with Nadam optimizer achieves the highest accuracy. To classify tumor and no tumor, Amena Mahmoud et al. [23] performed three pre-trained models, VGG16, VGG19, and Inception V3, with

Table 2
Summary of sections 2.3 and 2.4.

Type	Ref	Method	Dataset	Classes	Accuracy	Drawback
Com-binedML and DL Models	[28]	CNN, VGG19, CNN-SVM	14712 MR images, 5266 MR images	2, 3	87 %, 40 %, 98.31 % and 92 %, 93.31 %, 96.8 %	More improvement is required in CNN-SVM
	[29]	EfficientnetB0 and VGG-19 with cubic SVM	3000 MR images	2	99.7 %	Binary classification
	[30]	CNN, R-CNN, SVM-RBF	2764 MRI and 3064 TIW-CE MRI images	3	98.3 %, 98 %	RBF kernel is suitable for medium dimensional data
	[31]	AlexNet, ResNet-18 with SVM	3060 MRI images	4	95 %, 91 %	SVM cannot handle noisy data
	[32]	DenseNet-169, ShuffleNet V2, MnasNet, Inception V3, ResNeXt-50 with KNN, SVM, RF etc.	253, 3000, 3064 MR images	2, 3	98.5 %	Complex architecture
	[34]	FCM clustering with SVM	3064 TIW-CE MRI	3	98.2 %	Might have sensitivity to parameter settings
	[35]	PDCNN with SVM, K-NN, NB and Decision Tree	253, 3064, 2870 MRI images	2, 3, 4	98.6 %, 98.1 %, 98.3 %	—
	[36]	ADE_DieT	5249 MRI images	4	96.09 %	Limited generalizability
	[37]	ensemble convolutional neural network (EDCNN)	3064 TIW-CE MRI	3	97.7 %	Increase parameters
	[38]	Five DL models and an ensemble model	1430 T1W-MRI, 897 T2W-MRI, 1240 FLAIR-MRI	2	97.2 %	All datasets are of two classes
Deep Hy-brid Models	[39]	Feature fusion using InceptionV3, DenseNet201 and radiomics features and classification using PSO-KELM	3064 TIW-CE MRI, 3264 MR images	3, 4	98.2 %	—
	[40]	Ensemble IVX16	3264 MR images	4	96.9 %	—
	[41]	Hybrid model using EfficientNetB0 and ResNet-50	3064 TIW-CE MRI images	3	98.9 %	Excessive enhancement can deteriorate Average information
	[42]	MASK RCNN using ResNet-50 and DenseNet-41	3064 TIW-CE MRI, 253 MR images	3, 2	98.3 %, 97.7 %	Model may not work better for small dataset

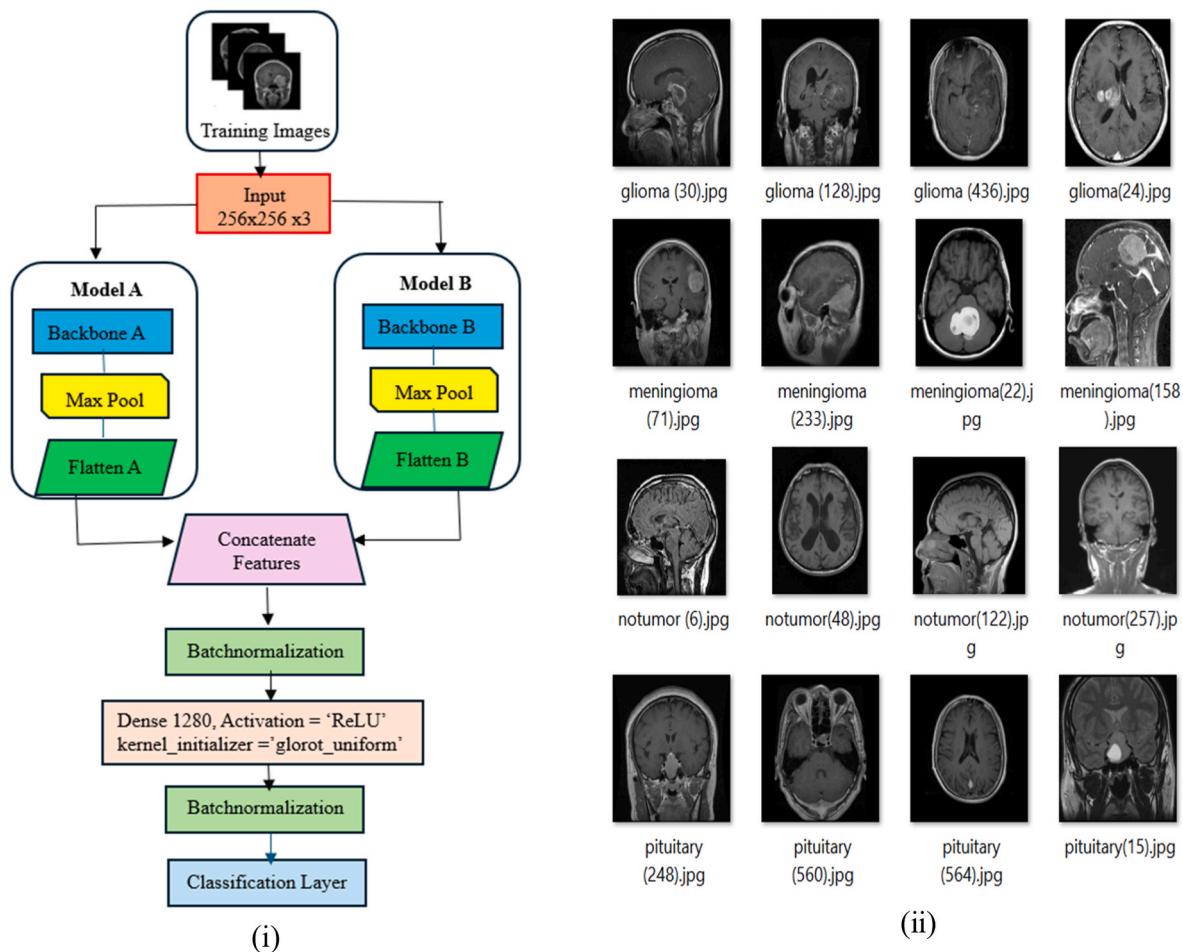


Fig. 1. (i) The proposed MobDenseNet architecture. (ii) Samples of brain MRI images.

Table 3
Distribution table of brain MRI images in each dataset

Dataset	Glioma	Meningioma Class 1	Pituitary	Notumor	Total
	Class 0		Class 2	Class 3	
Dataset-1 (Figshare dataset) [43]	1420	708	930	–	3064
Dataset-2 (Sartaj-Bhuvaji dataset) [44]	926	937	901	500	3264
Dataset-3 (Br35 dataset) [45]	300	300	300	–	900
Dataset-4 (Crystal brain MRIs) [46]	300	300	300	–	900

Table 4
Parameters in data augmentation

Type of Augmentation	Value setting
Rescaling	1/255
Rotation range	15
Width shift range	0.2
Height shift range	0.2
Zoom range	0.2
Horizontal flip	True
Vertical flip	True
Fill mode	Nearest

Table 5
List of hyper-parameter values in the study

Hyper-Parameter name	Value setting
Activation Function	ReLU
Kernel Initializer	Glorot uniform
Initial Learning rate	0.002
Optimizer	Nadam
Batch Size	64
Epochs	80
Train-test split	90-10 %
Output activation Function	Softmax
Output kernel initializer	Random uniform

Aquila Optimizer and achieved the highest accuracy of 98.95 % from VGG19. Alamin et al. [24] used four transfer learning architectures on multiclass datasets with the accuracy of models varying from 98.72 % to 99.68 %, and the best accuracy was achieved from Resnet50V2. To address the problem of limited datasets, Yapici et al. [25] presented an approach in which synthetic data is generated using Cycle GAN and used to train a fine-tuned DenseNet121 for multiclass classification and mean accuracy increased from 0.9595 to 0.9947. Sohaib et al. [27] investigated five fine-tuned DL frameworks: DenseNet201, DenseNet121, ResNet152V2, Xception, and InceptionResNetV2 on two different datasets. Xception architecture achieved an excellent classification accuracy of 99.67 % for the 3-class dataset and 95.87 % for the 4-class dataset. A summary of Sections 2.1 and 2.2 is given in Table 1.

