FAMPO
FEDERATION OF AFRICAN MEDICAL PHYSICS ORGANIZATIONS

African Journal of Medical Physics
2022; 4(1): 63-68. doi: 10

Research Paper

Evaluating the Performance of Transfer-Learning Approaches for Multiclass Classification of Glioma, Meningioma and Pituitary Tumour

Adewole Maruf [1, 2, 3, ✉], Ige A. Taofeeq [4], Anazodo Udunna [2, 5], Akpochafor O. Michael [1, 3], Adewole Adetunji[6], Adeyomoye Adekunle [1, 7] and Aweda Moses [1, 2]

1. Department of Radiation Biology, Division of Radiotherapy and Radiodiagnosis, University of Lagos, Lagos, Nigeria.
2. Medical Artificial Intelligence Laboratory, Lagos, Nigeria.
3. Marcelle Ruth Cancer Centre, Victoria Island, Lagos, Nigeria.
4. University of Abuja, Abuja, Nigeria.
5. Montreal Neurological Institute, McGill University, Montreal, Canada.
6. Department of Computer Sciences, University of Lagos, Lagos, Nigeria.
7. Lagos University Teaching Hospital, Lagos, Nigeria

✉ Corresponding author: Adewole Maruf, Medical Artificial Intelligence Laboratory, Lagos, Nigeria, Email: marufmore@gmail.com

© AJMP is the official journal of the Federation of African Medical Physics Organizations (FAMPO). This is registered under Nigerian company number (CAC/IT/No 54182). See http://fampo-africa.org terms for full terms and conditions.
ISSN 2643-5977

Received: 2022.09.15; Accepted: 2023.03.21; Published: 2023.07.30

Abstract

Brain tumours are one of the deadliest types of cancers as only 36% of brain tumour patients survive five years after diagnosis. Brain tumours are detected and classified by biopsy -an invasive procedure with the potential to impede brain function and introduce infections. Diagnostic imaging approaches using anatomical magnetic resonance imaging (MRI) have minimized the use of pre-treatment biopsies for detection, but current visual classification of brain tumour images using expert readers have not improved classification accuracy enough to eliminate biopsies. Transfer learning is a machine learning technique where a previously trained model serves as the foundation for a model on a new problem. Recent advances in Convolution Neural Network (CNN), offer promise in using brain MRI to accurately classify brain tumours. Here, we evaluated the performance of 26 CNN models previously developed for general image classification in classification of brain tumours. We retrained 3064 T1-weighted contrast-enhanced MR images from 233 patients with either meningioma (708 images), glioma (1426 images), or pituitary tumour (930 images) using pre-trained weights from the ImageNet dataset and compared classification accuracies of the CNN models. This study provides an exhaustive evaluation of various state-of-the-art CNN models using a large publicly available multiclass brain MRI dataset. EfficientNetB3 had the highest accuracy of 98.98% in classifying tumor type among the 26 models tested. DenseNet121, EfficientNetB2, EfficientNetB5, and EfficientNetB4 also showed high accuracy in identifying the tumor type, with all models achieving an accuracy of more than 97%.

Keywords: Transfer Learning; Brain tumour classification; MRI; Convolution Neural Network (CNN); Keras Applications.

Introduction

Brain tumors are a serious condition that can directly impact human life and pose a threat to one's well-being. Brain tumors are a mass of abnormal tissue that grows uncontrollably and without following the natural processes that regulate normal cells (American

Association of Neurological Surgeons , 2021). More than 150 different brain tumours have been documented, but the two main groups of brain tumours are primary and metastatic. Most common types of adult brain tumours are Glioma, Meningioma and Pituitary tumour (Louis, et al., 2021).

In the United States alone, The American Cancer Society estimates that 73% of all diagnosis of brain tumor would lead to death (American Cancer Society , 2022). Also, the 5-year survival rate for people with a brain tumour is 36% (ASCO, 2022). This contrasts with people diagnosed with cancers of other part of the body such as breast (90%), prostate (98%) and cervix (66%) (American Cancer Society, 2022). This shows that Brain tumour contributes significantly to the global burden of diseases.

