Comparative Study of Deep Learning, Transfer Learning, and Transformer-Based Models for Brain Tumor Classification Using MRI Images

PRASHANTHA J[1], Dr. SRIDEVI MALIPATIL[2],

PG SCHOLAR[1], Associate Professor[2] Department of Computer Science and Engineering,

Rao Bahadur Y Mahabaleswarappa Engineering College, Ballari, Karnataka, India

jprashanth58181@email.com

Abstract— Brain tumor detection by Magnetic Resonance Imaging (MRI) is a significant topic in the field of medical image analysis due to its importance in the improvement of treatment methods and patient survival rates. This paper introduces a com- parative study between the use of deep learning, transfer learn- ing, transformers, and hybrid transformers for multiclass brain tu- mor classification. Four classes of MRI brain tumors have been taken into considera-tion, namely glioma, meningioma, pituitary tumors, and normal MRI brain images. Numerous deep learning models including CNN, VGG19, ResNet50, DenseNet121, Ef- ficientNetB4, ViT, DeiT, BEiT, Swin Transformer, SegForm- er, HViDT, Trans-EffNet, and U-Net Trans-former have been implemented in identical experiments. According to experimental results, transformer models significantly outperform convolu- tional neural networks. Among all mod-els, Vision Transformer (ViT) gives highest accuracy rate on testing data sets with 98.76nism in extracting global contextual information in MRI ima

Index Terms— Brain Tumor Detection, MRI, Deep Learning, Vision Transformer, Medical Image Analysis, Transfer Learn- ing, Hybrid Transfor

I. INTRODUCTION

Brain tumors are amongst the most critical neurological diseases that require early detection and diagnosis for effective treatment. Images from MRIs help to visualize brain tissue in a non-radiation way. However, segmentation and interpretation of MRIs take much time and require professional skills of radiologists.

With the help of recent developments in AI and DL techniques automatic classification of brain tumors became possible. At first, CNN models provided good results in analysis of medical images. The last developments include transformer models since they can consider longterm dependencies between data.

In this research, a comparative study is performed to con- sider the performance of deep learn-ing, transfer learning, transformers, and hybrid transformers in multiclass brain tumor classifi-cation

II. RELATED WORK

Many deep learning models have been applied in various brain tumor classification tasks. The performance of CNN models is quite good but it is not easy to analyze complicated tumors. Models that apply transfer learning with pretrained models like VGG19 and ResNet50 provide better performance.

Transformer-based models like ViT, DeiT, Swin Transformer, and SegFormer showed better results than deep learning models due to their global attention. Hybrid transformers provided better results because of integration of local and global features.

III. LITERATURE REVIEW

The classification of brain tumors from MRIs has become one of the most critical uses of Ar-tificial Intelligence (AI) technology in medicine today. Accurate identification of brain tumors is essential as the earlier the diagnosis, the greater the impact on treatment planning and pa-tients' survival rates. Recent advancements in learning-based methods and transfer learning approaches have significantly enhanced automated medical image analysis.

Traditional Machine Learning Approaches

The initial brain tumor classifiers were created using traditional machine learning algorithms in combination with manual feature extraction approaches. Commonly used features included texture measures, Histograms of Oriented Gradients (HOG), wavelet transformations, edge detections, and texture analysis based on statistics.

Despite obtaining satisfactory accuracy levels, the above men- tioned approaches had several shortcomings, including:

- Need for manual creation of features
- Lack of generalization ability
- Image noise sensitivity
- Complex tumor structures processing difficulties
- Lack of scalability

This resulted in researchers opting for automated deep learning-based feature extraction ap-proaches.

Deep Learning and CNN-Based Approaches

CNN has transformed the field of medical image analysis by learning the hierarchical structures of images directly from their pixel values. CNNs proved effective in applications like

CNN architectures demonstrated strong performance in:

- Tumor detection
- Tumor segmentation
- MRI classification
- Disease prediction
- Image enhancement

The success of CNNs inspired researchers to employ different architectures for brain tumor classification tasks.

VGGNet

Simonyan and Zisserman [3]. The VGGNet used very deep convolutions with small-sized receptive fields. VGG19 has become widely adopted by researchers because of its strong per-formance simplicity and strong feature extraction capabilities.

However, VGG19 suffers from:

- Computational complexities
- Number of parameters
- Increased memory usage
- Slower training time

ResNet

Residual Networks (ResNet) were proposed by He et al. [2], and they resolved the vanish-ing gradient problem using residual connections. With ResNet architectures, deep learning networks could be trained more effectively.