2.3. Deep models with a combination of ML and DL

Shweta et al. [28] utilized CNN, VGG19, and CNN-SVM (the model uses CNN to extract deep features and SVM for classification) models to classify multiclass MR images of two datasets. CNN, VGG19, and CNN-SVM showed accuracy values of 87 %, 40 %, and 98.31 %, respectively, on dataset 1 and 92 %, 93.31 %, and 96.8 %, respectively, on dataset 2. Bourennane et al. [29] combined two different CNN architectures: EfficientNetB0 and VGG-19 to extract features from 3000 MR images of binary class and obtained 99.78 % classification accuracy with cubic SVM. Ejaz et al. [30] performed basic CNN to extract features, R-CNN to localize tumor region and finally CNN and SVM-RBF algorithm is used to classify the MRI images with around 98 % accuracy on both datasets. Similarly, Ebrahim et al. [31] achieved 95.10 % accuracy with AlexNet and SVM, Jaeyong Kang et al. [32] achieved 98.5 % accuracy using different pretrained models (ShuffleNet V2, DenseNet-169 and MnasNet) and SVM-RBF classifier on four class datasets. Ullah et al. [33] extracted features by discrete wavelet transform and used advanced DNN to classify brain MRI images as normal or pathological with 95.8 % accuracy. Alkahtani et al. [34] utilized Fuzzy C-Means (FCM) clustering and Support Vector Machine (SVM) classifier to process contrast enhancement on a dataset of 3064 MRI images and ensured 98.2 % accuracy in classifying brain tumors as glioma, meningioma and pituitary. Takwa Rahman et al. [35] introduced a dilated parallel deep CNN (PDCNN) utilizing parallel processing paths with variable dilation rates to extract both fine and coarse image characteristics. The study considered multiple classifiers such as SVM, K-NN, NB and Decision Tree to classify images from multiple datasets. Reported accuracy was 98.6 % for binary classification, 98.1 % for dataset-2 and 98.3 % for the

Table 6
Comparison of proposed models with separate models on Dataset-1

Models with input image size	Accuracy	AUC	Loss	Precision	Recall	Inference time (in miliseconds)
MobileNetV1 (224 × 224)	0.928	0.987	0.221	0.931	0.925	108.523 ms
Xception (256 × 256)	0.908	0.976	0.300	0.908	0.902	283.287 ms
DenseNet121 (224 × 224)	0.928	0.984	0.234	0.922	0.921	280.662 ms
InceptionV3 (256 × 256)	0.941	0.991	0.176	0.950	0.934	210.440 ms
VGG19 (224 × 224)	0.876	0.971	0.315	0.887	0.868	825.781 ms
ResNet101V2 (256 × 256)	0.895	0.969	0.342	0.880	0.863	456.275 ms
Proposed MobDenseNet (256 × 256)	0.984	0.999	0.058	0.986	0.984	308.340 ms

Table 7
Comparison of proposed model with separate models on Dataset-2

Models with input image size	Accuracy	AUC	Loss	Precision	Recall	Inference time (in miliseconds)
MobileNetV1 (224 × 224)	0.938	0.992	0.201	0.941	0.932	116.895 ms
Xception (256 × 256)	0.889	0.977	0.343	0.906	0.862	285.484 ms
DenseNet121 (224 × 224)	0.917	0.989	0.250	0.918	0.912	264.340 ms
InceptionV3 (256 × 256)	0.918	0.977	0.343	0.915	0.893	228.422 ms
VGG19 (224 × 224)	0.828	0.975	0.530	0.836	0.798	808.361 ms
ResNet101V2 (256 × 256)	0.908	0.985	0.266	0.915	0.985	477.291 ms
Proposed MobDenseNet (256 × 256)	0.960	0.994	0.177	0.963	0.957	336.950 ms

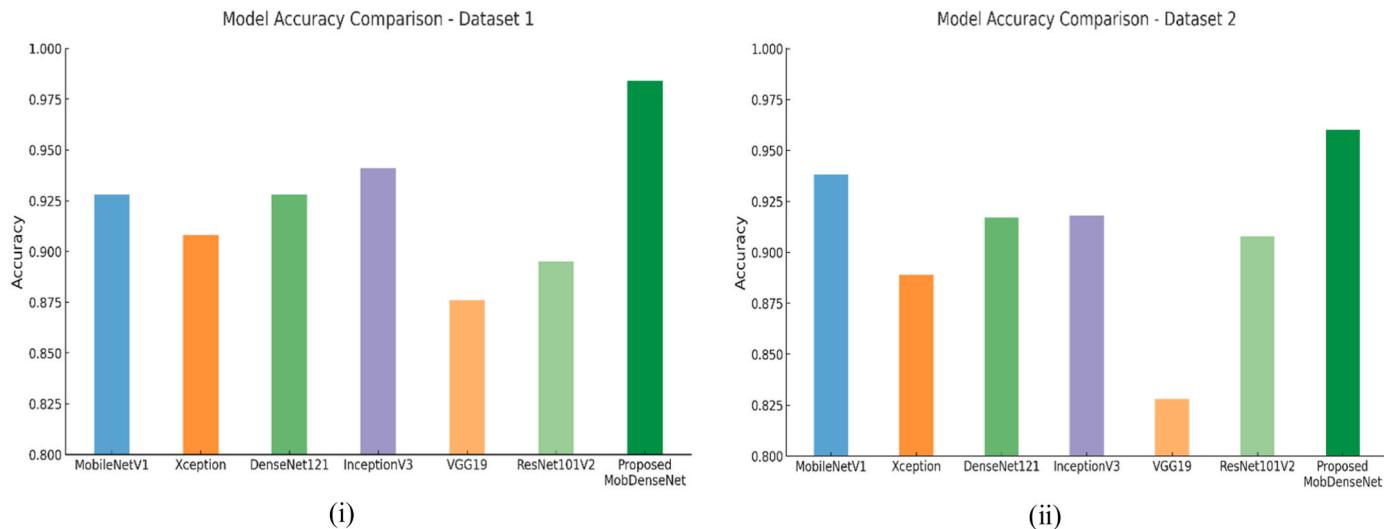


Fig. 2. (i): Comparison of classification accuracy values of the proposed model along with other models for Dataset-1. **Fig. 2** (ii): Comparison of classification accuracy values of the proposed model along with other models for Dataset-2.