Brain tumours have proved challenging to diagnose at an early stage because most patients often notice subtle changes rather than symptoms (Walter, et al., 2019). The difficulty in diagnosing brain tumors at an early stage can be attributed to patients usually noticing subtle changes instead of apparent symptoms, which can limit treatment options. (Aldape, et al., 2019)

To gain a complete understanding of brain tumours, Magnetic Resonance Imaging (MRI) scan and histopathology is often performed, this produces information that aids proper diagnosis and staging. However, the process of obtaining biopsy tissue sample for histopathology is an invasive procedure and can affect normal brain function. Brain tumour classification is essential for accurate diagnosis, it helps to determine an effective approach towards the best line of treatment.

Convolutional Neural Network (CNN) is an aspect of deep learning that centers around imaging tasks. It has gathered lot of interest since 2012 when AlexNet won the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) (Krizhevsky, Sutskever, & Hinton, 2012). This has given rise to lots of models. Transfer learning is a machine learning technique where a previously trained model serves as the foundation for a model on a new problem without retraining. Keras Applications provides a collection of 26 pre-trained deep learning models that can be downloaded along with their weights in Tensorflow. The ImageNet dataset was used to train these models, and it can be utilized for prediction, feature extraction, and fine-tuning (Keras, 2021). Some research in brain tumour classification (Badža & Barjaktarovic, 2020), (Khan, Jue, Mushtaq, & Mushtag, 2020), (Sachin, Punn, Sonbhadra, & Agarwal, 2021) often features the creation of a new custom network architecture and comparing its performance with only few randomly chosen existing network. There has never been a comprehensive comparison of the performance of various models to determine which one is best for classifying brain tumor images.

This study aims to evaluate the effectiveness of Keras applications in classifying MRI images of brain tumors namely meningioma, glioma, and pituitary tumors. This aim is achieved by retraining and evaluating the CNN models to determine the best performers based on accuracy, recall, precision, and training time.

Material and Methods

The dataset was originally published by Cheng Jun on figshare under the creative commons Attribution 4.0 International (CC BY 4.0) (Cheng J. , 2017). The dataset includes 3064 T1weighted contrast-enhanced images from 233 meningioma (708 slices), glioma (1426 slices), and pituitary tumor patients (930 slices). This data comprises 1025 Sagittal slices, 994 Axial slices, and 1045 Coronal slices and is arranged in MATLAB data format. With the help of CV2, it was imported into arrays to contain the images and labels. The images were resized and saved as .jpg files.

Although considered large compared to other available MRI image databases, this database is still far smaller than databases generally used in the field of artificial intelligence (Badža & Barjaktarovic, 2020). The resolution of the images was also down-sampled from 512 x 512 to 224 x 224. Training set consists of 2355 images, validation images consist of 587 images while the Testing data consist of 122 images.

The dataset was prepared using Jupyter notebook divided into training and testing dataset. A HP core i5 CPU was used for data training, visualization, validation and analysis. A NVIDIA Tesla K80 GPU (Google Collaboratory) was used for model training and evaluation.

Keras (https://keras.io/) based on Tensorflow was used for the model training. OpenCV (Cv2) was used for the image handling, conversion and augmentation. Numpy and pandas were used for numerical computations. Matplotlib and seaborn was used for plotting and graphing.

The models were added as an output layer for each of the ImageNet trained Keras applications. The base model was supported with a GlobalAveragePooling2D with a dropout of 0.5 followed by a dense layer to give the classification output of 3. Accuracy, Precision and Recall was used as the evaluation metrics. The hyper parameters tunning values are as displayed in the table 1.

