2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT) | 979-8-3503-7024-9/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICCCNT61001.2024.10725957

Pretrained DeIT for Brain Tumor Classification: A Fine-Tuning Approach with Label Smoothing

Shankar Singh Mahanty
Dept. of Computer Science & Engineering,
C.V. Raman Global University, Bhubaneswar, Odisha
shankarmahanty1412@gmail.com

Debendra Muduli
Dept. of Computer Science & Engineering
C.V. Raman Global University, Bhubaneswar, Odisha
debendra.muduli@cgu-odisha.ac.in

Ankur Kumari
Dept. of Computer Science & Engineering
C.V. Raman Global University, Bhubaneswar, Odisha
ankurkumari1830@gmail.com

Santosh Kumar Sharma
Dept. of Computer Science & Engineering
C.V. Raman Global University, Bhubaneswar, Odisha
santoshku.sharma17@gmail.com

***Abstract*—This research introduces a novel approach for brain tumor detection utilizing transformer-based deep learning architectures. Conventional methods and CNN-based techniques suffer significant scalability and performance issues, highlighting the necessity for more sophisticated solutions. By leveraging transformer architectures, our methodology integrates precise preprocessing techniques such as contour analysis, endpoint detection, channel-wise pixel value normalization, AdamW optimization, and the LabelSmoothingCrossEntropy criterion for improved model performance. We utilized SMOTE for class balancing, and our augmentation strategies further increased dataset diversity and generalization. The model is based on a pre-trained DeIT framework with a fine-tuned, redesigned head architecture and final layers, achieved an outstanding accuracy of 97.42% in brain tumor detection, along with notable average precision and recall metrics. These outcomes exceed those of existing models, promising enhanced treatment strategies, patient care, and early tumor detection capabilities in medical imaging, representing a significant advancement in the field.**

***Index Terms*—Brain Tumor, DeIT Transformer, Contour Analysis, LabelSmoothing CrossEntropy, SMOTE, Early Stopping**

I. Introduction

Brain tumors, including glioma, meningioma, and pituitary tumors, pose a major challenge in healthcare due to their uncontrolled cell proliferation, leading to increased intracranial pressure. This necessitates precise diagnosis and tailored treatments to achieve optimal clinical outcomes.

Brain tumors' variety and severity are influenced by their location, size, and malignancy. Gliomas, produced from glial cells, range in grade from low to high, with high-grade gliomas like glioblastoma being aggressive. Meningiomas are benign tumors in the meninges but can cause problems if grown too large. Pituitary tumors, located in the pituitary gland, can disrupt hormonal balance and cause various health problems. Understanding these tumor categories is crucial for research and treatment development [1].

Traditional methods for diagnosing brain tumors often struggle with accurate detection, localization, and classification. Non-invasive imaging techniques like MRI have become crucial, providing superior visual results on soft tissues and comprehensive data on tumor characteristics and distribution within the brain.

Deep learning methodologies such as CNN and transformers have significantly improved the efficiency and precision of brain tumor diagnosis using magnetic resonance images by extracting intricate details, aiding in precise tumor localization, classification, and prognostic assessment [2].

The proposed transformer model uses pre-trained models like the Data-efficient Image Transformer (DeIT) with fine-tuned final layers and a redesigned head, eliminating the need for manual feature extraction. This results in a robust and reliable model with low parameters for learning discriminant features. The model accurately distinguishes between brain tumors and non-tumorous tissue on magnetic resonance imaging, making it a valuable tool for medical professionals to make accurate diagnosis and informed treatment decisions in brain tumor cases.

This is how the rest of the paper is structured. The relevant work is summarized in Section II. Section III illustrates the stated methodology. The comparative analysis and experimental findings are covered in Section IV. Finally, the concluding remarks were included in Section V.