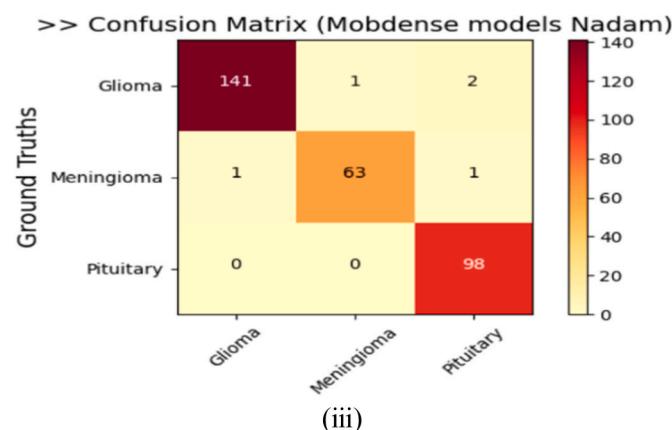
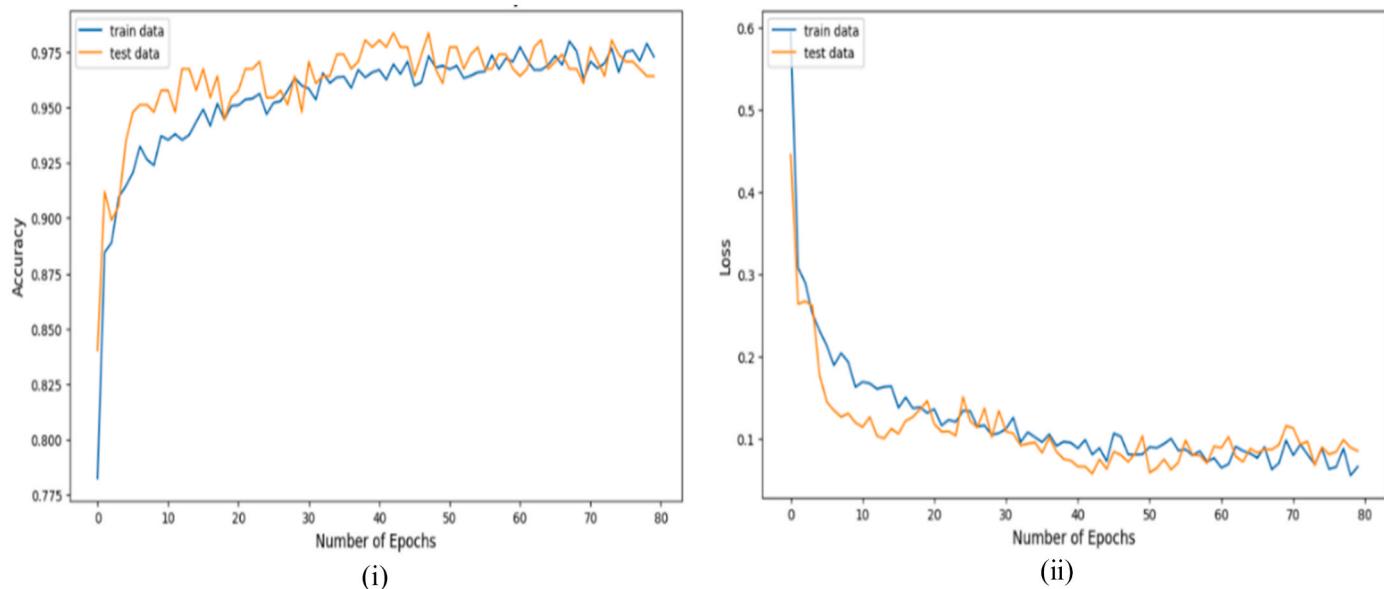


Fig. 3. Training and Testing analysis within 80 epochs on Dataset-1 (Figshare Dataset): (i) Accuracy survey, (ii) Loss survey, (iii) Confusion matrix of proposed MobDenseNet.

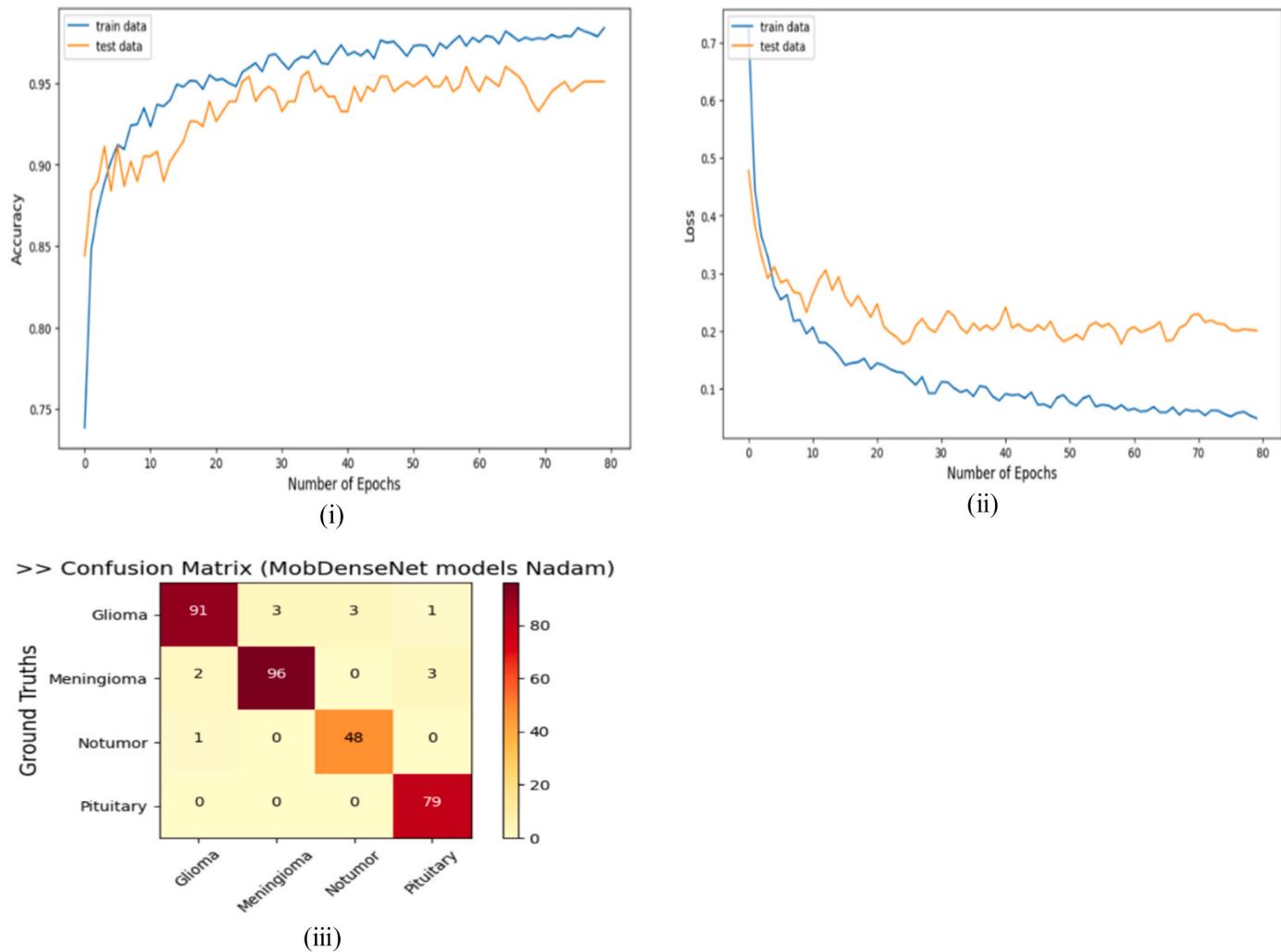


Fig. 4. Training and Testing analysis within 80 epochs on Dataset-2 (SartajBhuvaji Dataset): (i) Accuracy survey, (ii) Loss survey, (iii) Confusion matrix of proposed MobDenseNet.

