Multi Model Comparison Deep Learning Image Classification on Multiclass Brain Tumor Disease

Md. Alahi Almin Tansen
Department of CSE
European University of Bangladesh
Dhaka, Bangladesh
aatansen@gmail.com

Tara Moni
Department of CSE
European University of Bangladesh
Dhaka, Bangladesh
taramoni607@gmail.com

Habibur Rahman
Department of CSE
European University of Bangladesh
Dhaka, Bangladesh
habiburrahman102001@gmail.com

Abstract:

A brain tumor is a mass of abnormal cells that grow in the brain. There are different types of brain tumors, including glioma tumors, which develop in glial cells, and meningioma tumors, which develop from the meninges. Pituitary tumors grow in the pituitary gland, and can cause hormonal imbalances. MRI (Magnetic Resonance Imaging) is a type of diagnostic imaging that uses a magnetic field and radio waves to produce detailed images of the brain's structures, allowing doctors to visualize any abnormal growths or masses in the brain. In this study, we evaluated the performance of seven different convolutional neural network (CNN) models for detecting brain tumors in MRI images. We used two datasets, one from Kaggle with 3264 images and another from Roboflow with 6786 images. Seven pre-trained models are used including Xception, VGG16, VGG19, ResNet50V2, InceptionV3, MobileNetV2, DenseNet201. Our results show that on Kaggle dataset ResNet50V2 get the highest accuracy of 76.65%, while the DenseNet201 model achieved the highest accuracy of 89.23% on the Roboflow dataset. The outcome of this study can lead to earlier diagnosis, more accurate treatment planning, and doctors can provide faster and more effective treatments.

Keywords: Tumor detection; MRI images; Xception; VGG16; VGG19; ResNet50V2; InceptionV3; MobileNetV2; DenseNet201; Deep learning; Transfer learning; CNN; Multi Model;

1. Introduction:

Brain tumors are a complex and challenging medical condition that requires accurate diagnosis and treatment. MRI imaging is commonly used to detect brain tumors, but manual interpretation of these images can be time-consuming and error-prone. The use of deep learning CNN model, has shown promise in improving the accuracy and efficiency of brain tumor detection. Brain tumors are abnormal growths of cells that occur within the brain or spinal cord. They can be classified into different types based on their location, origin, and characteristics. Glioma tumor, meningioma tumor, and pituitary tumor are three common types of brain tumors. Glioma tumors originate from glial cells which provide support and protection for neurons. Meningioma tumors develop in the meninges, the protective layers surrounding the brain and spinal cord. Pituitary tumors arise from the pituitary gland which is responsible for regulating hormone production.

2. Literature Review:

Previous studies have demonstrated the effectiveness of CNN models for brain tumor detection. However, there is still a need for further evaluation of these models using larger and more diverse datasets. Many research has done before on image classification of MRI brain disease.

Recent study by Bitto et al[1] five pre-trained models, including VGG-16, VGG-19, ResNet-50, Xception, and Inception-V3. The model with the highest accuracy was ResNet-50, which performed at 96.76%. The model with the highest precision overall was Inception V3, with a precision score of 98.83%.

Another research where Alturki et al[3] used a voting classifier that combines logistic regression with stochastic gradient descent and deep convolutional features to classify tumorous victims from normal patients. The proposed approach achieved a high accuracy of 99.9%, which is better than cutting-edge methods.

Seetha and Selvakumar[4] proposed a CNN architecture is designed with small kernels and low complexity, achieving a high accuracy rate of 97.5%, outperforming other state-of-the-art methods. This proposed method shows promise for accurate and efficient brain tumor detection using MRI scans.

Cheng et al[5] uses three feature extraction methods, and the proposed method outperformed traditional methods with accuracies ranging from 82.31% to 91.28%. These results demonstrate that the proposed method is feasible and effective for classifying brain tumors using T1-weighted CE-MRI images.

Shafi et al[10] proposed method includes pre-processing, feature extraction, feature selection, and classification using support vector machine (SVM) classifier and prediction model with majority voting. The proposed system outperforms other state-of-the-art methods with high sensitivity, specificity, precision, and accuracy rates of 97.5%, 98.838%, 98.011%, and 98.719%, respectively. The overall training and testing accuracy of the proposed model are 97.957% and 97.744%, respectively. These findings demonstrate the potential of the proposed ensemble learning method for detecting lesions coexisting with tumors in neuro-medicine diagnosis.