ResNet-based approaches improved:

- Improved gradient propagation
- Enhanced feature reuse
- Stable convergence
- Improved classification performan

Even after solving such problems, CNNs cannot model long range spatial relationships that exist in MRI images.

DenseNet

In DenseNet architecture, each layer is connected to all the upcoming layers. It showed good utilization of parameters and less redundancy in feature learning.

Useful aspects of the DenseNet approach include:

- Reusing features
- Less parameter redundancy
- Efficient gradients
- Generalization

Transformer-Based Architectures

The transformer models were first developed for natural language processing (NLP) purposes. Later on, the transformer models were applied in computer vision as well due to their capability to learn relationships in data with the use of self-attention.

Unlike convolutional neural networks, the transformers have the ability to understand the con-text of the whole image and are hence suitable for medical image analysis.

Vision Transformer (ViT)

Dosovitskiy et al. [1] proposed Vision Transformer (ViT), which divides the image into patches and employs transformers for processing the image patches.

The advantages offered by the ViT architecture include:

- Global information extraction
- Dependency modeling between distant patches
- Excellent generalization ability
- Contextual understanding

Recent literature proved that transformers outperformed CNN-based architectures in the clas-sification of MRI images

Data-Efficient Image Transformer (DeiT)

Touvron et al. [6] proposed DeiT, which uses the knowledge distill- ation approach for improv-ing transformer training efficiency.

The primary strengths of DeiT include:

- Small dataset training
- Fast convergence
- Improved computational efficiency
- Enhanced stability during training

BEiT

Bao et al. [7] introduced BEiT, which employed self-supervision in transformer architecture for representation learning from images.

The following tasks demonstrated the strengths of BEiT:

- Classification
- Medical imaging
- Visual representation learning
- Transfer learning

Swin Transformer

Swin Transformer used shifted windows of attention mechanism for hierarchical feature repre-sentation.

Strengths of Swin Transformer include:

- Computational efficiency
- Multi-resolution representation learning

- Hierarchical representation
- Scalability

Segmentation and Hybrid Transformer Models

Recent research focused on developing segmentation-oriented and hybrid architecture trans-formers for analyzing medical images.

SegFormer

SegFormer integrates light-weight transformers with efficient contextual learning mechanisms. SegFormer exhibited excellent results for the tasks of medical image segmentation and classi-fication.

SegFormer Offers:

- Efficient feature fusion
- Light-weight architecture
- Strong context learning capability
- Improved ability to segment images

Hybrid Transformer Architectures

Hybrid transformer architecture involves the combination of CNN based feature extraction techniques with the transformer-based global context learning approach.

Some examples of hybrid transformers are:

- HViDT
- Trans-EffNet
- U-Net Transformer

The key benefits of using hybrid transformers include:

- Improved fusion of local global features
- Superior spatial understanding capability
- Improved tumor localization
- Strong generalization ability

These models achieved promising results in complex medical imaging tasks.

Research Gap

Many previous studies have presented individual architectures for brain tumor classification, limited research has performed a comprehensive comparative analysis across:

- Deep Learning models
- Transfer Learning models
- Pure Transformer architectures
- Hybrid Transformer architectures

More over, many earlier studies primarily concentrate on only on classification accuracy while overlooking several important evaluation factors, including:

- Training stability
- Loss convergence
- Generalization capability
- Class-wise performance
- Computational efficiency

This study aims to overcome these limitations by conducting a detailed comparative study of multiple architectures under identical experimental settings. The paper aims at analyzing the performance of the models as well as their training processes, convergence characteristics, and generalization ability to identify the most effective architecture for multiclass brain tumor classification from MRI images.

Motivation Behind the Proposed Study

The motivation behind The main goal of this study is to find out the best way architecture for multiclass brain tumor classification using MRI images.

This study aims to:

- Compare CNN and transformer architectures
- Analyze classification performance
- Evaluate training convergence
- Study loss behavior
- Investigate generalization capability
- Identify the best-performing architecture

The findings of this study contribute to the development of accurate computer-aided diagnosis systems for clinical applications.