II. Related works

Many techniques for computer-assisted diagnosis (CAD) have been put forth to help classify brain tumors. A large number of these models are based on labeled data and supervised learning approaches, in which convolutional neural networks (CNNs) and traditional machine learning classifiers are developed [3]. For example, Arunkumar et al. conducted a study on brain tumor categorization using artificial neural networks. They removed high-frequency components from brain MRIs using the Fourier transform, extracted the region of interest using thresholding, and removed texture features and directed gradient histograms for classification [4]. User Ghassemi et al. utilized a pre-trained GAN to acquire additional MRI slices [5]. Tandel et al. also used a pairwise GAN to improve brain MRI data and generated patient-level data using

majority voting, demonstrating a successful classification of gliomas [6]. Maheshwari et al. achieved 95% accuracy in classifying data on brain tumors using CNN models and the Resnet50 pre-trained approach, utilizing transfer learning. [7]. The authors of [8] provided a modified CNN technique with customized layers, yielding an accuracy rate of 86.56%. The VggNet-Lstm hybrid CNN model, which has the best accuracy rate of 84%, was used by the authors in [9]. They made comparisons between it and other hybrid models, including ResNet-Lstm and Alexnet-Lstm. Mohsen et al. [10] provided a novel CAD model that utilized DWT as a feature extractor and DL methods, achieving a precision rate of 93.94%. Zhang et al. [11] have presented a number of hybrid algorithms for the classification of brain images, with strong classification precisions. Das et al. presented a unique multiscale geometric analysis (MGA) technique termed the Ripplet transform (RT) [12] to extract important features from brain magnetic resonance images. The brain pictures were classified using least square SVM (LSSVM), while PCA was utilized for feature reduction. Parnian Afshar et al. proposed capsule networks(CapsNets) to overcome the shortcomings of CNN for brain tumor (BT) classification [13]. The study by [14] introduces the Transformer-Enhanced Convolutional Neural Network (TECNN), a hybrid deep learning model for effective BT classification using MRI images, combining self-attention mechanisms for long-range dependencies with CNNs for local feature extraction.

It is evident that the inherent benefits of feature extraction and reduction in deep learning make these principles extensively applicable in a variety of applications. Inspired by this, we suggest a DeIT transformer-based method for magnetic resonance imaging based brain lesion classification.

III. Proposed Methodology

The primary objective of this study is to automate the identification of brain tumors using brain MRI data. The system architecture, depicted in **fig. 1**, employs a data-efficient image transformer (DeIT) model, pre-trained on datasets like ImageNet. In which the final layers are adapted with a modified head for fine-tuning and classification purposes. Additionally, the methodology integrates the LabelSmoothingCrossEntropy criterion, AdamW optimizer, ReduceLROnPlateau scheduler, and EarlyStopping to enhance model performance.

This section details a comprehensive methodology for brain tumor classification, including dataset gathering, image preprocessing, dataset balance, augmentation approaches, suggested model, regularization methods, and hyperparameter configuration. Each stage of the proposed system is detailed as follows:

A. *Dataset used*

The brain tumor data set used in this study is publicly available on Kaggle [15]. It comprises magnetic resonance imaging (MRI) scans classified into four distinct categories: glioma, meningioma, notumor and pituitary tumors. A combined collection of 5712 MR imaging scans makes up the

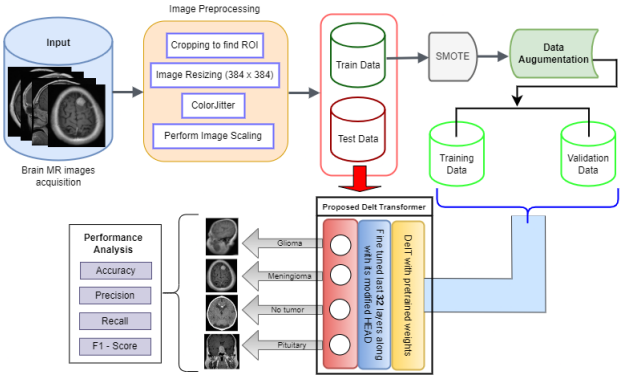

Fig. 1. The overall work flow of the proposed model

training set, while the testing set is made up of 1311 MRI images distributed between the four classes as tabulated in **table I**.