Özkaraca et al[11] They tested a simple CNN architecture, as well as VGG16Net and DenseNet models, with and without transfer learning. Using the proposed model with dense layers, they achieved a success rate of 94-97%. However, the model had a long processing time due to using both dense and convolutional networks.

Many more author write their paper on brain tumor classification. In our study we are trying to find out how the different types of CNN models interact with different range of dataset and find out which model fit best on those dataset.We use two range of dataset one is from Kaggle and another is from Roboflow.

3. Proposed Methodology:

We used two datasets consisting of MRI images. The Kaggle dataset contained 3264 images and the Roboflow dataset contained 6786 images. We trained and tested seven pre-trained CNN models, including VGG16, VGG19, MobileNetV2, ResNet50V2, InceptionV3, Xception, and DenseNet201. We used the Keras library with TensorFlow backend to implement the models and trained them using a batch size of 32 and 50 epochs.

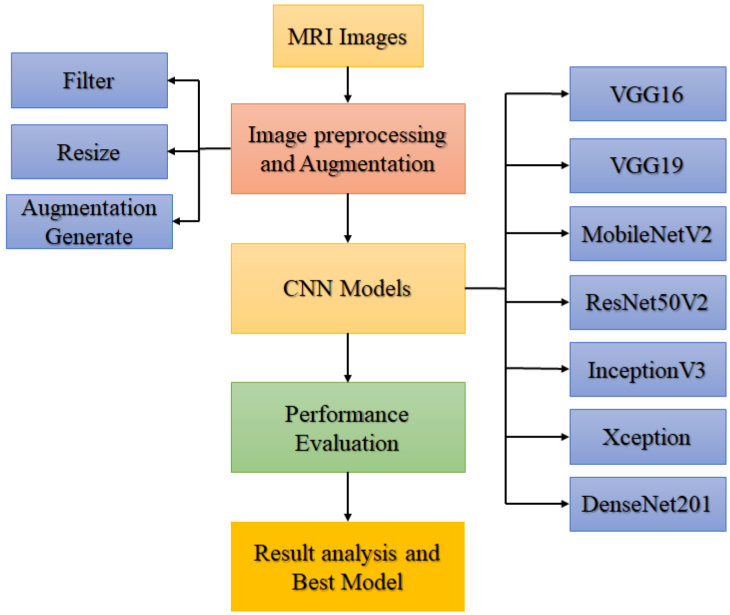

Figure 1 System Procedure Diagram

3.1 Dataset Description

We use two different range of data set that include Kaggle and Roboflow dataset. In Kaggle dataset total 3264 brain MRI images are present and in Roboflow dataset 6786 brain MRI images are present. Brain tumors are a serious disease that affects both children and adults. They make up the majority of primary Central Nervous System tumors. Around 11,700 people are diagnosed with brain tumors each year, with a 5-year survival rate of approximately 34% for men and 36% for women. Types of brain tumors include benign, malignant, and pituitary tumors. Accurate diagnostics, proper treatment planning, and early detection are crucial to improve patient outcomes. Magnetic Resonance Imaging (MRI) is the best technique to detect brain tumors, but it generates a large amount of image data that must be examined by a radiologist. Manual examination of these images can be challenging and prone to errors due to the complexity of brain tumors and their properties. Both dataset has two separate folder that contain training and testing which include four subfolder Glioma tumor, Meningioma tumor, No tumor and Pituitary tumor. Fig 2 represent those images of dataset.