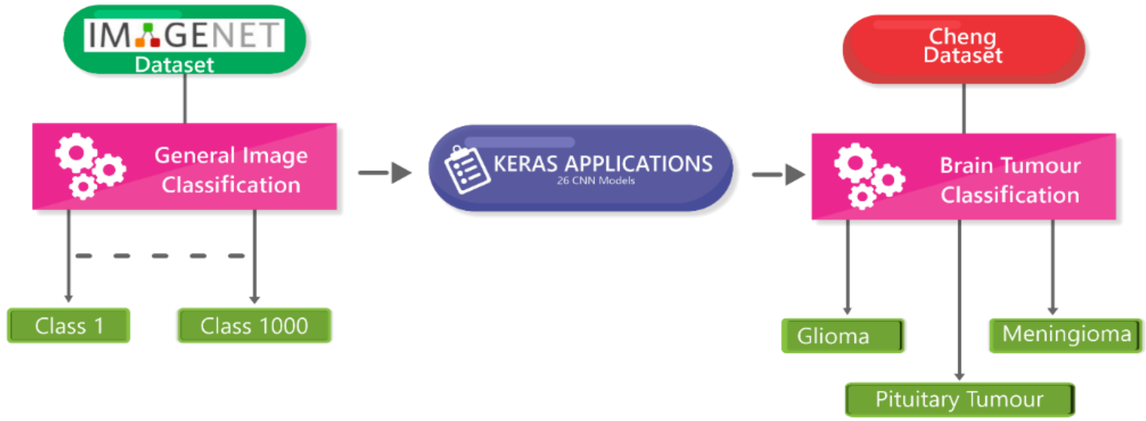

Figure 1. Research process flowchart.

Table 1 . Hyper-Parameters for Model Training

Parameter	Value
Input size	224 x 224 x 3
Weights	ImageNet
Epochs	10
Learning Rate	0.001
Batch Size	32
Optimizer	Adam
Loss Function	categorical_crossentropy
Metrics	Accuracy

Results

Different metrics were utilized to assess the models, comprising training and testing accuracy, recall, precision, and the duration needed to complete the model training (table 2). All models were executed for a total of 10 epochs, and the accuracy per epoch can be seen in Figure 2. Each model ran for a total of 10 epochs, the accuracy per epoch is as displayed in figure 2.

The ranking of the top-performing models based on their training and testing accuracy, namely EfficientNetB3, DenseNet121, EfficientNetB5, EfficientNetB4, and EfficientNetB2, is consistent with their performance when evaluated using recall and precision as the metrics, as evidenced by Figures 3 and 4.

The result obtained from this study is as compared with existing literatures in table 3.