TABLE I
Distribution of Brain Tumor MRI Images in Training and Testing Sets

Class	Training Set	Testing set
Glioma	1321 images	300 images
Meningioma	1339 images	306 images
No tumor	1595 images	405 images
Pituitary tumors	1457 images	300 images

B. *Pre processing of images*

This section includes a thorough explanation of the procedures utilized during the pre-processing stage:

1) Cropping of images: To effectively delineate the Region of Interest (ROI), contour analysis and the detection of extreme points are used to crop every image in the collection. **fig. 2** shows a visual representation of this procedure.

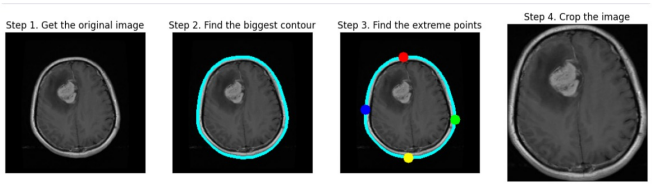

Fig. 2. Illustration of the process of cropping to find the Region of Interest

2) Image resizing: The scans were amended to a set 384x384 pixel size for streamlined uniform input image dimensions and minimize computational overhead. This step is essential when employing pre-trained deep learning models. Additionally, resizing to a standardized size improves computational efficiency during training and inference, leading to an overall more efficient process.

3) ColorJitter transformation: We utilized PyTorch's torchvision library for the ColorJitter transformation, which randomizes adjustments in brightness, contrast, saturation, and hue during training. This technique introduces variability to the images, aiding in enhancing model generalization capabilities.

4) Image scaling: We used ImageNet means to execute picture scaling in order to standardize and normalize the pixel values channel-by-channel.The procedure entailed subtracting the RGB channels' mean pixel values of 0.485, 0.456, and 0.406, respectively, then dividing the result by the standard deviation values of 0.229, 0.224, and 0.225 for every channel, in order to align the pixel values within a uniform range. This normalization step ensures that the pixel values are centered around zero and have a standard deviation of one for each color channel, which helps in the convergence of subsequent deep learning models.

5) Data balancing using SMOTE: We utilized the Synthetic Minority Over-sampling Technique (SMOTE) to tackle the class imbalance within the data set. With a k-nearest neighbor (kNN) parameter set to 5, SMOTE generated synthetic samples for minority classes by interpolating between existing samples. This approach balanced the class distribution, enhancing the model's capacity to generalize effectively across all classes.

6) Data Augmentation: Data augmentation plays a vital role in improving the diversity and robustness of the training data set. Employing transformations like rotation, flipping, scaling, and shearing to the training pictures, we artificially increase the dataset's size and introduce variations that help the model generalize better to unseen data. For each class in the training dataset, we applied augmentation techniques four times, effectively quadrupling the size of the training data. This approach not only increases the diversity of training samples but also aids in preventing overfitting by exposing the model to a wider range of variations within the data. After applying augmentation on the balanced training dataset, **Table II** presents a tabulation of the number of pictures in each category.

TABLE II
QUANTITY OF POST-AUGMENTATION IMAGES FOR EVERY CLASS IN THE TRAINING DATASET

Classes	Glioma	Meningioma	Notumor	Pituitary
Training dataset	6380	6380	6380	6380

7) Data Splitting using StratifiedShuffleSplit: Adding data to training and validation sets is an essential step in creating robust deep learning models. In our approach, we allocate 30% of the training dataset for validation purposes, ensuring that the model performance is assessed on unseen data during training. The model was trained using the remaining 70% of the dataset, allowing the model to a diverse range of cases. To maintain class balance and ensure representative samples in both the training and validation sets, we employed the StratifiedShuffleSplit technique. This technique splits the dataset into training and validation sets at random, thereby preserving the distribution of classes within the dataset. By stratifying the data based on class labels, we mitigate the risk of introducing bias and ensure that the model receives sufficient exposure to each class during training.

