ENHANCED BRAIN TUMOUR PREDICTION USING IMAGE AND NUMERIC DATASETS

R Rekha Sharmily*, B Karthik, T Vijayan

Dept of Electronics and Communication Engineering, Bharath Institute of Higher Education and Research, Chennai, India

ABSTRACT

Detecting brain tumors accurately remains challenging due to the diverse appearances of tumors, their variable sizes, shapes, and structures. The early diagnosis of brain tumor is helpful for planning the treatment. This research utilizes a blend of image and data-derived features and deployed them in two phases. The first phase, of the study employs numeric features in addition to machine learning classifiers such as Random Forest, Logistic Regression, SVC and XGBoost. The VGG16 model is employed for extracting the features and classifying the presence of brain tumor from the Magnetic Resonance Image dataset in the second phase. The recorded accuracy score of 98.96% and 98.5 % attained by both suggested models demonstrates their exceptional performance. Additionally, when comparing the results with other supervised learning algorithms and cutting-edge models, it validates their effectiveness.

Keywords: Brain Tumor, Machine Learning, Transfer Learning, Numeric features, VGG16, Random Forest, Logistic Regression, SVC, XGBoost

1. INTRODUCTION

Brain tumors can either be cancerous or noncancerous and can impact individuals across various age groups, including both children and adults. Regardless of their cancerous nature, brain tumors can affect brain function when they reach a size that exerts pressure on surrounding area. Tumors vary in their growth rates; some proliferate rapidly, while others exhibit a slower growth pace. Approximately one-third of brain tumors are malignant or cancerous. Tumors originating within the brain are referred to as primary tumors. Medical imaging involves a diverse set of procedures that serve as nonsurgical methods to examine the interior of the body. In medical image processing, the primary objectives of image segmentation are to identify tumors or lesions, employ effective machine vision, and achieve accurate and dependable results. Benign tumors, are non-cancerous tissues. The cancerous malignant growth is an uncontrollable growth within the brain, and spread to other tissues. Grade I tumors are small tumors that are typically managed through analyzing and understanding the content of an image. However, the Grade II tumors progresses by exhibiting a small deviation from the normal or typical characteristics. Grade III tumors are malignant and display aggressive growth. Grade IV tumors, characterized by rapid reproduction, pose the highest level of danger. Radiologists frequently opt for the MRI method because of its effectiveness in identifying abnormal cell growth, notably the presence of brain tumors.[1]

The objective of this research is creating a ML and TL model utilizing MRI images and numerical features to distinguish individuals with tumor or not. This research introduces a comprehensive approach for identifying the brain tumors by utilizing both picture and numeric features. The forecasting of brain tumors from MRI images is conducted using VGG16 for both feature extraction and classification.

This study assesses and compares the performance of models that leverage convolutional features with those relying on the original features. A variety of ML and TL, such as Random Forest, Logistic Regression, SVC, XGBoost, and VGG16, are employed for the performance comparison.

2. LITERATURE SURVEY

Oumaima et.al., [2], introduced a model for detecting the presence of brain tumor from both image and the numerical feature datasets. U net and Mobile Net were used for detecting the presence of brain tumor from image dataset. The voting classifier together with. stochastic gradient descent and logistic regression were used for detection of brain tumor from the feature dataset. This model achieved the highest accuracy level when compared with the existing models. The dataset included both images and features obtained from openly accessible Kaggle dataset. Data augmentation leads to noise and optimization of the voting classifier was found to be very difficult.

Shanaka et.al.,[3] conducted a study based on RCNN for brain tumor classification. The Chan-Vese algorithm was used for detecting the outline of the tumor for segmentation. The images were classified using a CNN. Moreover, the region of interest was determined employing a RCNN. The Chan–Vese segmentation method is deployed for segmenting the brain tumor. The results were compared with existing image segmenting models and this model outperformed other existing models.it was determined that further segmenting of brain tumor can be improved by using progressive contour methods.