IV. DATASET DESCRIPTION

The data used in this research contains 12,064 pre-processed T1-weighted contrast-enhanced MRI brain images collected for multi-class brain tumor classification. The images are grouped into four different categories.:

Table 1: Dataset Classes

Class	Description
Glioma	Malignant brain tumor
Meningioma	Tumor arising from the meninges
Pituitary	Pituitary gland tumor
No Tumor	Healthy brain MRI images

The dataset is separated into training and testing subsets using an 80:20 ratio.All MRI images were adjusted to a resolution 224×224 pixels before training to ensure uniformity in model input dimensions.

Dataset Composition

Table 2: Training and Testing Distribution

Class	Training Images	Testing Images
Glioma	3018	755
Pituitary	2504	546
Meningioma	2183	626
No Tumor	1945	487
Total	**9650**	**2414**

Dataset Structure

```
/Brain_Tumor_MRI_Dataset/ train/
    Glioma/ Meningioma/
    Pituitary/ No_Tumor/
 test/
    Glioma/ Meningioma/
    Pituitary/ No_Tumor/
```

Image Details

- Modality: T1-weighted contrast-enhanced MRI
- Image Format: JPEG/PNG
- Image Type: Grayscale or RGB
- Pre-processing: Resized and organized into labeled folders

Applications

The dataset is suitable for:

- Brain tumor classification
- Deep learning and transfer learning research
- Computer-aided diagnosis (CAD)
- Medical image analysis and radiology research

Dataset Source

The data set used in this paper can be found at:

- **Dataset Title:** Brain Tumor MRI Dataset
- **Source Link:** https://data.mendeley.com/datasets/zwr4ntf94j/1

The dataset contains 12,064 preprocessed T1-weighted contrast enhanced brain MRIs classified into four groups.

The dataset is provided under the terms of Creative Commons Attribution 4.0 International (CC BY 4.0).

V. METHODOLOGY

This section discusses preprocessing methods used, experimental configuration, model training approach, and evaluation criteria adopted for multiclass brain tumor classification based on MRI images.

Preprocessing

Images taken by MRI devices from diverse resources usually have variation in terms of inten-sity, orientation, contrast, and noise. Thus, it becomes very important to. preprocess images to improve their quality and ensure consistency in results obtained from the models.

The following preprocessing approaches were performed prior to training the models:

- **Image Resizing:** All the MRI images were adjusted to a resolution of 224 × 224 pixels 224×224 pixels to ensure that all models received uniform input sizes.
- **Normalization:** The pixel values were normalized to a range of $[0, 1]$.
- **Data Augmentation:** Methods were employed to make the model more generalized as well as increase data set diversity.
- **Horizontal Flipping:** The images were horizontally flipped randomly to ensure that the model would recognize the patterns irrespective of their direction.
- **Rotation Augmentation:** The images were randomly rotated to deal with different ori-entations present in the MRI scans.
- **Zoom Augmentation:** Different zooms were applied on the images to enable the model to recognize the tumors at different zoom levels.
- **Tensor Transformation:** The images were transformed to tensor form prior to being fed into the neural network models.

The preprocessing pipeline improved feature consistency, enhanced robustness, and reduced the possibility of overfitting during training.

Dataset Splitting

the dataset was organized into three sets:

- **Training Set:** Used for training purposes by the model and optimal tuning of the model parameters.
- **Validation Set:** Used for tuning the hyperparameters and assessing the performance of the model

during training.

- **Testing Set:** Used for measuring the final performance of the model Four

different types of MRI images were used in the experiment:

- Glioma
- Meningioma
- Pituitary Tumor
- No Tumor

Experimental Settings

Regarding the models' comparison, similar experimental setups were employed in training all the proposed neural networks

Table 3: Training Parameters

Parameter	Value
Image Size	224 × 224
Batch Size	32
Optimizer	Adam
Epochs	30
Learning Rate	0.0001
Loss Function	Cross Entropy Loss
Device	CUDA GPU
Framework	PyTorch

In turn, the Adam optimization algorithm was applied because of its flexibility in learning and fast convergence.As for the loss function, Cross Entropy Loss was chosen since the problem is multi-class classification of brain tumor MRI images.

Evaluated Architectures

The architectures tested in this study were classified into the following four categories:

Deep Learning Networks

- CNN

Transfer Learning Networks

- VGG19
- ResNet50
- DenseNet121
- EfficientNetB4

Transformer-based Networks

- ViT
- DeiT
- BEiT
- Swin Transformer
- SegFormer

Hybrid Transformer Networks

- HViDT
- Trans-EffNet
- U-Net Transformer

Training Process

All models tested were trained for 30 epochs on the training dataset. For each training cycle:

- Predictions were calculated by performing forward propagation.
- Cross-Entropy Loss was calculated based on predictions and actual labels.
- Parameters were optimized through backpropagation
- Accuracy and loss on validation data were evaluated after each epoch.
- The optimal weights that gave the highest accuracy were saved.