C. *Proposed model*

In the realm of deep learning for image classification, Convolutional Neural Networks (CNNs) have been the cornerstone due to their ability to extract spatial features effectively. However, with the emergence of transformers like Data-efficient Image Transformer (DeIT), a paradigm shift has occurred. DeIT builds upon the strengths of traditional CNNs but introduces revolutionary concepts such as attention mechanisms and positional encodings, inspired by the success of transformers in natural language processing. These innovations enable DeIT to excel in tasks like medical image analysis, particularly in scenarios such as brain tumor identification, where capturing global dependencies and intricate details is paramount. One of DeIT's standout features is its data efficiency, making it adept at achieving high performance with limited training data, a common challenge in medical imaging. Additionally, DeIT seamlessly integrates with transfer learning techniques, leveraging pre-trained models to enhance its capabilities further. This combination of advanced architectural elements and efficient learning strategies positions DeIT as a game-changer in image classification, offering superior performance and adaptability, especially in critical domains like medical diagnostics.

Inspired by this, in the proposed model for image classification, we integrated the deit_base_patch16_384 varriant of the Data-efficient Image Transformer (DeIT) architecture with transfer learning techniques. This involved fine-tuning the last 32 layers of the DeIT model, incorporating a modified head architecture, and concluding with a linear layer having nodes equal to the number of classes for multiclass classification tasks of brain tumor identification. This approach overcomes the need for a separate classifier, Additionally, it capitalizes on DeIT's advanced attention mechanisms and positional encodings, tailoring the model to the specific complexities of medical image processing. This approach streamlines the model and enhances its efficiency.

```
Sequential(
  (0): Linear(in_features=768, out_features=512, bias=True)
  (1): BatchNorm1d(512, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
  (2): ReLU()
  (3): Linear(in_features=512, out_features=256, bias=True)
  (4): BatchNorm1d(256, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
  (5): ReLU()
  (6): Dropout(p=0.5, inplace=False)
  (7): Linear(in_features=256, out_features=4, bias=True)
)
```

Fig. 3. Summary of proposed DeIT model head

The architectural layout of the head module used in our transformer-based model for image classification tasks is shown in **Fig 3** The simple designed head comprises sequential layers: a linear layer with 768 input features, derived from the output of last layers of the DeIT architecture, and 512 output features, followed by batch normalization and rectified linear unit (ReLU) activation for introducing non-linearity and stabilizing training. Subsequently, a second linear layer with 512 input features and 256 output features is incorporated, again followed by batch normalization and ReLU activation. A dropout layer with a probability of 0.5 is applied to prevent

overfitting, and finally, a linear layer with 256 input features and 4 output features, representing our classification classes, completes the head architecture. This design enables the model to learn intricate patterns in the data while regularizing its training for faster convergence and improved generalization performance.

In the current investigation, the following regularization methods and components were utilized.

Batch Normalization Batch normalization facilitates autonomous learning for each model layer by normalizing the inputs across a mini-batch. This normalization helps each layer's inputs converge more quickly and enhances generalization by minimizing internal covariate shift. The primary goal is to keep the activation values distributed consistently during training.

Dropout A random percentage of input units is set to zero throughout each training cycle as a regularization strategy to minimize overfitting. By encouraging the model to acquire more robust features and minimizing co-adaptation across neurons, this method improves the model's generalization performance on unseen data.