Table 8

False Positive Rate (FPR) and False Negative Rate (FNR) analysis of the proposed model on both datasets

Dataset	Dataset1 (Figshare)				Dataset2 (Sartaj Bhuvaji)			
Class	Glioma	Meningioma	Pituitary		Glioma	Meningioma	No-tumor	Pituitary
FPR (%)	0.61	0.41	1.43		1.33	1.36	1.12	1.67
FNR (%)	2.08	3.07	0.00		4.21	4.95	2.04	0.00

dataset-3. Amin et al. [36] proposed an advanced model ADE_DieT integrating the DieT Transformer with Principal Component Analysis and the Adaptive Differential Evolution algorithm. This method outperformed other pre-trained models like MobileNetV3 and ResNet50 with 96.09 % accuracy in classifying four types of brain tumors on a Kaggle dataset.

2.4. Deep hybrid models

In the literature, multiple experiments [37,39,41,42] were done on the dataset consisting of 3064 MRI images of three classes (glioma, meningioma, pituitary). Suraj Patil et al. [37] provides an idea of a deep ensemble CNN (EDCNN) to perform multiclass classification that achieved 97.7 % accuracy. Sandhiya et al. [39] fused features from InceptionV3 and DenseNet201 with radiomics features and performed classification by using PSO-KELM, achieving an accuracy of 98.2 %. After enhancing image quality using optimal contrast and nonlinear

techniques, Amir et al. [41] proposed a framework combining extracted features from EfficientNetB0 and ResNet50 using the partial least squares (PLS) method with 98.9 % classification accuracy. Masood et al. [42] developed and compared two architectures named Mask RCNN using backbones ResNet-50 and DenseNet-41 that were evaluated on two datasets. Gopal et al. [38] had done 18 experiments to compare the performance of five different DL models with a hybrid ensemble model for different datasets. In the study, the classification accuracy values of the individual models varied from 93 % to 96 % on three datasets but majority voting outperformed other models with an average accuracy of 97.2 % for FLAIR-MRI dataset. Combining three transfer learning models (InceptionV3, VGG16 and Xception), Shahriar et al. [40] proposed an ensemble model named IVX16 that gained a classification accuracy of 96.9 % on 3264 MRI images of four classes. A summary of Sections 2.3 and 2.4 is given in Table 2.

Table 9

Comparison of the proposed model with others for dataset1 (glioma, meningioma, pituitary)

Ref	Dataset	Method	Performance Metrics
[30]	Figshare dataset: 3064 MR images of 233 patients (glioma: 1426)	Deep CNN with SVM- RBF classifier FCM clustering with SVM	Accuracy: 98 % 98.2 %
[34]	Meningioma: 708	PDCNN with multiple classifiers	Accuracy: 98.1 %
[35]	Pituitary: 930	Ensemble Convolutional Neural Network (EDCNN)	Accuracy: 97.7 %
[37])	Extracted features from InceptionV3 and DenseNet201 and classify using PSO-KELM	Accuracy: 98.21 %
Proposed method		MobDense model	Accuracy: 98.40 %

Table 10

Comparison of proposed model with others for dataset2 (glioma, meningioma, notumor, pituitary)

Ref	Dataset	Method	Performance Metrics
[14]	Kaggle sartaj bhuvaji dataset: 3264 MRI (826 of glioma, 937 of meningioma, 500 of no tumor, 901 pituitary)	2D CNN and a convolutional autoencoder with six ML models	Accuracy: 93.4 %
[27]		Xception, DenseNet201, DenseNet121, ResNet152V2, and InceptionResNetV2	Accuracy: 95.87 %
Proposed Method		MobDense model	Accuracy: 96.02 %

Table 11

Comparison of the proposed model with other work for several datasets

Dataset	Class	Ref.	Method	Accuracy
Kaggle Dataset: 253 MR tumor: 155, no tumor: 98 class) [49]	Binary classification	[26] 2022 [32] 2021 [42] 2021 —	Inception V3, ResNet-50, VGG16 DenseNet169 with RBF-SVM MASK RCNN	78 %, 95 %, 96 % 98.04 % 97.7 % 98.39 %
Kaggle Dataset: 3000 MR images of 2 classes tumor:1500, notumor:1500 [50]	Binary classification	[20] 2024 [22] 2023 [23] 2023 [29] 2024	ResNet50 with SPP ResNet50 with Nadam VGG-19, VGG-16, Inception V3 EfficientNetB0 and VGG-19 to extract features, cubic SVM classifier	99.02 % 99.8 % 98.9 %, 97 %, 98.5 % 99.7 %
Kaggle Dataset (subset of Sartaj dataset): 3060 MRI (826 of glioma, 937 of meningioma, 396 of notumor, 901 pituitary) [43]	Multiclass classification four classes	[31] 2022 —	AlexNet + SVM ResNet-18+SVM Proposed MobDenseNet	95 %, 91 % 95.59 %

3. Materials and methods

This section provides details about the availability of the datasets, followed by the preprocessing and augmentation methods applied to enhance model performance for the datasets. These steps help to ensure data quality and improve the robustness of the classification model.

3.1. Dataset availability

Nowadays numerous open-source brain MRI datasets are available on Kaggle repository. This study utilizes images from four datasets. Two benchmark datasets are used to train and evaluate the hybrid model using hold-out method, splitting the data into 90 % for training and 10 % for testing. This approach ensures a subset of data being reserved for testing the model's generalization to unseen data, allowing it to learn patterns and features more successfully. Dataset-1 (Figshare) contains 3064 MRI images with the categories of glioma as class 0 (1426 samples), meningioma as class 1 (708 samples) and pituitary as class 2 (930 samples) [43]. Dataset-2 (Sartaj Bhuvaji) contains a total of 3264 MRI images with the categories of glioma as class 0 (926 samples), meningioma as class 1 (937 samples), notumor as class 2 (500 samples) and pituitary as class 3 (901 samples) [44]. Fig. 1 shows some of the brain tumor MRI images from Dataset-1 as well as Dataset-2. Subsets of two additional dataset each having three classes is also used in the study to perform cross dataset evaluations on the hybrid model [45,46]. The distribution of images in each dataset is shown in Table 3.

3.2. Dataset preprocessing and augmentation

This section discusses the preprocessing techniques applied to the datasets to prepare the data samples for application of various DL models. Before extracting features, MRI images were resized to 256x256. Each pixel was normalized to a standard [0,1] range using a scaling factor of 1/255. Keras ImageDataGenerator class was used to convert 256x256 input dimensions to the desired input size of specific CNN's. As part of data augmentation, different techniques including rotation, scaling, flipping and translation were considered. These parameters were selected based on medical imaging standards and to keep consistency with the existing relevant literature. Table 4 presents the parameters used for data augmentation. Table 4 indicates a 15° rotation, along with width/height shifts (0.2) and zoom (0.2); these values are helpful for the model to adapt to slight variations in patient positioning and differences in scanning. Flipping images horizontally and vertically intensifies spatial invariance while the nearest fill mode avoids interpolation artifacts. In the experiments, we used 64 images in each batch over a total of 80 epochs. The selected parameter settings reduce overfitting by making natural variations that help the model to learn important tumor features rather than memorizing patterns. The hyperparameters were selected based on empirical testing and previous DL literature. Parameter values are adjusted based on validation accuracy, loss trends, and convergence behavior through manual fine-tuning. Table 5 lists the hyperparameters of the study.