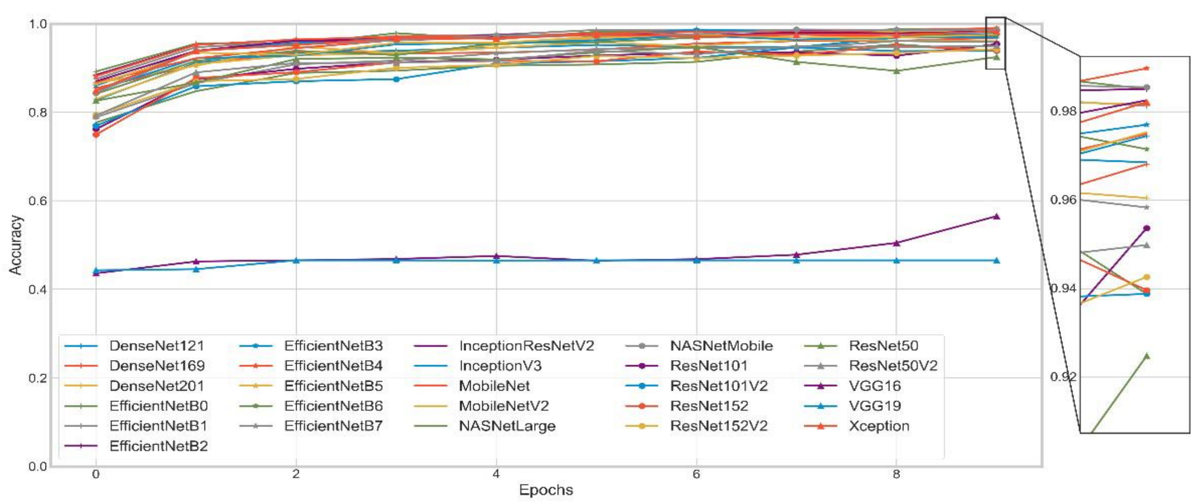

Figure 2. Training Accuracy vs Epochs.

Table 2. Table of result of performance evaluation

Model	Training Accuracy	Test Accuracy	Recall	Precision	Time (s)	Parameters
1. DenseNet121	0.9745	0.9672	0.9590	0.9750	1011	7040579
2. DenseNet169	0.9682	0.8443	0.8361	0.8430	1119	12647875
3. DenseNet201	0.9605	0.7705	0.7705	0.7705	1222	18327747
4. EfficientNetB0	0.9851	0.9180	0.9180	0.9180	959	4053414
5. EfficientNetB1	0.9813	0.9098	0.9098	0.9098	1109	6579082
6. EfficientNetB2	0.9851	0.9590	0.9590	0.9669	2188	7772796
7. EfficientNetB3	0.9771	0.9836	0.9836	0.9836	1165	10788146
8. EfficientNetB4	0.9898	0.9590	0.9590	0.9669	1593	7772796
9. EfficientNetB5	0.9817	0.9590	0.9590	0.9669	2719	28519674
10. EfficientNetB6	0.9716	0.8279	0.8279	0.8279	2487	40967058
11. EfficientNetB7	0.9584	0.8607	0.8525	0.8595	3815	64105370
12. InceptionResNetV2	0.9826	0.8689	0.8689	0.8760	1454	54341347
13. InceptionV3	0.9686	0.8197	0.8197	0.8264	1222	21808931
14. MobileNet	0.9750	0.9344	0.9344	0.9344	832	3231939
15. MobileNetV2	0.9754	0.7869	0.7869	0.7934	858	2261827
16. NASNetLarge	0.9389	0.4672	0.4672	0.4672	4248	84928917
17. NASNetMobile	0.9856	0.2295	0.2295	0.2295	960	4272887
18. ResNet101	0.9537	0.8607	0.8525	0.8740	2074	42664323
19. ResNet101V2	0.9389	0.7213	0.7213	0.7213	1273	42632707
20. ResNet152	0.9397	0.8033	0.8033	0.8033	2260	58377091
21. ResNet152V2	0.9427	0.4262	0.4262	0.4298	2379	58337795
22. ResNet50	0.9248	0.8197	0.8197	0.8197	1065	23593859
23. ResNet50V2	0.9499	0.8197	0.8197	0.8197	1099	23570947
24. VGG16	0.5652	0.6967	0.5492	0.8072	1473	14716227
25. VGG19	0.4654	0.4672	0.0000	0.0000	1048	20025923
26. Xception	0.9822	0.9262	0.9098	0.9407	1404	20867627