The training process leveraged the computational capabilities of CUDA GPUs.

VI. PERFORMANCE METRICS

the results obtained from each model was evaluated using multiple classification metrics to ensure comprehensive analysis.

The following metrics were used:

- Accuracy
- Precision
- Recall
- F1-Score
- Training Loss
- Validation Loss
- Confusion Matrix

Accuracy

Accuracy measures the overall percentage of correctly classified samples.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

where:

- TP = True Positive
- TN = True Negative
- FP = False Positive
- FN = False Negative

Precision

Precision refers to the percentage of positive samples which are correctly detected by the model.

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

Higher precision means that fewer positive errors have been made by the model.

Recall

Recall represents how effectively the model identifies positive samples.

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

Higher recall indicates fewer false negatives.

F1-Score

The F1-score combines precision and recall into a single performance measure using their harmonic mean.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{4}$$

F1-Score provides balanced evaluation when class distribution is uneven.

Confusion Matrix

A confusion matrix was used to analyze classwise prediction performance and identify mis-classification patterns among tumor categories.

The confusion matrix provides a clearer understanding of:

- Correct classifications
- False positives
- False negatives
- Class-specific confusion

Training and Validation Loss

Training and validation loss values were monitored during the training process to determine if the model converges or if there is any overfitting problem.

Lower loss values generally reflect better optimization and improved generalization capability.

VII. EXPERIMENTAL RESULTS

Figure 1 presents the comparative performance analysis of all tested models based on the following evaluation metrics:

- Training Accuracy

Table 4: Performance Comparison of All Models

Model	Category	Training Accuracy	Training Loss	Validation Accuracy	Validation Loss	Test Accuracy
ViT	Transformer	0.9994	0.3502	0.9876	0.3503	**0.9876**
HViDT	Hybrid Transformer	0.9994	0.3513	0.9872	0.3511	0.9872
Trans-EffNet	Hybrid Transformer	0.9990	0.3517	0.9872	0.3520	0.9872
SegFormer	Transformer	0.9988	0.3527	0.9867	0.3527	0.9867
BEiT	Transformer	0.9991	0.3504	0.9863	0.3504	0.9863
Swin Transformer	Transformer	0.9990	0.3524	0.9863	0.3524	0.9863
U-Net Transformer	Hybrid Transformer	0.9979	0.0063	0.9847	0.0063	0.9847
DeiT	Transformer	0.9992	0.3511	0.9793	0.3511	0.9793
DenseNet121	Transfer Learning	0.9631	0.1043	0.9631	0.1670	0.9631
VGG19	Transfer Learning	0.7601	0.7499	0.7601	0.7499	0.7601
CNN	Deep Learning	0.7229	0.7035	0.7229	0.7035	0.7229
ResNet50	Transfer Learning	0.5468	0.5173	0.5468	0.5173	0.5468
EfficientNetB4	Transfer Learning	0.4582	1.2624	0.4582	1.2448	0.4582