Priyanka et.al.,[4], suggested a brain tumor classification approach based on VGG-16 and Efficientnet CNN. The classification was performed using Kaggle dataset. The dataset is preprocessed and features are extracted from the input images. The models were trained with the given dataset. The PCA was used for reducing the dimensions. The Efficientnet CNN outperformed the VGG16 approach. Ullah et.al.,[5], presented a frame work for overcoming the limitations of deeplearning. Deeplearning models suffer from overfitting, inorder to overcome overfitting pretrained nine transfer learning approaches. Nine transfer learning models were used for detection and classification of brain tumor and inceptionresnetv2 outperformed all other models.

Sailunaz et.al., [6], has introduced an user friendly web interface for the segmentation of brain tumor. The publically available BRATS dataset was utilized for segmentation. The 2D and 3D segmentation were done through UNet and U Net++.The 3D Unet scored the higher performance metrics. The web interface was incorporated with certain DL model alone.

Hossain et.al.,[7] have conducted an experiment with realtime dataset to segment and detect the brain tumor using two approaches. In the first approach segmentation is performed by Fuzzy C Means method and for classification six ML models were employed and in the second approach segmentation and classification is done with the other rough CNN model. The CNN model has scored better than the other model. Qasem et al., [8] analyzed a machine learning model for classifying of the brain tumor images. The model consists of a preprocess module, morphological operation module and a segmentation module. The segmentation is performed using Watershed algorithm and the KNN machine learning model is used for classification. The research work was experimented with a larger dataset, however the results are less accurate. Lotlikar et.al.,[9] reviewed on various preprocessing methods, machine learning and deep learning models for detecting and classifying the MRI images. The researchers reviewed on five types of machine learning models and found that the boosting methods like GBML gave better results compared to other models. They also investigated various deeplearning and transfer learning models and found that transfer learning models scored higher accuracies with small datasets.

Swarup et.al.,[10] proposed and compared the performance of two models namely the GoogleNet and AlexNet. The Convolutional Neral Network model was used for preprocessing the MRI images. The Googlenet outperformed the Alexnet in performance metrics. The Googlenet consists of more number of layers and less number of parameters parameters when compared with the Alexnet model.

Abdolkarimzadeh et. al., [11] analyzed a method forbrain tumor detection with the optimization of in finite element analysis, where the inverse dynamic approach is fully coupled. This method aims to estimate the variable pressure boundary and its resultant effects. The experiments are conducted with varying and boundaries where the pressure is consistently maintained at a uniform level. The varying pressure boundaries are found to be less error prone. Detailed information concerning the structure and organization of living organisms, especially in terms of anatomy.

Table 1. Comparison of related works

S.No	Methodology	Remarks
[2]	U Net, Mobile Net, Voting classifier	Noise prone, optimization was hard
[3]	RCNN	Can be further improved by active contouring
[4]	VGG-16 and Efficientnet	Less accurate.
[5]	Inceptionresnetv2 and 8 DL methods	Overfitting problem
[6]	UNet and U Net++	Can work on few models only
[7]	Six ML methods, CNN	-
[8]	KNN	Less accurate
[9]	Review on ML, DL, TL models	TL models performed better with less dataset
[10]	Google Net, AlexNet	Reduced computational requirement
[11]	Variable pressure boundary	Anatomy details needed

3. MATERIALS AND METHOD

Brain tumor detection was done in two stages using machine learning and transfer learning methods. The numerical data and image data were given as inputs. The openly accessible Kaggles Brain Tumor dataset was chosen as the input dataset. The dataset comprises of both numerical and corresponding image datasets. The number of features taken in to account are thirteen extracted from 3762 representations. The dataset consists of tumor and no tumor data in terms of numbers and corresponding images.