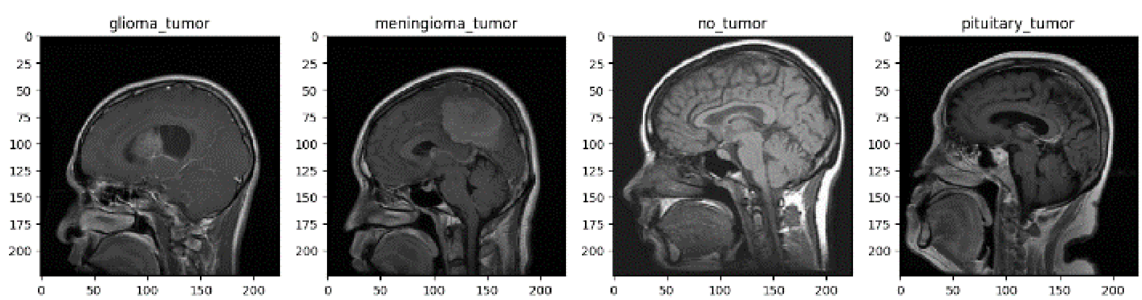

Figure 2 Sample images of dataset

3.2 Preprocessing and Augmentation

We preprocessed MRI images by filtering and preprocessed function is applied to each image to standardize the pixel values to improve training efficiency. Augmentation is done by rotating, zooming, flipping, shifting which is referred to table 1. All images are resized into 224x224 pixel. After preprocessing and augmentation all our proposed models are used to train. In preprocessing we found that in Roboflow dataset there are black images that does not contain brain MRI which are measured by using min-max pixel size method and filter out those images and use the correct image that contain brain MRI. After filterout total number of images in Roboflow and Kaggle dataset in table 2.

Rotation range	Width shift range	Height shift range	Zoom range	Horizontal flip
10	0.1	0.1	0.1	True

Table 1 : Data Augmentation Parameters

Filtering	Kaggle Dataset	Roboflow Dataset
Before	3264	6786
After	3264	6205

Table 2 : Filtering images

We can see that in table 1 ; Kaggle dataset does not change but in roboflow that contain some black images has been decreased from 6786 to 6205.

We used training data by splitting 80% for training and 20% for validation. For both dataset that contain the number of training, validation and testing in table 3.

Dataset Measurement	Kaggle Dataset	Roboflow Dataset
Training Data	2297	3941
Validation Data	573	983
Testing Data	394	1281

Table 3 : number of dataset

3.3 Model Implementation

In our proposed deep learning based CNN model we used seven model including VGG16, VGG19, MobileNetV2, ResNet50V2, InceptionV3, Xception, and DenseNet201.

VGG16 model The architectural design of VGG16 is based on stacking multiple 3x3 convolutional layers, which is more computationally efficient than using larger filters. The first two layers have 64 filters each, followed by four layers with 128 filters each, and three layers with 256 filters each. The final two convolutional layers have 512 filters each. All convolutional layers are followed by rectified linear unit (ReLU) activation functions and 2x2 max pooling layers. It has over 138 million trainable parameters [ref]

VGG19 is a deep convolutional neural network model that contains 19 layers. includes 16 convolutional layers, followed by three fully connected layers. The first two sets of convolutional layers in VGG19 are similar to VGG16, consisting of two or three convolutional layers followed by max pooling. The third set consists of four convolutional layers, followed by a max pooling layer. Finally, there are five convolutional layers, followed by a max pooling layer, in the fourth set.[ref]

MobileNetV2 model includes a series of inverted residual blocks with linear bottleneck layers. In each block, the input feature map is first passed through a 1x1 convolutional layer to reduce the number of channels, and then a depthwise separable convolutional layer is applied to capture spatial information. The output of the depthwise convolution is then passed through another 1x1 convolutional layer to increase the number of channels. Finally, a skip connection is added between the input and output of the block to facilitate gradient flow.[ref]

ResNet50V2 is a variant of the ResNet (Residual Network) architecture that contains 50 layers and is a deep convolutional neural network designed to improve accuracy on image classification tasks while reducing the impact of vanishing gradients during training.[ref]

InceptionV3 model consists of a series of "Inception" modules, which include multiple parallel convolutional layers with different filter sizes. These parallel convolutional layers are combined through concatenation to capture features at different scales.[ref]

Xception model uses a similar approach to the Inception model, which consists of a series of "Inception" modules. However, in the Xception model, each Inception module is replaced by a collection of depthwise separable convolutions. This means that instead of using a standard convolutional layer followed by a pointwise convolutional layer, the Xception model first applies a depthwise convolution to each input channel separately, followed by a pointwise convolution to combine the output channels.[ref]

DenseNet201 is a convolutional neural network architecture that was introduced by the authors of the DenseNet family of models. It has 201 layers in total and is specifically designed for image classification tasks. It consists of multiple dense blocks, where each dense block includes multiple layers that are densely connected to each other. The dense connections allow information to flow more easily through the network and facilitate feature reuse, which helps to reduce the number of parameters needed.[ref]

All those models are used in same way so that result could be compared and analyzed perfectly. We freeze the pre-trained layers in the base model and train only the new layers. ImageNet weights are used. Models are compiled with categorical cross-entropy loss and a lower learning rate of 0.0001 with Adam optimizer. In output layer softmax activation is used. Total number of epoch used is 50.