4. Proposed architecture

In this section, the proposed model architecture will be discussed. The MobDenseNet model is a hybrid model that uses two pretrained models, DenseNet121 [47] and MobileNetV1 [48], to extract deep features from the magnetic resonance images, fuse them, and finally send the features to the classification layers. So, in the following subsection, the concept of the proposed MobDenseNet will be discussed.

4.1. Architecture of Proposed MobDenseNet

This subsection discusses the hybrid MobDenseNet model, shown in Fig. 1 (i), for image classification. The input images are RGB with a size

Table 12

Cross-dataset information with accuracy result

Train		Test		Accuracy					
Database	Images	Database	Images	MobDenseNet		Dense121Net		MobileNetV1	
				Train	Test	Train	Test	Train	Test
Figshare [43]	3064 (glioma, meningioma, Pituitary)	Sartaj [44] Br35 [45] [46]	540 900 900	97.9 % 98.1 % 97.7 %	91.4 % 91.8 90.8	92.5 93.2 92.8	87 % 88 % 84.1 %	92.8 93.7 93.2	88.2 % 89.7 % 85.6 %

of $256 \times 256 \times 3$. After some preprocessing, these images are fed into two parallel models to extract deep features; Model A (used pretrained DenseNet121 as backbone A) and Model B (used pretrained MobileNetV1 as backbone B) that are followed by max pooling layer, flatten layer, concatenation layer and classification layer. In the feature extraction phase, Model A utilizes dense connections among layers, allowing the reuse of feature maps from earlier layers to all subsequent layers to capture fine-grained, highly detailed patterns. Moreover, Model B leverages depthwise separable convolutions to learn efficient features while preserving complex cross-channel patterns. Max pooling layers (2×2 filter size, stride 2) reduce the spatial dimensions of extracted features by half in dual model architecture. These parallel extracted features are concatenated and passed through a fully connected layer for classification purposes. Our fully connected layer starts with a batch normalization layer, then a dense layer having 1280 filters with the activation function ‘ReLU’ and kernel_initializer as glorot_uniform, followed by another batch normalization layer. After all these settings, the classification layer is added, which has a dense layer. For Dataset-1, a dense layer with 3 filters, the activation function ‘softmax’, and kernel_initializer as random_uniform is added. The same settings are done for Dataset-2 with only a change in the number of filters added, as dataset 2 has four types of images to classify. This hybrid model combines the high-quality feature extraction of DenseNet121 with the efficient feature processing of MobileNet that creates a complementary architecture to significantly improve accuracy and generalization. Fig. 1(ii) shows samples of brain MRI images.

5. Experimental results

This section discusses the experimental results in terms of different evaluation metrics. Open-source Google Collaboratory is a cloud-based service that allows users to run Python code for experimental purposes utilized to implement the architectures. We used ML tools like Scikit-learn, Numpy, Matplotlib, Scipy, etc. Tables 6 and 7 compare the results of popular pre-trained models such as MobileNet, DenseNet, ResNet101V2, InceptionV3, VGG19 and Xception with proposed MobDenseNet on the Dataset-1 and Dataset-2 respectively. In Dataset-1, the model obtained 0.984 accuracy that surpasses existing stand-alone models with a notable AUC of 0.999 and minimal loss of 0.058. Similarly, the model attained 0.960 accuracy with an AUC of 0.994 in Dataset-2. In this study, we assess the performance of the model using the concept of confusion matrix and some evaluation metrics such as training and testing accuracy, recall, precision, loss, AUC and F1 score. Fig. 2 shows bar plots of the classification accuracy values of the proposed model and existing models. Fig. 2(i) is for the case of Dataset-1, while Fig. 2(ii) is for Dataset-2. To evaluate the model’s inference performance, the average inference time was calculated by considering 100 random predictions, as shown in Tables 6 and 7. While MobDenseNet takes a longer inference time (308.340 ms for Dataset-1 and 336.950 ms for Dataset-2) compared to its underlying backbone architectures, the considerable progress in accuracy as well as other performance metrics justified the additional processing requirements. Note that the inference time shown here corresponds to the Google Colaboratory service used, and it may vary depending on the hardware configuration.

In the following, the performance of the proposed model is discussed.

At the 80th epoch, the proposed MobDenseNet model shows superior results in accuracy, precision, specificity, and loss reduction, as shown in Figs. 3 and 4. Fig. 3 displays the training and testing analysis on Dataset-1. Fig. 3(i) shows the accuracy curve for both training and test data. On the y-axis, accuracy is displayed between 0 and 1 (or 0 %–100 %), while the x-axis shows the number of epochs, up to 80. The training curve rises quickly, hitting 0.92 before slowing down and fluctuating around 0.98. The test curve climbs rapidly to 0.95, then slows and fluctuates till 0.985. Fig. 3(ii) presents loss curve, with training loss starting at 0.6 and dropping to 0.05 while test loss starts at 0.5 and drops to 0.1. Throughout the training process, the train and test accuracy curves are closely aligned which indicates that the model generalizes well on this dataset. The loss curves also follow a similar trend, further confirming that overfitting is not a major concern in this case. It is good to alleviate false-positive cases and false-negative cases to prevent misclassification of disorders in medical research. The confusion matrix shown in Fig. 3 (iii) describes the TP (true-positive), FP (false-positive), TN (true-negative), and FN (false-negative) levels to show the number of observations that were either affected or erroneously detected. In this case only 3 images are misclassified for glioma, 1 as meningioma and 2 as pituitary, 2 images are misclassified for meningioma, 1 as glioma and 1 as pituitary while pituitary class gets the perfect classification.

The curves for training and testing analysis on dataset-2 are shown in Fig. 4. Fig. 4(i) reflects the accuracy curve that indicates training accuracy starting from 0.75 and reaching the peak around 0.98 at the end, while the highest testing accuracy reaches 0.96, which defines good performance on unseen data. In the loss curve in Fig. 4(ii), we may notice that both curves drop quickly in the first few epochs, but after that, training loss decreases steadily over time, and testing errors decrease with fluctuations. The accuracy and loss curves show a slightly larger gap between training and testing performance, but the overall trend remains stable. For this dataset, the model might capture dataset-specific patterns, but it still maintains a reasonable generalization capability. Fig. 4(iii) represents the confusion matrix where model misclassifies 7 images for glioma, 5 images for meningioma, 1 image for no tumor but correctly classifies all images for pituitary. A high FPR can lead to irrelevant medical interventions as it represents the proportion of negative cases that were misclassified as positive. Similarly, a high FNR indicates that the model is failing to detect a considerable number of true cases as it represents the percentage of actual positive cases that were incorrectly classified as negative. Table 8 describes the FPR and FNR analysis for each class of both datasets. The model attained low FPR (Glioma: 0.61 %, Meningioma: 0.41 %, Pituitary: 1.43 %), indicating few false alarms and low FNR (Glioma: 2.08 %, Meningioma: 3.07 %), minimal misclassification of these classes in Dataset-1. For Dataset-2, glioma, meningioma, no-tumor and pituitary show low FPR of 1.33 %, 1.36 %, 1.12 %, and 1.67 %, respectively, minimizing false alarms. However, FNR values for glioma (4.21 %), meningioma (4.95 %), and no-tumor (2.01 %) suggest occasional misclassification. The Pituitary class (FNR: 0.00 %) was perfectly classified for both datasets.