3.1 Machine Learning and Deep Learning Classifiers

Moreover, machine learning classifiers namely Random Forest, Logistic Regression, SVC and XGBoost were employed. The Random Forest is an ensemble learning model used for classification and constructed from decision trees. The logistic regression is a supervised learning model based on estimating the likelihood that a certain instance will belong to a certain class, mainly used for binary classification. C-Support Vector Classification, SVC supports kernels namely linear, polynomial, radial basis function (RBF), and sigmoid. XGBoost is an optimized distributed gradient boosting model based on ensemble learning. The input data in the form of features are given as inputs to the machine learning models. The ML models perform training and testing in the 80 is to 20 splitted dataset.

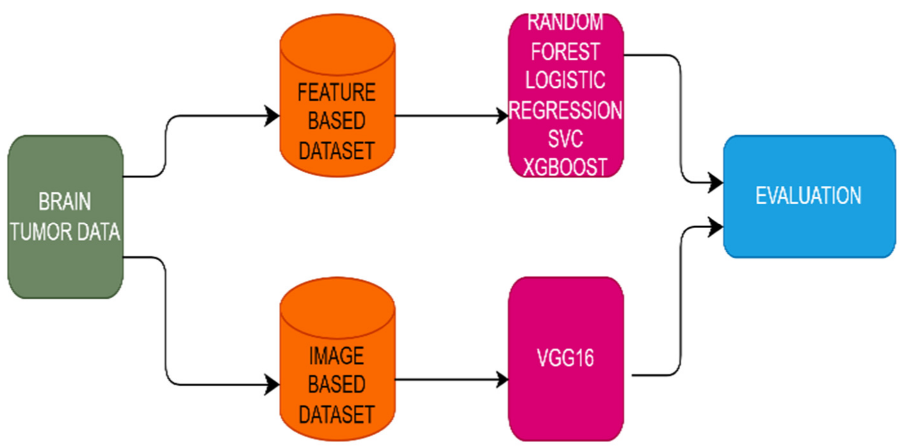


Fig.1.Proposed Brain tumor detection model using image and feature based dataset

VGG16 is a pretrained deep learning model used for feature extraction and classification. It consists of thirteen convolutional layers, five maxpooling layers and three fully connected layers. Moreover, the convolutional layers extract the informations of the brain tumor. The maxpooling layers refines the most significant features of the image, thus reducing the number of features. The output is fed in to the fully connected layer. The features are flattened to one dimensional feature array and were classified in to tumor and no tumor classes. The MRI images are given as input to the VGG16 approach and are trained and tested for accuracy. The network is trained for 30 epochs with a batchsize of 32. The accuracy of both ML and DL methods are compared.

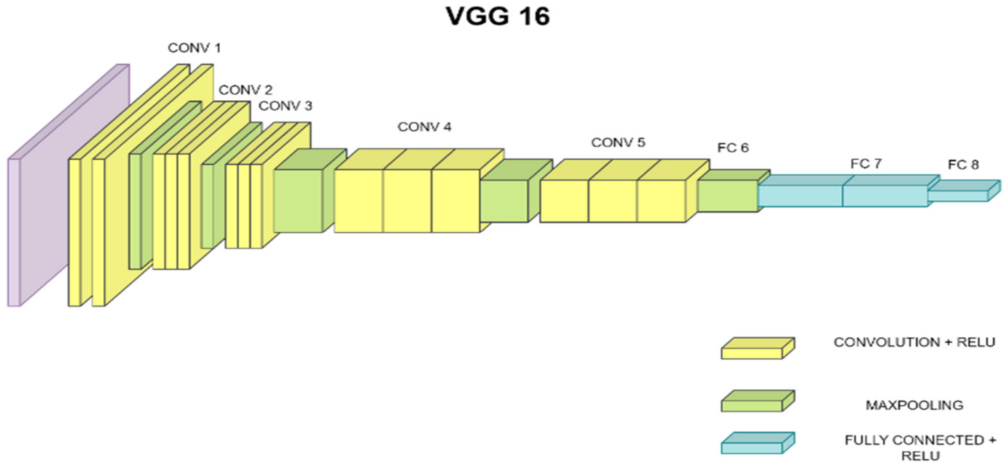
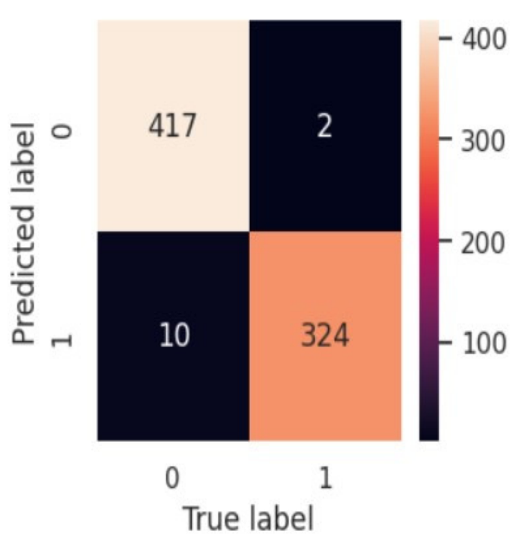