ReLU Activation Chosen for its simple structure and effectiveness in introducing non-linearity, assisting the model to learn complex patterns and for mitigating the gradient vanishing and gradient explosion problem. The ReLU function provides decent calculation and optimization performance. The function is described in equation (1) as follows:

$$\mathrm{Relu}(x) = \begin{cases} x, & \text{if } x \geq 0 \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

Early stopping is an optimization method designed to reduce overfitting without compromising the model's accuracy. Its primary goal is to halt the training process before overfitting leads to poor performance. In our proposed model, we set the number of epochs to 50. However, early stopping was triggered at epoch 30 based on validation accuracy.

AdamW Optimizer a variant of the Adam optimizer, stands out for its adaptive learning rate mechanism and weight decay regularization. This combination not only aids in faster convergence by adjusting learning rates for each parameter individually, but also helps prevent the model from getting stuck in local minima by effectively handling weight updates.

LabelSmoothingCrossEntropy Criterion Employed as a loss function to address noisy labels and enhance model robustness against overfitting by smoothing the ground truth distribution. It not only enhances model robustness against overfitting, but also improves generalization performance on unseen data, resulting in a more stable and reliable training process.

ReduceLROnPlateau Scheduler Used to dynamically reduce the learning rate when the validation metrics plateau, enabling the model to optimize its parameters and improve performance during training.

During the training phase, the hyperparameters of the proposed transformer model are tuned experimentally; **Table III** lists the optimal hyperparameters.

TABLE III
HYPER-PARAMETER SET UP FOR PROPOSED TRANSFORMER MODEL

Hyper parameters	Values
Training batch size	64
Validation batch size	32
Test batch size	16
Initial learning rate	0.0001
Weight decay	0.01
Patience	5
Scheduler Mode	max
Scheduler Factor	0.1

During the model learning process, the same hyper parameters are used with 30 epochs. The best-performing model weights were then saved for future use or evaluation.

IV. EXPERIMENTAL RESULT AND ANALYSIS

In this section we describe our research and findings for our approach to brain tumor detection. To evaluate the efficacy of our method, we constructed an experimental framework and measured important parameters such as accuracy, sensitivity, specificity, and the F1 score. Accuracy indicates the overall accuracy of the classification outcomes. Sensitivity measures correct positive identifications. Specificity assesses accurate negative identifications and the F1 score provides a fair evaluation. These measurements provided valuable information on the effectiveness of our method, allowing for well-informed evaluations and comparisons with other brain tumor detection methods.

A. *Environment setup*

The proposed model's efficacy was validated through simulation experiments conducted on a Kaggle notebook environment. The notebook is equipped with a cumulative CPU, a RAM size of 29 GB, a disk size of 73 GB, and utilizes a 16 GB GPU P100 accelerator for computations. These resources were instrumental in assessing the model's performance and scalability under realistic computing conditions.

B. *Performance evaluation metrics*

As discussed in Section III, the effectiveness of our proposed method is evaluated through various metrics. These metrics are utilized to gauge different facets of the method's performance. The definition of these metrics is presented as follows:

- Accuracy: The model's accuracy is calculated as the percentage of correctly identified samples relative to the total number of samples. The following equation (2) is used to carry out this calculation:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

- Precision: This measure evaluates the accuracy of positive predictions by calculating the percentage of correctly recognized positive instances among all anticipated positive cases, which include both true positives and false positives. The following formula (3) represents precision:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{3}$$

- Recall: This metric assesses the model's accuracy in identifying positive instances, calculated by dividing the number of true positives by the sum of true positives and false negatives. Mathematically, recall is given by equation (4).

$$\text{Recall} = \frac{TP}{TP + FN} \tag{4}$$

- F1-score: The F1-score offers a fair assessment of the model's performance by integrating recall and precision into a single statistic. The subsequent equation (5) demonstrates it.

$$\text{F1-score} = 2 \times \frac{(\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})} \tag{5}$$

- Confusion Matrix: It accurately classifies brain tumor instances, comparing actual and predicted positive/negative classifications such as true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN). The confusion matrix demonstrated in **fig. 4** corresponds to the proposed model applied to test data. This reflects the effectiveness of our approach in categorizing cases of brain tumors.