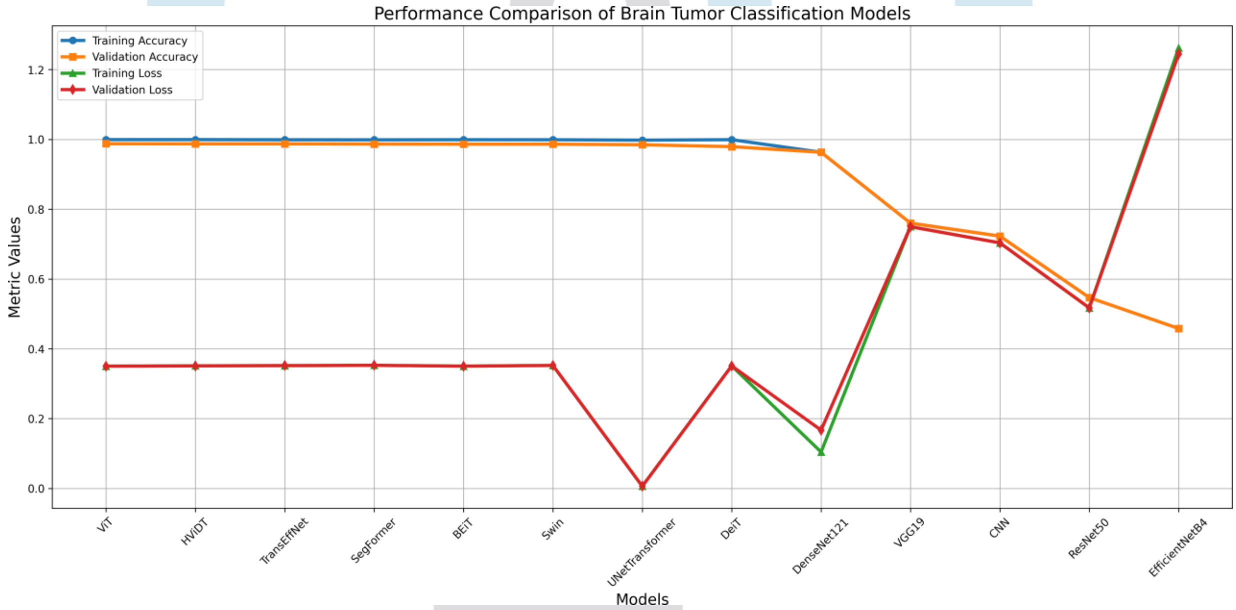

Figure 1: Comparison of Training Accuracy, Training Loss, Validation Accuracy, and Valida-tion Loss Across All Models

- Training Loss
- Validation Accuracy
- Validation Loss

The graph clearly demonstrates that transformer-based architectures and hybrid transformer models achieved significantly higher classification resulting in higher accuracies and lower loss values than conventional CNN and transfer learning methods.

Among all evaluated architectures, the Vision Transformer (ViT) achieved the best validation and testing accuracy, indicating strong generalization capability for multiclass brain tumor clas-sification using MRI images.

Hybrid transformer architectures such as HViDT and Trans-EffNet also exhibited highly stable convergence behavior and balanced overall performance during training. In contrast, classical CNN architectures showed comparatively lower accuracy and higher loss values, indicating limited capability in learning complex MRI tumor patterns.

The experimental findings suggest that transformer-based self-attention mechanisms and over-all contextual understanding feature learning substantially improve classification performance for brain tumor MRI analysis.

VIII. DISCUSSION

The obtained results demonstrated that transformer based architectures outperform traditional CNN and transfer learning based approaches in the task of brain tumor classification using MRI images.

The following important observations were derived from the experimental results:

- Of all models, Vision Transformer (ViT) performed better, with an accuracy rate of98.76%.
- Hybrid transformer architectures demonstrated balanced and stable performance across training and validation stages.
- Among all the categories, glioma tumor classification turned out to be the hardest one due to its complexity and variability.
- Pituitary tumors were classified with near-perfect accuracy by transformer-based models.
- Transformer self attention mechanisms contributed significantly to improved contextual feature extraction from MRI images.

Overall, the results indicate that transformer-based learning provides superior capability in cap-turing global dependencies and complex spatial patterns in brain MRI images, leading to im-proved multiclass tumor classification performance.

IX. FUTURE SCOPE

In fact, the suggested method of brain tumor classification has demonstrated excellent perfor-mance while automating MRI image analysis using deep learning and transformers. Although the brain tumor classification models designed here are highly accurate, there is a variety of ways to improve this methodology further.

Development in the areas of AI, imaging technologies, etc. may lead to creating even more effective brain tumor classification solutions.

Use of Advanced Deep Learning Architectures

Other models that can be studied in the future include such architectures as:

- ConvNeXt
- EfficientFormer
- MaxViT
- MobileViT
- Masked Autoencoders
- Vision-Language Models

Such approaches may bring better feature extraction, computationally efficient algorithms, and generalizability of the proposed solution. Ensemble learning methodologies may be considered as well.

Training with Larger and Multi-Institutional Datasets

The present research employs MRI scans gathered from relatively few sources. Future research could involve collecting larger and more diverse datasets from different hospitals and medical institutions.

Large and diverse datasets would be beneficial in improving:

- Generalizability of the model
- Model robustness
- Medical reliability of results
- Dataset diversity
- Immunity to dataset bias

Using diverse datasets may also help in adapting models to different types of MRI equipment, acquisition protocols, image resolutions, and varying clinical environments.