Fig.2. Architecture of VGG16

4.RESULTS AND DISCUSSION

The research was conducted in a 12th generation i5 core machine with Windows 11 and 1080 NVIDIA graphics support. The implementations were done using Python coding.

The dataset was applied to the machine learning approaches. The XGBoost ensemble model gained the highest accuracy of 98.5 %. The Random Forest model scored about 98.4 %, followed by SVC and Logistic Regression with 97.9 % and 97.3 % respectively. The performance of the machine learning models were valuated with confusion matrix.



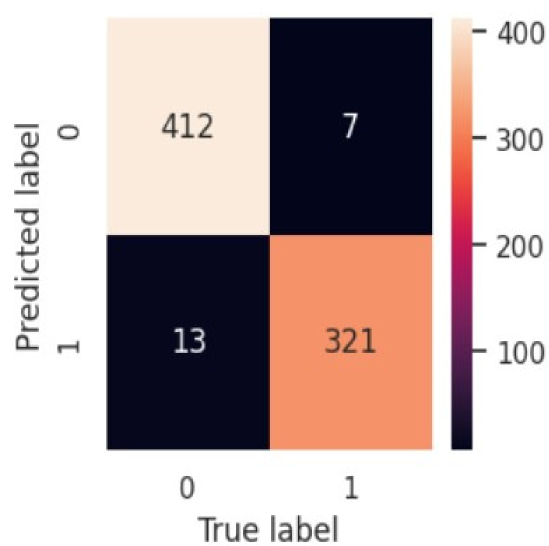
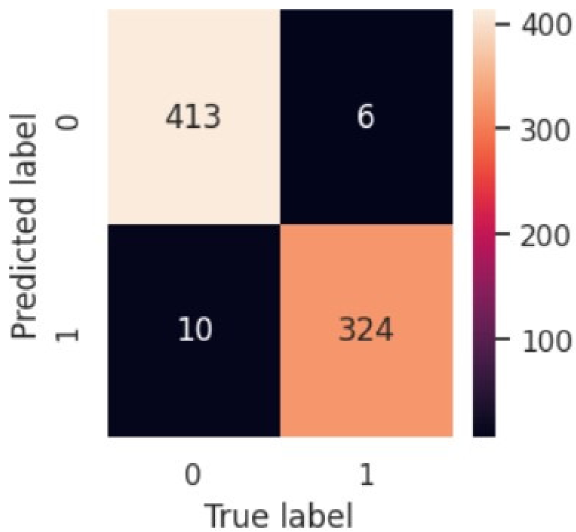
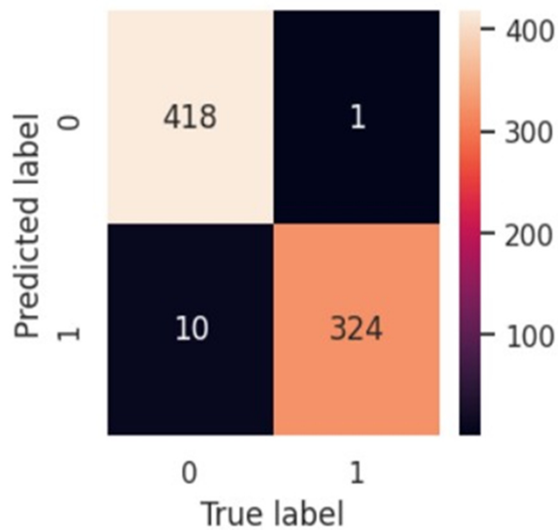