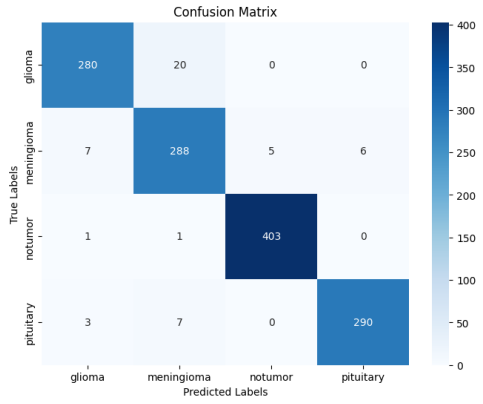

Fig. 4. Confusion matrix obtained for the model

Table IV presents the classification report for the proposed model, including the support values that demonstrate the natural distribution of classes in the test data.

TABLE IV
CLASSIFICATION REPORT FOR BRAIN TUMOR

Class	**precision**	**recall**	**f1-score**	**support**
glioma	0.98	0.93	0.95	300
meningioma	0.91	0.94	0.93	306
notumor	0.99	1.00	0.99	405
pituitary	0.98	0.97	0.97	300
accuracy			**0.97**	1311
macro avg	0.97	0.97	0.97	1311
weighted avg	0.97	0.97	0.97	1311

C. *Comparative analysis*

1) With existing CAD models: To demonstrate the potency of the proposed model with state-of-the-art CAD models, they are simulated in a similar environment. **Table V** display the comparison analysis.

TABLE V
PERFORMANCE COMPARISON WITH CAD MODELS

	Model	Accuracy
1	VGG16	92.50%
2	VGG19	90.00%
3	EfficientNet-B0	94.00%
4	DenseNet	91.00%
5	InceptionV3	93%
6	**Proposed DeIT**	**97.42%**

2) With previously proposed model: Comparison of outcomes achieved by recent deep learning approaches for automated brain tumor detection using the Brain MRI dataset is presented in **table VI**. The results clearly demonstrate the superior performance of the developed approach compared to other existing systems.

TABLE VI
PERFORMANCE COMPARISON WITH EXISTING MODELS

Authors	Methods	Accuracy
Mohammadi et al. [13]	Custom Model	93%
Charfi et al. [16]	PCA	9O%
Maheshwari et al. [7]	ResNet50	95%
Citak et al. [17]	SVM, multilayer perceptron and logistic regression	93%
Vani et al. [18]	SVM	81.48%
KV Durga et al. [19]	InceptionV3 + VGG19	96%
Proposed Method	**Fine tuned DeIT**	**97.42%**

Fig. 5 illustrates the training accuracy and validation accuracy for each epoch, while **Fig. 6** displays the convergence of loss during training and validation across different iterations for the proposed DeIT model.

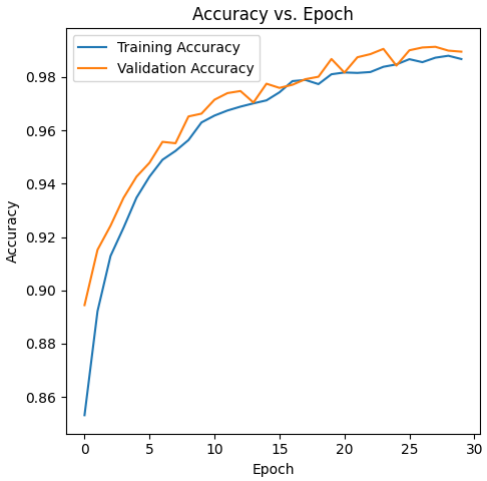

Fig. 5. Training & Validation accuracy plot of proposed model

It is evident that The proposed transformer model outperforms other methods in accuracy and requires less learning parameters, offering quick training speed with minimum training sample as compared to other pre-trained CNN models [20]. Therefore, this model offers several benefits compared to other current computer-aided diagnostic models for brain tumor identification, including:

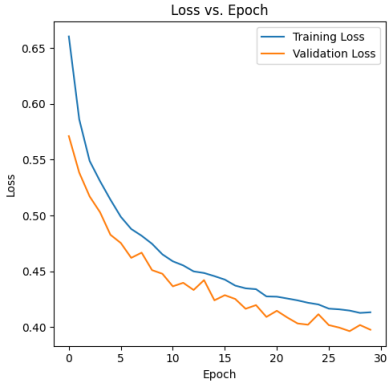

Fig. 6. Loss convergence plot obtained for proposed model

- The proposed transformer model streamlines brain tumor classification procedures by eliminating traditional feature extraction steps, reducing feature matrices' dimensionality, and facilitating classification more efficiently.
- The model efficiently extracts features during training, outperforming existing methods.
- The transformer model surpasses previous models with a validation accuracy of 99.14% in just 30 epochs, while using less training sample and computational power.
- The fine-tuning of the final layers and model head may mitigate overfitting, enhance task-specific adaptation, accelerate convergence, and address domain shift.
- In comparison to other existing CAD models, the DeIT model offers better accuracy.

V. Conclusion

In conclusion, this paper introduces a robust mechanism that encompasses a data-efficient image transformer (DeIT) model with a modified head, resulting in improved accuracy in brain lesion detection. The technique also amalgamates the LabelSmoothingCrossEntropy criterion, the AdamW optimizer, the ReduceLROnPlateau scheduler, and EarlyStopping to enhance model performance. However, we have obtained a test accuracy of 97.42% and along with it we have demonstrated the effectiveness of the model using different evaluation metrics. These findings highlight the promise of advanced machine learning approaches for addressing issues in medical field, paving the door for comparatively reliable and efficient detection systems. To sum up, an improved brain tumour detection is a big step in the direction of safer and more effective medical technology. Moreover, Further research can enhance detection and scalability for real-world deployment, ultimately leading to advancements in medical technology.

References

[1] Masoumeh Siar and Mohammad Teshnehlab. Brain tumor detection using deep neural network and machine learning algorithm. In *2019 9th international conference on computer and knowledge engineering (ICCKE)*, pages 363–368. IEEE, 2019.

[2] Sunanda Das, OFM Riaz Rahman Aranya, and Nishat Nayla Labiba. Brain tumor classification using convolutional neural network. In *2019 1st international conference on advances in science, engineering and robotics technology (ICASERT)*, pages 1–5. IEEE, 2019.

[3] Santosh Kumar Sharma, Debendra Muduli, Rojalina Priyadarshini, Rakesh Ranjan Kumar, Abhinav Kumar, and Jitesh Pradhan. An evolutionary supply chain management service model based on deep learning features for automated glaucoma detection using fundus images. *Engineering Applications of Artificial Intelligence*, 128:107449, 2024.

[4] N Arunkumar, Mazin Abed Mohammed, Salama A Mostafa, Dheyaa Ahmed Ibrahim, Joel JPC Rodrigues, and Victor Hugo C De Albuquerque. Fully automatic model-based segmentation and classification approach for mri brain tumor using artificial neural networks. *Concurrency and Computation: Practice and Experience*, 32(1):e4962, 2020.

[5] Navid Ghassemi, Afshin Shoeibi, and Modjtaba Rouhani. Deep neural network with generative adversarial networks pre-training for brain tumor classification based on mr images. *Biomedical Signal Processing and Control*, 57:101678, 2020.

[6] Gopal S Tandel, Ashish Tiwari, and Omprakash G Kakde. Performance optimisation of deep learning models using majority voting algorithm for brain tumour classification. *Computers in Biology and Medicine*, 135:104564, 2021.