6. Discussion

In this section, the performance of the proposed model will be evaluated through multiple comparative analyses. Models’ accuracy is

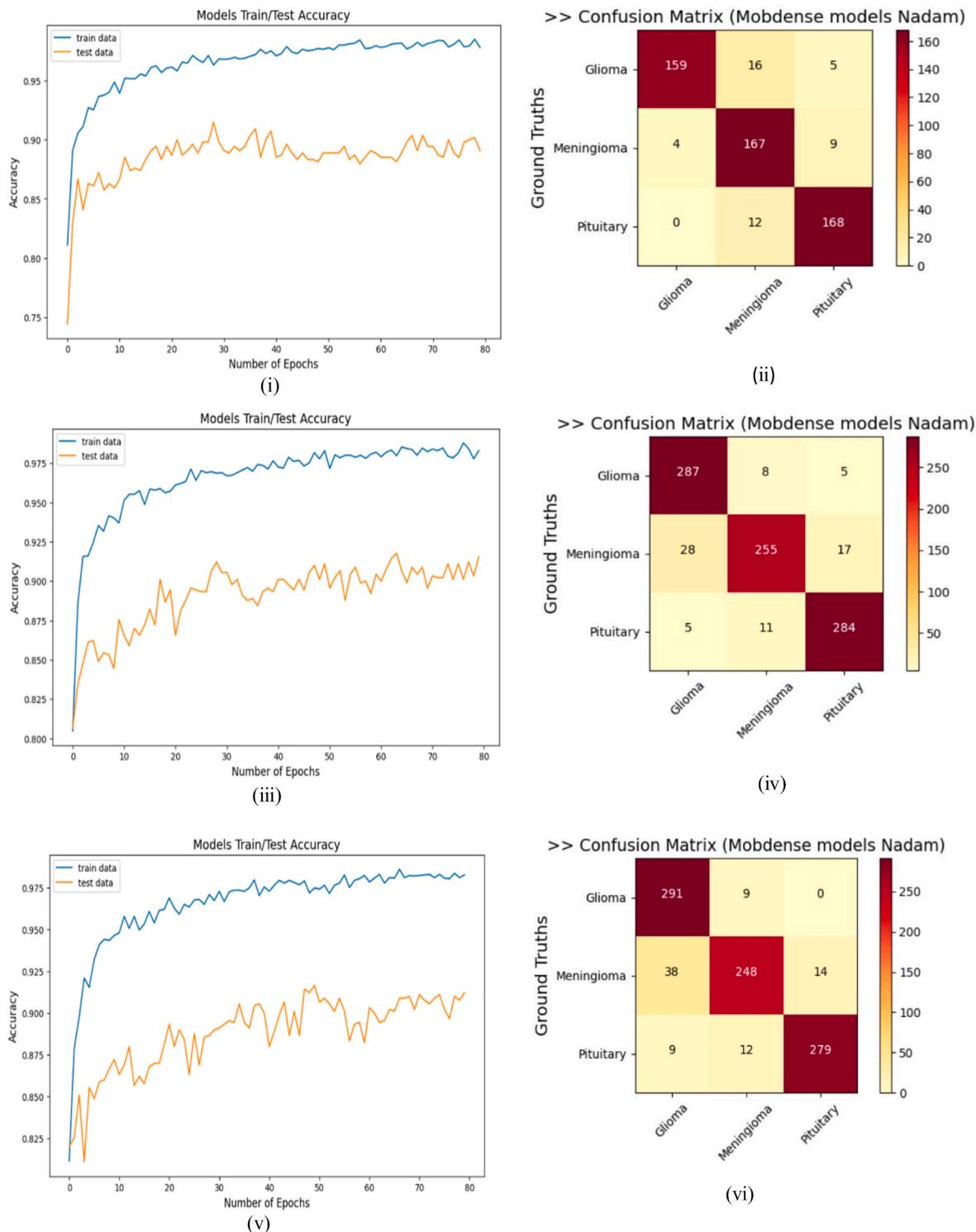


Fig. 5. Proposed MobDenseNet models training and testing accuracy curves using train images from Figshare dataset [43] and (i) 540 test images from Ref. [44], (ii) related confusion matrix (iii) 900 test images from Ref. [45], iv), related confusion matrix (v) 900 test images from Ref. [46] and vi), related confusion matrix.

compared with existing state-of-the-art methods on the same datasets; evaluations are applied across various datasets to reveal its versatility. Cross-dataset evaluations are also performed to assess the model's generalization capability. Finally, its performance on balanced datasets was validated.

6.1. Performance comparison of MobDenseNet model with existing work on same dataset

To establish a hybrid model, it is necessary to compare its performance with related state-of-the-art methods. In this section, MobDenseNet will be compared to a few other recent research works that are using similar datasets of this study. DL models are dataset-dependent, which means the performance of a DL model highly depends on the

Table 13

Summary of the model on a balanced dataset

Dataset	Class	No of images	Train 90 %	Test 10 %	Total	Accuracy		
						MobDenseNet	DenseNet121	MobileNetV1
Dataset-1 Figshare	glioma	708	1911	213	2124	97.3 %	92.02 %	91.04 %
	meningioma	708						
	pituitary	708						
Dataset-2 Sartaj	glioma	500	1800	200	2000	95.8 %	89.5 %	89.0 %
	meningioma	500						
	notumor	500						
	pituitary	500						

structure of the dataset. In [Table 9](#), we compare the accuracy result of the MobDenseNet model on Dataset-1 with the accuracy result of other hybrid methods [30,34,35,37,39] in which the same dataset was used. Similarly, [Table 10](#) shows the comparison of the proposed hybrid model for Dataset-2 with others [14,27].

6.2. Performance comparison of MobDenseNet model with existing work on various datasets

In this section, we review the classification accuracy attained by several DL models when applied to different datasets as described in the literature; following that, we present the performance results of the proposed MobDenseNet for the datasets. [Table 11](#) points out that the newly proposed hybrid model works better on binary and multiclass datasets [20,22,23,26,29,31,32,42].

6.3. MobDenseNet models accuracy on cross-dataset

To investigate the model's generalization capabilities, we conducted cross-dataset evaluations. We considered Dataset-1 (Figshare) to train the proposed model and tested it on the datasets from different domains [44–46]. For testing, specifically, we used 540 images from the Sartaj database, 900 from Ref. [45] and 900 from Ref. [46]. All the test datasets included images of four classes (glioma, meningioma, notumor and pituitary). Since the model was trained on three classes (glioma, meningioma, and pituitary), we excluded the 'no-tumor' class from the test data to evaluate the model's classification of unseen data. The model performed well on the source dataset, but the testing accuracy dropped notably on the external dataset. This drop can be attributed to domain shift, which refers to the difference in data distributions between the training and testing datasets. In our multi-dataset scenario, different factors such as differences in image acquisition, sample diversity, background artifacts, and noise, etc., can contribute to this domain shift. The Figshare dataset comprises MRI scans from Nanfang Hospital and Tianjin Medical University General Hospital, while the [44–46] datasets likely include images from other, undisclosed sources. Variations in MRI scanner types, imaging protocols, and contrast agent usage can result in significantly different image characteristics, including intensity levels, contrast, and anatomical detail. The proposed model exhibited superior performance across various cross-dataset scenarios, achieving test accuracies close to 92 %, outperforming both DenseNet121 and MobileNetV1 ([Table 12](#)). [Fig. 5](#) demonstrates train test accuracy curve for each experiment with their corresponding confusion matrix for better understanding of model's behavior on cross-dataset experiments. The testing accuracy curve gets lower than the training accuracy curve as we provided testing images from different domains, but the reported testing accuracy is still good enough to distinguish different brain tumor types. [Fig. 5\(ii\)](#) depicts the confusion matrix where the proposed model was tested on [44], which reveals correctly classified samples as glioma: 159, meningioma: 167, and pituitary: 168. On the other hand, for glioma images, 16 were misclassified as meningioma and 5 were incorrectly predicted as pituitary tumors. For meningioma cases, 4 were incorrectly classified as glioma and 9 were misclassified as pituitary. Among

pituitary tumor cases, 12 were misclassified as meningioma and none as glioma. [Fig. 5\(iv\)](#) displays 826 correctly classified cases and 74 misclassified cases as [45]. Here, 13 glioma images, 45 meningioma images, and 16 pituitary images were incorrectly predicted. Similarly, 818 correctly identified cases are reported in [Fig. 5\(vi\)](#) on [46], where 82 erroneous classifications were observed, of which 9 were from glioma, 52 from meningioma, and 21 from pituitary. However, the overall classification performance remains strong, with most samples correctly identified across all three classes. To further reduce these inter-class misclassifications, different domain adaptation techniques can be explored in the future.