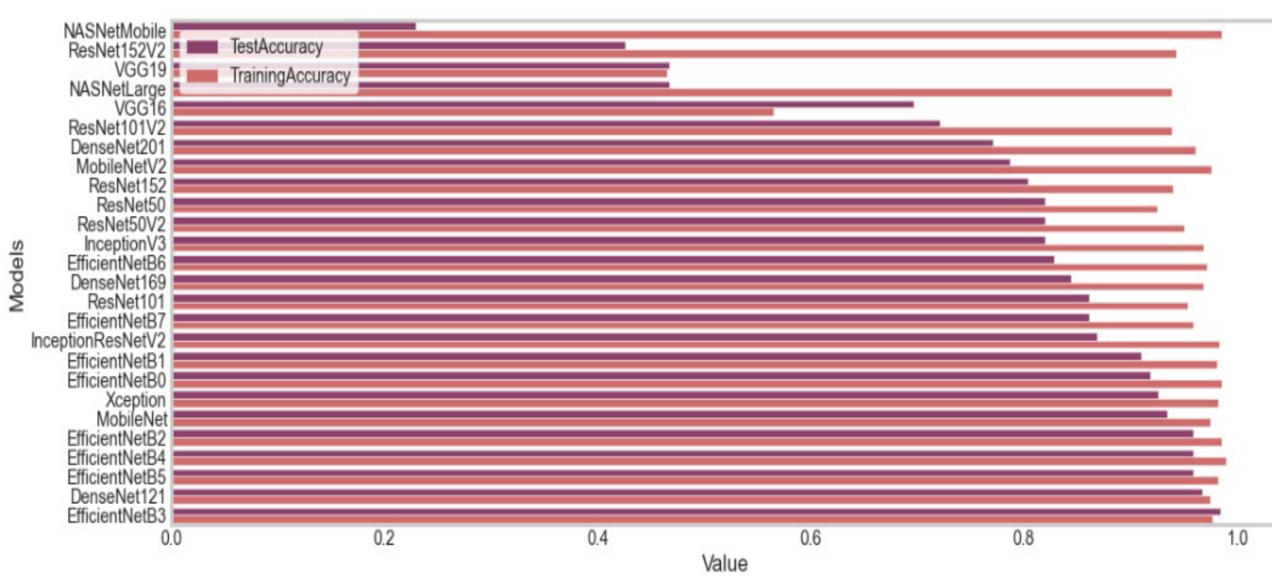

Figure 3. Accuracy plot of models: Training vs Testing.

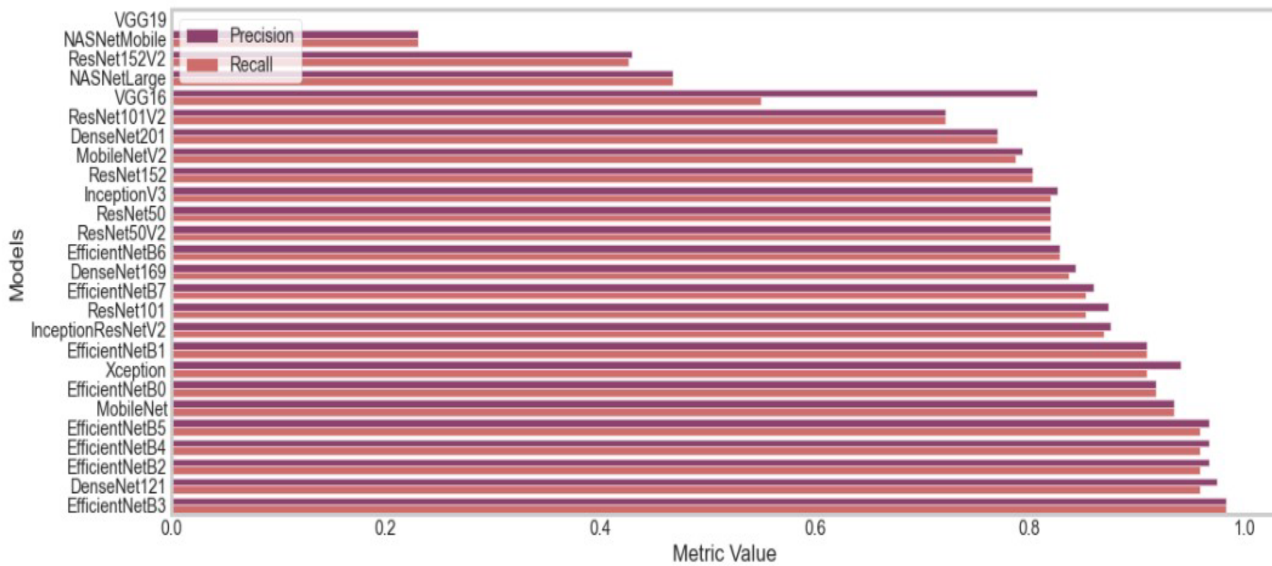

Figure 4. Precision and Recall Evaluation.