Incorporation of 3D MRI Volumes

The research is mainly based on 2D MRI slice classification. But brain tumors are volumetric data with valuable spatial information in multiple slices.

Possible approaches in future may include:

- 3D Convolutional Neural Networks (3D-CNNs)
- 3D Vision Transformers
- Volumetric Attention Networks

With this technique, classification can learn the spatial relationships and thus perform better on tasks.

Automatic Tumor Segmentation

Future research could include tumor segmentation prior to classification. Tumor segmentation algorithms can be used for:

- Localizing tumors automatically
- Eliminating extraneous background
- Achieving higher classification accuracies
- Improving model explainability

Well-known segmentation models, such as:

- U-Net
- Attention U-Net
- SegFormer
- TransUNet

can be included in the proposed scheme.

Explainable Artificial Intelligence (XAI)

Deep learning algorithms tend to be treated as black box models, thus limiting clinical con-fidence in them. Potential future research might include the use of explainable AI (XAI), approaches like:

- Grad-CAM
- SHAP
- LIME
- Attention Heatmaps
- Saliency Maps

The above will aid in visualizing the tumors that affect predictions made by the algorithms.

Real-Time Clinical Decision Support Systems

This approach may be used in a real time clinical decision support system for hospitals and diagnostics labs.

Future systems may include:

- Web-based MRI analysis platforms
- Cloud-based diagnosis systems
- AI-assisted radiology software
- Automated report generation

Such systems may reduce diagnosis time and improve early tumor detection.

Transfer Learning and Self-Supervised Learning

Medial data sets usually have limitations since annotating them needs expertise from a radiol-ogist.

For further studies, we can explore:

- Self-supervised learning
- Contrastive learning
- Few-shot learning
- Semi-supervised learning

Such techniques will help train visual representation using unlabeled MRI scans and decrease reliance on manually labeled data sets.

Tumor Grading and Severity Estimation

The current system only takes into account four categories of MRI image classification. In the future work, the system could be improved in order to allow:

- Tumor grade classification
- Tumor stage classification
- Survival rate prediction
- Prognosis analysis
- Evaluation of treatment efficacy

These future systems could provide deeper clinical insights and assist healthcare professionals in treatment planning, disease monitoring, and personalized medical decision making.

Multimodal Medical Data Integration

Potential areas for future research include combining magnetic resonance imaging with addi-tional medical data such as:

- CT scans
- Genetic information
- Medical histories
- Patient records
- Histopathological information

Multimodal learning techniques have proven useful in enhancing diagnostic accuracy and en-abling more comprehensive patient analysis for clinical decision-making.

Federated Learning and Privacy Preservation

The medical data is extremely sensitive and private in nature due to the privacy policies and laws for health care data protection.

Further work can be done on federated learning,which will allow for a collaborative training process by hospitals, without posing a risk of leakage of patient data.

The advantages of federated learning include the following:

- High level of privacy preservation
- Robustness of models
- Collaboration between multiple hospitals
- Distributed secure learning

These techniques can enhance the scalability and reliability of the system for AI based medical diagnosis systems while preserving patient privacy and security of their data.

Edge AI and Mobile Healthcare Applications

Transformers and CNNs that have been made lightweight can be optimized to work on the following platforms:

- Mobile devices
- Embedded systems
- Edge computing devices
- Mobile healthcare devices

It would facilitate remote diagnosis in rural regions.

Robustness Against Noise and Variability

The MRI images generally include the following components:

- Noise
- Motion artifacts
- Intensity variation
- Scanner variation

Methods for robust pre-processing and data augmentation can be explored further.

Cross-Dataset Generalization

Nevertheless, the performance of most clinical AI models is determined using one dataset that may affect their application in practice.

Further research can focus on:

- Cross-dataset validation
- Domain adaptation
- Generalization test
- External tests

These methods can enhance the clinical robustness and readiness for practical implementation of models.

Attention Visualization and Interpretability

Visualization of Attention in Transformers Transformers use self-attention for feature extrac-tion.

Further research can examine methods for attention visualization and find out which regions the model focuses on when it performs tumor segmentation.

This will help:

- Make models more transparent
- Enhance confidence of the model in clinical practice
- Explain the process
- Make models more interpretable

Adversarial Robustness

Deep learning algorithms are sensitive to attacks and noise. The future research can develop robust algorithms resistent to:

- Attacks
- Corrupted images
- Noise of MRIs
- Variability in scanning devices

This is essential for clinical implementation.