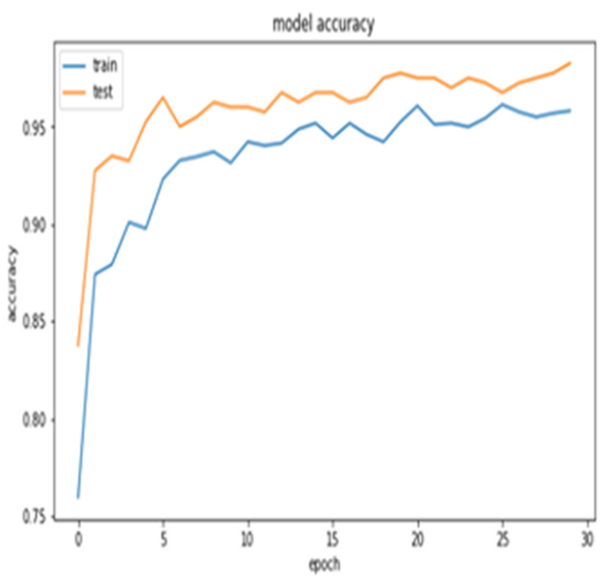
Fig 3.a.Confusion matrix of random forest b. Confusion matrix of Logistic Regression



c. Confusion matrix of SVC



d. Cofusion matrix of XGBoost



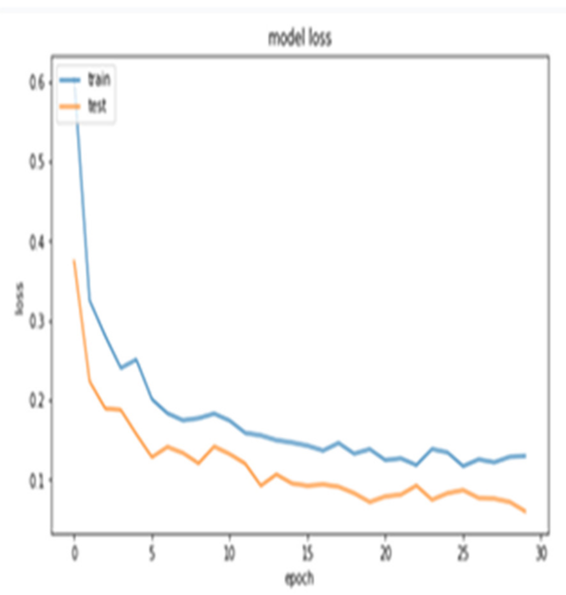


Fig.4. VGG16 accuracy and loss plot

The training and testing of VGG16 is done with 80% and 20% of the data respectively. However, the model was trained with the MRI image dataset and tested for classifying the presence of brain tumor and the VGG16 consists of convolutional layers utilized for extraction of features. However, the classification of tumor is done by the final densenet layer with the sigmoid function. Maxpooling layers are employed for feature reduction inorder to overcome overfitting. The

VGG16 model is trained with 30 epochs and batchsize of 32. This approach achieved a highest accuracy of 98.96%. Furthermore, the models performance is compared with the machine learning models performance. The comparison table for the machine learning and deep learning models are tabulated below.