Table 3. Comparison of result with existing literature

Ref	Model	Training Accuracy	Testing Accuracy
(Khan, Jue, Mushtaq, & Mushtag, 2020)	Custom-CNN	96%	89%
	VGG-16	90%	87%
	ResNet-50	92%	87%
	Inception-V3	93%	83%
(Soumik & Hossain, 2020)	Inception-V3	99.4 %	-
(Sachin, Punn, Sonbhadra, & Agarwal, 2021)	VGG-16	-	93.15%
	VGG-19 ResNet-50	-	93.8%
	Custom-CNN (MAG-Net)	-	94.2%
		-	98.04%
(Badža & Barjaktarovic, 2020)	Custom-CNN	-	97.28%
(Machiraju & Rao, 2021)	Inception-V3	99.34%	89.0%
This study top performers	DenseNet121	97.45%	96.72%
	EfficientNetB3	97.71%	98.36%
	EfficientNetB2	98.51%	95.90%
	EfficientNetB5	98.17%	95.90%
	EfficientNetB4	98.98%	95.90%

Discussion

Overfitting is a common problem when models show good result on training but perform below par when tested with separate dataset. It can be seen that some of the models as shown in table 2 showed excellent result after the completion of training and validation but when tested with the testing data, the accuracy is below par. Only 8 out of the top 10 best models based on training accuracy made it to the top-10 ranking table based on the test data. This research work, classified better than the original study done on this dataset, it was only able to achieved an accuracy of 91.28% (Cheng, et al., 2015), this is far lesser compared to the top-10 results obtained in this study.

Each of the models have different network architecture and as such different training parameters and consequently different size. This affects the time it takes to train for the same number of epochs. Fig. 5 shows that MobileNet, MobileNetV2, EfficientNetB0, NASNetMobile and DenseNet121 trains within the fastest time possible. Models with large layers and training parameters often takes longer time to train. This is as evidenced with NASNetLarge with over 80 million parameters taking up the longest time to train with a whooping 5000 seconds.

Large, layered models are expected to have optimum performance but in the case of NASNetLarge, this is not exactly the case, with a size of 343610240 bytes it has an accuracy of 0.9856 but when exposed to testing data it correctly classified with an accuracy of 46.74%. It is unclear why VGG 16 and VGG 19 performed woefully for this multiclass classification task, further study is recommended to unravel this.

To ensure that Artificial Intelligence serve as a useful tool for radiologists, it is recommended that upcoming studies evaluate the efficacy of transfer learning and Keras applications in modelling T1-CE, T2, and FLAIR MRI images. However, due to resource limitations, additional testing could not be conducted in the low-resource settings where this research was conducted.

Conclusions

The study confirms the importance of transfer learning as an adoptable approach for building models for brain tumour classification. It also shows that EfficientNetB3, DenseNet121, EfficientNetB2, EfficientNetB5 and EfficientNetB4 are the best pretrained models to use for brain tumour classification tasks.

Abbreviations

CC: Creative Commons; CNN: Convolutional Neural Network; CPU: Central Processing Unit; CV2: Computer Vision 2.0; DL: Deep Learning; GPU: Graphics Processing Unit; ILSVRC: ImageNet Large Scale Visual Recognition Challenge; HP: Hewlett Packard; ML: Machine Learning; MRI: Magnetic Resonance Imaging.

Author Contributions

All authors contributed to this study. All authors gave their final approval.

Competing Interests

We declare no competing interests.