[7] Priyansh Saxena, Akshat Maheshwari, and Saumil Maheshwari. Predictive modeling of brain tumor: a deep learning approach. In *Innovations in Computational Intelligence and Computer Vision: Proceedings of ICICV 2020*, pages 275–285. Springer, 2020.

[8] Mangena Venu Madhavan, Aditya Khamparia, and Sagar Dhanraj Pande. An augmented customized deep learning approach for brain tumour identification. *The Imaging Science Journal*, 71(4):331–342, 2023.

[9] Ahmet Çinar and Muhammed Yildirim. Detection of tumors on brain mri images using the hybrid convolutional neural network architecture. *Medical hypotheses*, 139:109684, 2020.

[10] El-Sayed A El-Dahshan, Heba M Mohsen, Kenneth Revett, and Abdel-Badeeh M Salem. Computer-aided diagnosis of human brain tumor through mri: A survey and a new algorithm. *Expert systems with Applications*, 41(11):5526–5545, 2014.

[11] Quan Bin Zhang, Xiao Yan Ji, Qiang Huang, Jun Dong, Yu De Zhu, and Qing Lan. Differentiation profile of brain tumor stem cells: a comparative study with neural stem cells. *Cell research*, 16(12):909–915, 2006.

[12] Suchismita Das, Gopal Krishna Nayak, Luca Saba, Mannudeep Kalra, Jasjit S Suri, and Sanjay Saxena. An artificial intelligence framework and its bias for brain tumor segmentation: A narrative review. *Computers in biology and medicine*, 143:105273, 2022.

[13] Parnian Afshar, Arash Mohammadi, and Konstantinos N Plataniotis. Bayescap: A bayesian approach to brain tumor classification using capsule networks. *IEEE Signal Processing Letters*, 27:2024–2028, 2020.

[14] Cheng Chen, Huilin Wang, Yunqing Chen, Zihan Yin, Xinye Yang, Huansheng Ning, Qian Zhang, Weiguang Li, Ruoxiu Xiao, and Jizong Zhao. Understanding the brain with attention: a survey of transformers in brain sciences. *Brain-X*, 1(3):e29, 2023.

[15] Msoud Nickparvar. Brain tumor mri dataset. 2021. available at. https://www.kaggle.com/dsv/2645886.

[16] Said Charfi, Redouan Lahmyed, and Lalitha Rangarajan. A novel approach for brain tumor detection using neural network. *International Journal of Research in Engineering and Technology*, 2(7):93–104, 2014.

[17] Elvan C Citak, Fatih Sagcan, Begumhan D Gundugan, Sevcan T Bozdogan, Eda B Yilmaz, Emel Avci, Yuksel Balci, and Yasemin Y Karabulut. Metachronous wilms tumor, glioblastoma, and t-cell leukemia in an child with constitutional mismatch repair deficiency syndrome due to novel mutation in msh6 (c. 2590g¿ t). *Journal of pediatric hematology/oncology*, 43(2):e198–e202, 2021.

[18] N Vani, A Sowmya, and N Jayamma. Brain tumor classification using support vector machine. *International Research Journal of Engineering and Technology (IRJET)*, 4(7):792–796, 2017.

[19] Kondepudi Venkata Durga, Debendra Muduli, K Rahul, Amballa Vijay Sai Charan Naidu, Majji Jayanth Kumar, and Santosh Kumar Sharma. Automated diagnosis of brain tumor based on deep learning feature fusion using mri images. In *2023 IEEE 3rd International Conference on Applied Electromagnetics, Signal Processing, & Communication (AESPC)*, pages 1–6. IEEE, 2023.

[20] Debendra Muduli, Ratnakar Dash, and Banshidhar Majhi. Automated breast cancer detection in digital mammograms: A moth flame optimization based elm approach. *Biomedical Signal Processing and Control*, 59:101912, 2020.