6.4. MobDenseNet models accuracy on balanced dataset

We tested our model on a balanced dataset. We considered a balanced subset of 2124 images with 708 images per class from Dataset-1 (Figshare) and a balanced subset of 2000 images with 500 images per class from Dataset-2 (Sartaj). [Table 13](#) depicts that the MobDenseNet model outperformed both DenseNet121 and MobileNetV1 in terms of accuracy in both cases.

7. Limitations

While dual-mode architectures naturally increase the number of trainable parameters, which results in higher memory usage and computational requirements, the enhanced feature extraction capabilities of the model balance this trade-off. The proposed model's inference speed is slightly higher than that of architectures such as DenseNet121, MobileNetV1, InceptionV3, and Xception, but given the improved accuracy, this is a reasonable compromise. In the case of cross-dataset evaluation, domain variations between datasets may still require fine-tuning or adaptation techniques for optimal generalization. Like most DL models, the effectiveness of the proposed model depends on hyperparameter tuning, as improper tuning can make training unstable.

8. Conclusion

This study represents MobDenseNet, a new hybrid DL model named MobDenseNet to predict the types of brain tumors combining feature fusion, transfer learning, classification layers, and batch normalization layers. The model was trained with a batch size of 64, a learning rate of 0.002, 80 epochs, and ReLU activation using Nadam optimizer. The proposed MobDenseNet achieved superior results on two benchmark datasets reporting 98.4 % accuracy, 99.9 % AUC, 98.6 % precision, 98.40 % recall, 98.5 % F1-score for Dataset-1 and 96.02 % accuracy, 99.4 % AUC, 96.3 % precision, 95.7 % recall and 95.9 % F1-score for Dataset-2. Experimental results indicate that the proposed model outperforms some of the existing models reported in the literature in identifying gliomas, meningiomas, pituitary and healthy brains. The paper also presents the performance of the proposed model applied to various datasets considered in the literature, ranging from binary to multiclass. Finally, the model's performance is evaluated using multiple cross-dataset experiments. To assess how the model behaves with

unbiased inputs and equal class representation, we also evaluate its performance on the balanced dataset. As future work, the proposed model can be trained with diverse and large datasets, which may further improve the data-driven diagnosis in a practical application. In addition, explainable AI methods can be applied to enhance the transparency of the decision-making process of the proposed model. Finally, the model can be further improved to be lightweight, which is for deployment in a practical resource-constrained scenario.

CRediT authorship contribution statement

Meher Afroj: Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **M. Rubaiyat Hossain Mondal:** Writing – review & editing, Supervision, Project administration, Conceptualization. **Md Riad Hasan:** Writing – review & editing, Visualization, Validation, Formal analysis. **Sworna Akter:** Writing – review & editing, Validation, Methodology, Conceptualization.

Funding

The work received no funding.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research work is part of the MSc (ICT) thesis of the author, Meher Afroj, under the supervision of the author, M. Rubaiyat Hossain Mondal, at IICT, BUET, Bangladesh.

Data availability

Data will be made available on request.