3.4 Performance Evaluation

For performance evaluation after training each model we test it through test dataset. We calculate accuracy, precision, recall and also measure the confusion matrix that calculated the number of true positive, true negative, false positive, and false negative predictions on test dataset. Furthermore ROC, AUC curve is measured to plots the true positive rate (TPR) against the false positive rate (FPR). Equation (1) to (4) is used to generate performance matrices.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall\ or\ TPR = \frac{TP}{TP + FN} \tag{3}$$

$$FPR = \frac{FP}{FP + TN} \tag{4}$$

where TP, TN, FP, FN, FPR and TPR are respectively the number of true-positives, true-negatives, false-positives, false-negatives, false-positive-rate and true-positive-rate respectively.

4. Result Analysis:

On the Kaggle dataset VGG16 highest training accuracy 98.35%, highest validation accuracy 81.68% and test accuracy 73.86%. VGG19 highest training accuracy 98.17%, highest validation accuracy 81.68% and test accuracy 74.37%. MobileNetV2 Highest training accuracy 99.00%, highest validation accuracy 83.60% and test accuracy 75.89%. ResNet50V2 highest training accuracy 97.91%, highest validation accuracy 84.12% and test accuracy 76.64%. InceptionV3 highest training accuracy 97.65% highest validation accuracy 82.20 and test accuracy 74.11%, Xception highest training accuracy 98.26%, highest validation accuracy 84.12% and test accuracy 72.84%. DenseNet201 highest training accuracy 99.13%, highest validation accuracy 87.61% and test accuracy 72.33%. On the Roboflow dataset VGG16 highest training accuracy 97.61%, highest validation accuracy 83.01% and test accuracy 88.05%. VGG19 highest training accuracy 97.26%, highest validation accuracy 81.69% and test accuracy 87.90%. MobileNetV2 Highest training accuracy 98.45%, highest validation accuracy 83.93% and test accuracy 88.60%. ResNet50V2 highest training accuracy 97.18%, highest validation accuracy 83.93% and test accuracy 88.83%. InceptionV3 highest training accuracy 96.73% highest validation accuracy 83.01 and test accuracy 87.43%, Xception highest training accuracy 97.49%, highest validation accuracy 84.84% and test accuracy 88.91%. DenseNet201 highest training accuracy 98.40%, highest validation accuracy 86.98% and test accuracy 89.22%.

Now in table 4 shows the matrices measured in Kaggle dataset

Model	Disease	precision	recall	f1-score	support	Accuracy
	Glioma_tumor	100.00%	25.00%	40.00%	100	
	Meningioma tumor	57.28%	99.13%	72.61%	115	
VGG16	No tumor	86.32%	96.19%	90.99%	105	73.86%
	Pituitary tumor	96.23%	68.92%	80.32%	74	
	Glioma_tumor	91.00%	20.00%	33.00%	100	
	Meningioma tumor	65.00%	99.00%	79.00%	115	
VGG19	No tumor	77.00%	98.00%	86.00%	105	74.37%
	Pituitary tumor	89.00%	76.00%	82.00%	74	
	Glioma_tumor	97.00%	29.00%	45.00%	100	
	Meningioma tumor	61.00%	98.00%	75.00%	115	
MobileNetV2	No tumor	85.00%	95.00%	90.00%	105	75.89%
	Pituitary tumor	93.00%	77.00%	84.00%	74	
	Glioma_tumor	97.22%	35.00%	51.47%	100	
	Meningioma tumor	64.77%	99.13%	78.35%	115	
ResNet50V2	No tumor	80.46%	98.10%	88.41%	105	76.65%
	Pituitary tumor	92.59%	67.57%	78.12%	74	
	Glioma_tumor	100.00%	30.00%	46.15%	100	
	Meningioma tumor	66.86%	100.00%	80.14%	115	
InceptionV3	No tumor	71.03%	98.10%	82.40%	105	74.11%
	Pituitary tumor	93.62%	59.46%	72.73%	74	
	Glioma_tumor	96.00%	22.00%	36.00%	100	
	Meningioma tumor	66.00%	97.00%	79.00%	115	
Xception	No tumor	70.00%	99.00%	82.00%	105	72.84%
	Pituitary tumor	89.00%	68.00%	77.00%	74	
	Glioma_tumor	96.00%	24.00%	38.00%	100	
	Meningioma tumor	64.00%	98.00%	77.00%	115	
DenseNet201	No tumor	71.00%	98.00%	82.00%	105	72.34%
	Pituitary tumor	98.00%	61.00%	75.00%	74	

Table 4 : Matrices of all models on Kaggle Dataset

Table 5 Shows the matrices measured in Roboflow dataset.