Table2. Machine Learning and Transfer Learning Models Performance

S.No	MODEL	ACCURACY
1.	VGG 16	98.96%
2.	XGBoost	98.5 %
3.	Random Forest	98.4 %
4.	SVC	97.9 %
5.	Logistic Regression	97.3 %

The graphical representation of the performance of analysed models is given as follows,

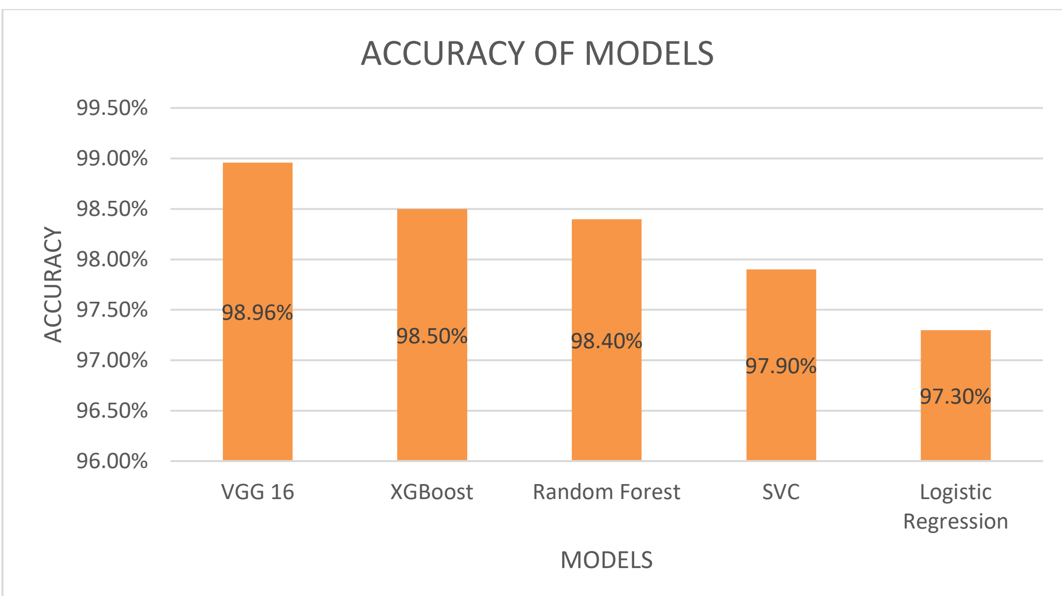


Fig.5. Graphical representation of the models

As mentioned in the above table, it is found that the pretrained deep learning models outcome is higher than the machine learning models. The XGBoost technique scored the higher accuracy than others.

5.CONCLUSION

This research work utilized two types of datasets numeric and images. The numeric features were classified by machine learning techniques like Random Forest, Logistic Regression and XGBoost. The XGBoost classifier scored accuracies as 98.4 %. The pretrained VGG16 classifier scored an accuracy of 98.96%. Although the attained accuracy is higher than the machine learning models, the training time for the Deep learning models are high. Moreover, it is helpful for the early diagnosis of brain tumor so that the treatment can be provided at early stage which is beneficial for mankind. This algorithm can be used for detecting types and grades of brain tumor. We intend to implement this approach in additional subject areas, extending beyond the current discourse on the abundance of extensive data. Alternatively, we can explore alternative learning that involve the exchange of information while adhering to the same proposed approach.