Integration with Hospital Information Systems

Future systems can be integrated with:

- Hospital Information Systems (HIS)
- Picture Archiving and Communications Systems (PACS)
- Electronic Health Records (EHR)

Such integration can facilitate automated workflow management and intelligent medical deci-sion support.

Human-AI Collaborative Systems

In futuristic systems, the results produced by AI systems may be combined with the knowledge of radiologists to create hybrid intelligent systems.

The collaboration between human and AI may lead to:

- Error-free diagnosis
- Clinical accuracy
- Credibility
- Expert decision making

Smart Healthcare Ecosystems

It may be potentially implemented in AI driven smart healthcare ecosystems for automated medical imaging analysis.

In general, in the future, smart healthcare systems will be able to deliver:

- Automated diagnostics
- Intelligent patient monitoring
- Cloud-based healthcare services
- On-site medical assistance
- Telemedicine services

All this implies that advances in deep learning, transformers, explainable AI, federated learn-ing, together with innovations in medical imaging technologies, are able to significantly en-hance brain tumors diagnosis system automation.

Integration of intelligent AI based healthcare solutions can provide many benefits for radiolo-gists and patients all around the globe

X. CONCLUSION

This work presented a comparative analysis of deep learning, transfer learning, transformer-based models, and hybrid transformer architectures for classifying multiple brain tumor types classification using MRI images.

The experimental results demonstrated that transformer based architectures outperformed tra-ditional CNN based models in terms of classification accuracy, convergence behavior, and gen-eralization capability. Among all evaluated models, the Vision Transformer (ViT) attained the highest accuracy of 98.76%.

The findings indicate that self-attention mechanisms are highly effective in capturing global contextual information from medical images, leading to improved brain tumor classification performance. The study also emphasizes the potential of transformer based approaches for future AI assisted medical imaging and clinical diagnostic systems.

REFERENCES

[1] A. Dosovitskiy et al., "An Image is Worth 16x16 Words: Transformers for Image Recog-nition at Scale," 2021.

[2] K. He et al., "Deep Residual Learning for Image Recognition," 2016.

[3] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," 2015.

[4] Z. Liu et al., "Swin Transformer: Hierarchical Vision Transformer using Shifted Win-dows," 2021.

[5] E. Xie et al., "SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers," 2021.

[6] H. Touvron et al., "Training Data-efficient Image Transformers and Distillation through Attention," 2021.

[7] H. Bao et al., "BEiT: BERT Pre-Training of Image Transformers," 2021.

[8] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Con-volutional Neural Networks," 2012.

[9] G. Huang et al., "Densely Connected Convolutional Networks," 2017.

[10] M. Tan and Q. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," 2019.

[11] C. Szegedy et al., "Going Deeper with Convolutions," 2015.

[12] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," 2015.

[13] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," 2015.

[14] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," 2015.

[15] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation," 2016.

[16] J. Long, E. Shelhamer, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," 2015.

[17] L. Wang et al., "Deep Learning for Identifying Metastatic Breast Cancer," 2016.

[18] S. Pereira et al., "Brain Tumor Segmentation Using Convolutional Neural Networks in MRI Images," 2016.

[19] M. Havaei et al., "Brain Tumor Segmentation with Deep Neural Networks," 2017.

[20] B. H. Menze et al., "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)," 2015.

[21] A. Esteva et al., "Dermatologist-level Classification of Skin Cancer with Deep Neural Networks," 2017.

[22] G. Litjens et al., "A Survey on Deep Learning in Medical Image Analysis," 2017.

[23] T. N. Kipf and M. Welling, "Semi-Supervised Classification with Graph Convolutional Networks," 2017.

[24] J. Devlin et al., "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," 2019.

[25] A. Vaswani et al., "Attention Is All You Need," 2017.

[26] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a Convolutional Neural Network," 2017.

[27] Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," Nature, vol. 521, no. 7553, pp. 436–444, 2015.

[28] I. Goodfellow, Y. Bengio, and A. Courville, "Deep Learning," MIT Press, 2016.

[29] R. Yamashita et al., "Convolutional Neural Networks: An Overview and Application in Radiology," Insights into Imaging, 2018.

[30] M. T. Islam et al., "Brain Tumor Detection in MRI Images Using Deep Learning Tech-niques," 2020.