Model	Disease	precision	recall	f1-score	support	Accuracy
	Glioma_tumor	86.00%	82.00%	84.00%	299	
	Meningioma tumor	77.00%	77.00%	77.00%	301	
VGG16	No tumor	99.00%	93.00%	96.00%	381	88.06%
	Pituitary tumor	88.00%	99.00%	93.00%	300	
	Glioma_tumor	91.00%	84.00%	87.00%	299	
	Meningioma tumor	81.00%	76.00%	78.00%	301	
VGG19	No tumor	98.00%	92.00%	95.00%	381	87.90%
	Pituitary tumor	81.00%	99.00%	89.00%	300	
	Glioma_tumor	85.00%	94.00%	89.00%	299	
	Meningioma tumor	92.00%	6100.00%	73.00%	301	
MobileNetV2	No tumor	94.00%	98.00%	96.00%	381	88.60%
	Pituitary tumor	84.00%	100.00%	91.00%	300	
	Glioma_tumor	87.00%	89.00%	88.00%	299	
	Meningioma tumor	81.00%	77.00%	79.00%	301	
ResNet50V2	No tumor	98.00%	90.00%	94.00%	381	88.84%
	Pituitary tumor	88.00%	99.00%	93.00%	300	
	Glioma_tumor	81.00%	90.00%	85.00%	299	
	Meningioma tumor	87.00%	66.00%	75.00%	301	
InceptionV3	No tumor	98.00%	94.00%	96.00%	381	87.43%
	Pituitary tumor	83.00%	97.00%	90.00%	300	
	Glioma_tumor	95.00%	82.00%	88.00%	299	
	Meningioma tumor	81.00%	76.00%	78.00%	301	
Xception	No tumor	93.00%	97.00%	95.00%	381	88.91%
	Pituitary tumor	87.00%	99.00%	92.00%	300	
	Glioma_tumor	88.00%	93.00%	90.00%	299	
	Meningioma tumor	92.00%	67.00%	77.00%	301	
DenseNet201	No tumor	96.00%	96.00%	96.00%	381	89.23%
	Pituitary tumor	81.00%	99.00%	89.00%	300	

Table 5 : Matrices of all models on Roboflow Dataset

From Table 4 and 5 we can see the Accuracy on test dataset are lower in Kaggle dataset but higher accuracy in Roboflow dataset.

Highest accuracy achieved by DenseNet201 which is 89.23%. More measurement of DenseNet201 on Roboflow dataset showed below.

DenseNet201 model training and validation accuracy and loss plot respectively fig 3 fig 4

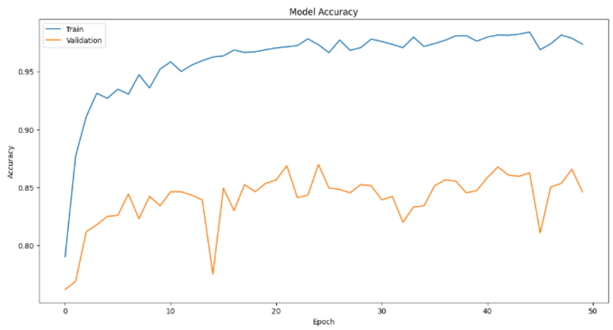

Figure 3 DenseNet201 training and validation accuracy

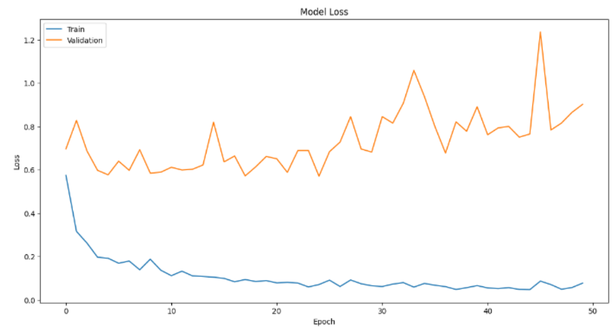

Figure 4 DenseNet201 training and validation loss

Confusion matrix and ROC,AUC curve of DenseNet201 Model respectively fig 5 and fig 6