References

- [1] World Health Organization. Brain tumors. <https://www.who.int/news-room/fact-sheets/detail/cancer>; 2023.
- [2] Srinivasan S, Francis D, Mathivanan SK, Rajadurai H, Shivahare BD, Shah MA. A hybrid deep CNN model for brain tumor image multi-classification. *BMC Med Imag* 2024;24(1):21.
- [3] Cinar N, Kaya M, Kaya B. A novel convolutional neural network-based approach for brain tumor classification using magnetic resonance images. *Int J Imag Syst Technol* 2023;33(3):895–908.
- [4] Rasheed Z, Ma YK, Ullah I, Al Shloul T, Tufail AB, Ghadi YY, Khan MZ, Mohamed HG. Automated classification of brain tumors from magnetic resonance imaging using deep learning. *Brain Sci* 2023;13(4):602.
- [5] Rizwan M, Shabbir A, Javed AR, Shabbir M, Baker T, Obe DAJ. Brain tumor and glioma grade classification using Gaussian convolutional neural network. *IEEE Access* 2022;10:29731–40.
- [6] Kaur R, Doege A, Upadhyaya GK. An ensemble learning approach for brain tumor classification using MRI. In: Soft computing: theories and applications: proceedings of SoCTA 2020, vol. 1. Springer Singapore; 2022. p. 645–56.
- [7] Priyadarshini P, Kanungo P, Kar T. Multigrade brain tumor classification in MRI images using fine tuned EfficientNet. e-Prime-Adv Electr Eng, Electr Energy 2024; 8:100498.
- [8] Mijwil MM, Doshi R, Hiran KK, Unogwu OJ, Bala I. MobileNetV1-based deep learning model for accurate brain tumor classification. *Mesopotamian J Comput Sci* 2023;29–38.
- [9] Alsaif H, Guesmi R, Alshammari BM, Hamrouni T, Guesmi T, Alzamil A, Belguesmi L. A novel data augmentation-based brain tumor detection using convolutional neural network. *Appl Sci* 2022;12(8):3773.
- [10] Garcea F, Serra A, Lambert F, Morra L. Data augmentation for medical imaging: a systematic literature review. *Comput Biol Med* 2023;152:106391.
- [11] Li M, Kuang L, Xu Z, Sha Z. Brain tumor detection based on multimodal information fusion and convolutional neural network. *IEEE Access* 2019;7: 180134–46.
- [12] Das S, Bose S, Nayak GK, Saxena S. Deep learning-based ensemble model for brain tumor segmentation using multi-parametric MR scans. *Open Comput Sci* 2022;12(1):211–26.
- [13] Showrov AA, Aziz MT, Nabil HR, Jim JR, Kabir MM, Mridha MF, Asai N, Shin J. Generative adversarial networks (GANs) in medical imaging: advancements, applications and challenges. *IEEE Access* 2024;12:35728–53.
- [14] Saeedi S, Rezayi S, Keshavarz H, Niakan Kalhori RS. MRI-based brain tumor detection using convolutional deep learning methods and chosen machine learning techniques. *BMC Med Inf Decis Making* 2023;23(1):16.
- [15] Simo AMD, Kouanou AT, Monthe V, Nana MK, Lonla BM. Introducing a deep learning method for brain tumor classification using MRI data towards better performance. *Inform Med Unlocked* 2024;44:101423.
- [16] Diaz-Pernas FJ, Martinez-Zarzuela M, Anton-Rodriguez M, Gonzalez-Ortega D. A deep learning approach for brain tumor classification and segmentation using a multiscale convolutional neural network. *Healthcare* 2021;9(2):153. MDPI.
- [17] Khan MSI, Rahman A, Debnath T, Karim MR, Nasir MK, Band SS, Mosavi A, Dehzangi I. Accurate brain tumor detection using deep convolutional neural network. *Comput Struct Biotechnol J* 2022;20:4733–45.
- [18] Kurdi SZ, Ali MH, Jaber MM, Saba T, Rehman A, Damasevicius R. Brain tumor classification using meta-heuristic optimized convolutional neural networks. *J Personalized Med* 2023;13(2):181.
- [19] Xu M, Guo L, Wu H-C. Novel robust automatic brain-tumor detection and segmentation using magnetic resonance imaging. *IEEE Sens J* 2024;24(7): 10957–64. 1 April1.
- [20] Neamah K, Mohamed F, Waheed SR, Kurdi WHM, Taha AY, Kadhim KA. Utilizing deep improved ResNet50 for brain tumor classification based MRI. *IEEE Open J Computer Soc* 2024;5:446–56.
- [21] Gomez-Guzman MA, Jimenez-Beristain L, Garcia-Guerrero EE, Lopez-Bonilla OR, Tamayo-Perez UJ, Esqueda-Elizondo JJ, Palomino-Vizcaino K, Inzunza-Gonzalez E. Classifying brain tumors on magnetic resonance imaging by using convolutional neural networks. *Electronics* 2023;12(4):955.
- [22] Ahmed S, Podder P, Mondal MRH, Rahman SA, Kannan S, Hasan MJ, Rohan A, Prosvirin AE. Enhancing brain tumor classification with transfer learning across multiple classes: an in-depth analysis. *BioMed Informatics* 2023;3(4):1124–44.
- [23] Mahmoud A, Awad NA, Alsubaie N, Ansarullah SI, Alqahtani MS, Abbas M, Usman M, Soufiene B, Saber A. Advanced deep learning approaches for accurate brain tumor classification in medical imaging. *Symmetry* 2023;15(3):571.
- [24] Talukder MA, Islam MM, Uddin MA, Akhter A, Pramanik MAJ, Aryal S, Almoyad MAA, Hasan KF, Moni MA. An efficient deep learning model to categorize brain tumor using reconstruction and fine-tuning. *Expert Syst Appl* 2023;230: 120534.
- [25] Yapiç M, Karakıç R, Gürkahraman K. Improving brain tumor classification with deep learning using synthetic data. *Comput Mater Continua (CMC)* 2023;74(3).
- [26] Srinivas C, Ks NP, Zakariah M, Althaibi YA, Shaukat K, Partabane B, Awal H. Deep transfer learning approaches in performance analysis of brain tumor classification using MRI images. *J Healthc Eng* 2022;2022(1):3264367.
- [27] Asif S, Zhao M, Tang F, Zhu Y. An enhanced deep learning method for multi-class brain tumor classification using deep transfer learning. *Multimed Tool Appl* 2023; 82(20):31709–36.
- [28] Suryawanshi S, Patil SB. Efficient brain tumor classification with a hybrid CNN-SVM approach in MRI. *J Adv Inf Technol* 2024;15(3).
- [29] Bourennane M, Naimi H, Mohamed E. Deep feature extraction with cubic-SVM for classification of brain tumor. *Stud Eng Exact Sci* 2024;5(1):19–35.
- [30] Haq EU, Jianjun H, Huarong X, Li K, Weng L. A hybrid approach based on deep CNN and machine learning classifiers for the tumor segmentation and classification in brain MRI. *Comput Math Methods Med* 2022;2022(1):6446680.
- [31] Senan EM, Jadhav ME, Rassem TH, Aljaloud AS, Mohammed BA, Al-Mekhlafi ZG. Early diagnosis of brain tumour mri images using hybrid techniques between deep and machine learning. *Comput Math Methods Med* 2022;2022(1):8330833.
- [32] Kang J, Ullah Z, Gwak J. MRI-based brain tumor classification using ensemble of deep features and machine learning classifiers. *Sensors* 2021;21(6):2222.
- [33] Ullah Z, Farooq MU, Lee SH, An D. A hybrid image enhancement based brain MRI images classification technique. *Med Hypotheses* 2020;143:109922.
- [34] Alqhtani SM, Soomro TA, Ali A, Aziz A, Irfan M, Rahman S, Jalalah M, Almwagani AH, Eljak LAB. Improved brain tumor segmentation and classification in brain MRI with FCM-SVM: a diagnostic approach. *IEEE Access* 2024;12: 61312–35.
- [35] Rahman T, Islam MS, Uddin J. MRI-based brain tumor classification using a dilated parallel deep convolutional neural network. *Digit* 2024;4(3):529–54.
- [36] Amin M, Nahar KM, Gharaibeh H, Nasareyh A, Alsalmanc NA, Alomar A, Malkawi M, Alqasem N, Smerat A, Zitar RA, Nusier S. DieT transformer model with PCA-ADE integration for advanced multi-class brain tumor classification. *Intell Based Med* 2025;11:100192.
- [37] Patil S, Kirange D. Ensemble of deep learning models for brain tumor detection. *Procedia Comput Sci* 2023;218:2468–79.
- [38] Tandel GS, Tiwari A, Kakde OG, Gupta N, Saba L, Suri JS. Role of ensemble deep learning for brain tumor classification in multiple magnetic resonance imaging sequence data. *Diagnostics* 2023;13(3):481.
- [39] Sandhiya B, Raj SKS. Deep learning and optimized learning machine for brain tumor classification. *Biomed Signal Process Control* 2024;89:105778.
- [40] Hossain S, Chakrabarty A, Gadekallu TR, Alazab M, Piran MJ. Vision transformers, ensemble model, and transfer learning leveraging explainable AI for brain tumor detection and classification. *IEEE J Biomed Health Inf March* 2024;28(3):1261–72.
- [41] Aamir M, Rahman Z, Dayo ZA, Abro WA, Uddin MI, Khan I, Imran AS, Ali Z, Ishfaq M, Guan Y, Hu Z. A deep learning approach for brain tumor classification using MRI images. *Comput Electr Eng* 2022;101:108105.

- [42] Masood M, Nazir T, Nawaz M, Mahmood A, Rashid J, Kwon HY, Mahmood T, Hussain A. A novel deep learning method for recognition and classification of brain tumors from MRI images. *Diagnostics* 2021;11(5):744.
- [43] Figshare Brain Tumor Dataset. Available online: <https://www.kaggle.com/datasets/ashkhagan/figshare-brain-tumor-dataset> (accessed on 12 March 2024).
- [44] SartajBhuaji Brain Tumor classification MRI. Available online: <https://www.kaggle.com/sartajbhuvaji/brain-tumor-classification-mri> (accessed on 24 March 2024).
- [45] Kaggle Brain Tumor Dataset. Available online: <https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset> (accessed on 2 October 2024).
- [46] Kaggle Brain Tumor Dataset. Available online: <https://www.kaggle.com/datasets/mohammadhosseini77/brain-tumors-dataset> (accessed on 3 October 2024).
- [47] Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2017. p. 4700–8.
- [48] Howard AG. Mobilenets: efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*. 2017.
- [49] Chakrabarty N. Brain MRI images for brain tumor detection dataset. Available online: <https://www.kaggle.com/datasets/navoneel/brain-mri-images-for-brain-tumor-detection>. [Accessed 12 June 2024].
- [50] Br35H. Brain tumor detection. Available online: <https://www.kaggle.com/datasets/ahmedhamada0/brain-tumor-detection>. [Accessed 2 June 2024].