References

[1] Aldape, K., Brindle, K. M., Chesler, L., Chopra, R., Gajjar, A., Gilbert, M. R., Gottardo, N., Gutmann, D.H., Hargrave, D., Holland, E. C. & Gilbertson, R. J. (2019). Challenges to curing primary brain tumours. *Nature reviews Clinical oncology*, *16*(8), 509-520. doi:10.1038/s41571-019-0177-5.

[2] American Association of Neurological Surgeons. (2021). Brain Tumors. (American Association of Neurological Surgeons) Retrieved 02 19, 2022, from https://www.aans.org/en/Patients/Neurosurgical-Conditions-and-Treatments/Brain-Tumors.

[3] American Cancer Society . (2022, 1 12). Key Statistics for Brain and Spinal Cord Tumors. (American Cancer Society) Retrieved 02 19, 2022, from https://www.cancer.org/cancer/brain-spinal-cord-tumors-adults/about/key-statistics.htm.

[4] American Cancer Society. (2022). Survival Rates for Breast Cancer. (American Cancer Society) Retrieved 2 20, 2022, from https://www.cancer.org/cancer/breast-cancer/understanding-a-breast-cancer-diagnosis/breast-cancer-survival-rates.html.

[5] ASCO. (2022). Cancer.net. Retrieved february 12, 2022, from http://www.cancer.net/cancer-types/brain-tumor/statistics.

[6] Badža, M. M., & Barjaktarović, M. Č. (2020). Classification of brain tumors from MRI images using a convolutional neural network. *Applied Sciences*, *10*(6), 1999.. doi:10.3390/app10061999ww.

[7] Cheng, J. (2017). Brain Tumor Dataset. doi:10.6084/m9.figshare.1512427.v5.

[8] Cheng, J., Huang, W., Cao, S., Yang, R., Yang, W., Yun, Z., Wang, Z. & Feng, Q. (2015). Enhanced performance of brain tumor classification via tumor region augmentation and partition. *PloS one*, *10*(10), e0140381. doi:10.1371/journal.pone.0140381.

[9] Keras. (2021). Keras Application. Retrieved February 21, 2022, from https://keras.io/api/applications/.

[10] Khan, H. A., Jue, W., Mushtaq, M., & Mushtaq, M. U. (2021). Brain tumor classification in MRI image using convolutional neural network. *Mathematical Biosciences and Engineering*. doi:10.3934/mbe.2020328.

[11] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, *60*(6), 84-90.

[12] Louis, D. N., Perry, A., Wesseling, P., Brat, D. J., Cree, I. A., Figarella-Branger, D., Hawkins, C., Ng, H. K., Pfister, S. M., Reifenberger, G. & Ellison, D. W. (2021). The 2021 WHO classification of tumors of the central nervous system: a summary. *Neuro-oncology*, *23*(8), 1231-1251. doi:10.1093/neuonc/noab106.

[13] Machiraju, J. L., & Rao, S. N. (2021). Brain Tumor Magnetic Resonance Image Classification: A Deep Learning Approach. doi:10.21203/rs.3.rs-673106/v1.

[14] Sachin, G., Punn, N. S., Sonbhadra, S. K., & Agarwal, S. (2021). MAG-Net: Mutli-task attention guided network for brain tumor segmentation and classification. arXiv:2107.12321v1 [eess.IV] 26 Jul 2021.

[15] Soumik, M. F. I., & Hossain, M. A. (2020, June). Brain tumor classification with inception network based deep learning model using transfer learning. In *2020 IEEE Region 10 Symposium (TENSYMP)* (pp. 1018-1021). IEEE. doi:10.1109/TENSYMP50017.2020.9230618.

[16] Walter, F. M., Penfold, C., Joannides, A., Saji, S., Johnson, M., Watts, C., Brodbelt, A., Jenkinson, M. D., Price, S. J., Hamilton, W. & Scott, S. E. (2019). Missed opportunities for diagnosing brain tumours in primary care: a qualitative study of patient experiences. *British Journal of General Practice*, *69*(681), e224-e235. doi:10.3399/bjgp19X701861