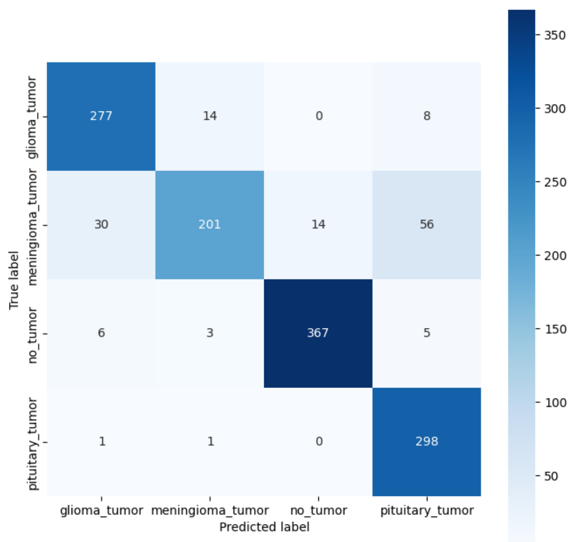

Figure 5 DenseNet201 model confusion matrix

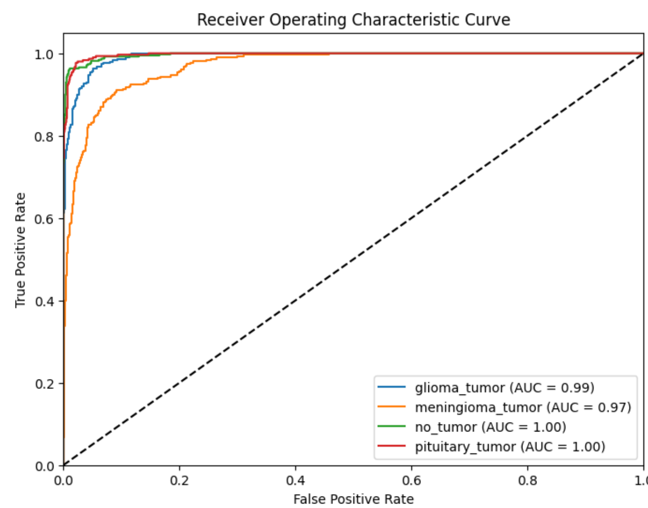

Figure 6 DenseNet201 model ROC AUC curve

5. Discussion:

We can see that our result improves on Roboflow dataset where number of total images are 6205 on the other hand model perform poorly on Kaggle where total number of images are 3264. We also observe that even when the training has higher percentage it does not mean that it will perform well on test dataset. So in our final best perform model which is DenseNet201 with the training accuracy of 98.40% and validation accuracy of 86.98% and finally on test dataset test accuracy of 89.23% fit well on Roboflow Dataset and give the best result.

6. Conclusion:

This study purpose is to show the deference of various model performance on two difference range of dataset of brain tumor MRI images. Here we used seven different deep learning models that includes VGG16, VGG19, ResNet50V2, InceptionV3, MobileNetV2, Xception, DenseNet201. Where we used Kaggle dataset of 3264 images and Roboflow dataset of 6786 images. We found that Kaggle dataset fit poorly on that model when we measured the test accuracy but on Roboflow dataset our model perform very well and the best result we found on DenseNet201 with the training accuracy of 98.40% and validation accuracy of 86.98% and finally on test dataset test accuracy of 89.23%. Model will perform more better if we have big amount of dataset of images. In future we can make this model more powerful on larger dataset and compare with more powerful deep learning models.

References:

1. Bitto, Abu Kowshir & Bijoy, Md. Hasan Imam & Yesmin, Sabina & Mahmud, Imran & Jueal Mia, Md & Biplob, Khalid. (2023). Tumor-Net: convolutional neural network modeling for classifying brain tumors from MRI images. International Journal of Advances in Intelligent Informatics. 9. 148. 10.26555/ijain.v9i2.872.
2. Shadin, Nazmus & Sanjana, Silvia & Chakraborty, Sovon & Sharmin, Nusrat. (2022). Performance Analysis of Glaucoma Detection Using Deep Learning Models. 190-195. 10.1109/ICISET54810.2022.9775828.
3. Alturki, Nazik & Umer, Muhammad & Ishaq, Abid & Abuzinadah, Nihal & Alnowaiser, Khaled & Mohamed, Abdullah & Saidani, Oumaima & Ashraf, Imran. (2023). Combining CNN Features with Voting Classifiers for Optimizing Performance of Brain Tumor Classification. Cancers. 15. 1767. 10.3390/cancers15061767.
4. Seetha, J. & Selvakumar, Raja. (2018). Brain Tumor Classification Using Convolutional Neural Networks. Biomedical and Pharmacology Journal. 11. 1457-1461. 10.13005/bpj/1511.
5. Cheng, Jun & Huang, Wei & Cao, Shuangliang & Yang, Ru & Yang, Wei & Yun, Zhaoqiang & Feng, Qianjin. (2015). Enhanced Performance of Brain Tumor Classification via Tumor Region Augmentation and Partition. PLoS ONE. 10. 10.1371/journal.pone.0140381.
6. Qassim, Hussam & Feinzimer, David. (2018). Compressed residual-VGG16 CNN model for big data places image recognition. 169-175. 10.1109/CCWC.2018.8301729.
7. Zülch, K.. (1965). Brain Tumors: Their Biology and Pathology. 10.1007/978-1-4899-6571-4.
8. Sonawane, Sumedh & Banne, Sujit & Bhor, Aditya & Barhate, Parag. (2022). Brain Tumor Classification using CNN. International Journal for Research in Applied Science and Engineering Technology. 10. 1526-1528. 10.22214/ijraset.2022.41560.
9. Kumar, Susheel & Bansal, Amishi & Singh, Nagendra. (2023). Brain Tumor Classification Using Deep Learning Techniques. 10.1007/978-3-031-24367-7_6.

10. Shafi, A. S. M. & Bayazid Rahman, Md & Anwar, Tanjilul & Halder, Rajkumar & Kays, H.M.. (2021). Classification of brain tumors and auto-immune disease using ensemble learning. Informatics in Medicine Unlocked. 24. 100608. 10.1016/j.imu.2021.100608.
11. Özkaraca, Osman & Bağrıaçık, Okan & Gürüler, Hüseyin & Khan, Faheem & Hussain, Jamil & Khan, Jawad & Laila, Umme. (2023). Multiple Brain Tumor Classification with Dense CNN Architecture Using Brain MRI Images. Life. 13. 349. 10.3390/life13020349.
12. Shamrat, F M & Chakraborty, Sovon & Afrin, Saima & Moharram, Md & Amina, Mahdia & Roy, Tonmoy. (2022). A Model Based on Convolutional Neural Network (CNN) for Vehicle Classification. 10.1007/978-981-16-9416-5_37.
13. Shamrat, F M & Chakraborty, Sovon & Moharram, Md & Roy, Tonmoy & Rahman, Masudur & Aronya, Biraj. (2022). A Transfer Learning Approach for Face Recognition Using Average Pooling and MobileNetV2. 10.1007/978-981-16-9416-5_38.
14. Akter, Shamima & Shamrat, F M & Chakraborty, Sovon & Karim, Asif & Azam, Sami. (2021). COVID-19 Detection Using Deep Learning Algorithm on Chest X-ray Images. Biology. 10. 10.3390/biology10111174.
15. Tasnim, Zarrin & Chakraborty, Sovon & Shamrat, F M & Chowdhury, Ali & Nuha, Humaira & Karim, Asif & Zahir, Sabrina & Billah, Md. (2021). Deep Learning Predictive Model for Colon Cancer Patient using CNN-based Classification. International Journal of Advanced Computer Science and Applications. 12. 687-696. 10.14569/IJACSA.2021.0120880.
16. Shamrat, F M & Chakraborty, Sovon & Billah, Md & Kabir, Moumita & Shadin, Nazmus & Sanjana, Silvia. (2021). Bangla numerical sign language recognition using convolutional neural networks. Indonesian Journal of Electrical Engineering and Computer Science. 23. 405-413. 10.11591/ijeecs.v23.i1.pp405-413.
17. Chakraborty, Sovon & Shamrat, F M & Billah, Md & Jubair, Md & Alauddin, Md & Ranjan, Rumesh. (2021). Implementation of Deep Learning Methods to Identify Rotten Fruits. 10.1109/ICOEI51242.2021.9453004.