REFERENCES

1. Zahoor, M.M., Qureshi, S.A., Khan, A., Rehman, A.u., & Rafique, M. (2022). A novel dual-channel brain tumor detection system for MR images using dynamic and static features with conventional machine learning techniques. *Waves in Random and Complex Media*, pp. 1–20.
2. Saidani, O., Aljrees, T., Umer, M., Alturki, N., Alshardan, A., Khan, S.W., Alsubai, S., & Ashraf, I. (2023). Enhancing prediction of brain tumor classification using images and numerical data features. *Diagnostics (Basel)*, 13(15), 2544. https://doi.org/10.3390/diagnostics13152544
3. Chatterjee, S., Nizamani, F.A., Nürnberger, A., et al. (2022). Classification of brain tumours in MR images using deep spatio-spatial models. *Scientific Reports*, 12, 1505. https://doi.org/10.1038/s41598-022-05572-6
4. Gunasekara, S.R., Kaldera, H.N.T.K., & Dissanayake, M.B. (2021). A systematic approach for MRI brain tumor localization and segmentation using deep learning and active contouring. *Journal of Healthcare Engineering*, 2021, 6695108. https://doi.org/10.1155/2021/6695108
5. Yan, F., Chen, Y., Xia, Y., Wang, Z., & Xiao, R. (2023). An explainable brain tumor detection framework for MRI analysis. *Applied Sciences*, 13, 3438. https://doi.org/10.3390/app13063438
6. Modiya, P., & Vahora, S. (2022). Brain tumor detection using transfer learning with dimensionality reduction method. *International Journal of Intelligent Systems and Applications in Engineering*, 10(2), 201–206. https://ijisae.org/index.php/IJISAE/article/view/1310

7. Ullah, N., Khan, J.A., Khan, M.S., Khan, W., Hassan, I., Obayya, M., Negm, N., & Salama, A.S. (2022). An effective approach to detect and identify brain tumors using transfer learning. *Applied Sciences*, 12, 5645. https://doi.org/10.3390/app12115645

8. Saeedi, S., Rezayi, S., Keshavarz, H., et al. (2023). MRI-based brain tumor detection using convolutional deep learning methods and chosen machine learning techniques. *BMC Medical Informatics and Decision Making*, 23, 16. https://doi.org/10.1186/s12911-023-02114-6

9. Sailunaz, K., Bestepe, D., Alhajj, S., Özyer, T., Rokne, J., & Alhajj, R. (2023). Brain tumor detection and segmentation: Interactive framework with a visual interface and feedback facility for dynamically improved accuracy and trust. *PLoS ONE*, 18(4), e0284418. https://doi.org/10.1371/journal.pone.0284418

10. Meshram, P.A., Joshi, S.S., & Mahajan, D.A. (2023). Image and video processing; computer vision and pattern recognition. *arXiv preprint arXiv:2305.06025*. https://doi.org/10.48550/arXiv.2305.06025

11. Hossain, T., Shishir, F., Ashraf, M., Nasim, M.A., & Shah, F. (2019). Brain tumor detection using convolutional neural network. *2019 International Conference on Advanced Science and Engineering (ICASERT)*. https://doi.org/10.1109/ICASERT.2019.8934561

12. Qasem, S.N., Nazar, A., Qamar, A., Shamshirband, S., & Karim, A. (2019). A learning based brain tumor detection system. *Computers, Materials & Continua*, 59(3), 713–727.

13. Lotlikar, V.S., Satpute, N., & Gupta, A. (2022). Brain tumor detection using machine learning and deep learning: A review. *Journal of Intelligent & Fuzzy Systems*, 18(6), 604–622.

14. Swarup, C., Singh, K.U., Kumar, A., Pandey, S.K., Varshney, N., & Singh, T. (2023). Brain tumor detection using CNN, AlexNet & GoogLeNet ensembling learning approaches. *Electronic Research Archive*, 31(5), 2900–2924. https://doi.org/10.3934/era.2023146

15. Abdolkarimzadeh, F., Ashory, M.R., Ghasemi-Ghalebahman, A., & Karimi, A. (2021). Inverse dynamic finite element-optimization modeling of the brain tumor mass-effect using a variable pressure boundary. *Computer Methods and Programs in Biomedicine*, 212, 106476.