STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **16**, December 2026, pp 5290–5299.
Published online in International Academic Press (www.IAPress.org)

TransMedNet: A Hybrid Transformer-CNN Framework for Accurate Brain Tumor Classification from MRI Scans

E. Aravind*, Alle Shantharaju

Chaitanya–Deemed to be University, Himayatnagar, Hyderabad, Telangana 500075, India

Abstract The diagnosis of brain tumors through MRI analysis presents significant challenges due to the heterogeneous nature of tumor morphology and subtle differences across tumor types. To address these complexities, we introduce TransMedNet, a hybrid deep learning model that synergizes convolutional and transformer-based architectures for enhanced brain tumor classification. By combining the spatial precision of CNNs with the global context awareness of Vision Transformers, the model effectively captures both localized abnormalities and long-range dependencies within MRI scans. A dual attention mechanism further refines feature selection, while multiscale processing supports representation of tumor patterns at different spatial scales. Experimental evaluation on the public MRI dataset shows that TransMedNet provides competitive classification performance across the reported accuracy, AUC, and F1-score measures. Because the present study is based on a single public dataset and does not include independent external clinical validation, the results should be interpreted as evidence of technical feasibility within the evaluated experimental setting rather than proof of clinical deployment readiness.

Keywords Brain Tumor Classification, MRI, Vision Transformer (ViT), Hybrid Attention Network, Deep Learning, CNN, Self-Attention, Medical Image Analysis, Automated Diagnosis

AMS 2020 subject classifications 68T07, 68T45, 92C55

DOI: 10.19139/soic-2310-5070-4228

1. Introduction

The brain, a vital and complex organ composed of billions of cells, can be severely affected by abnormal cell growth that forms brain tumors, impairing normal functionality and destroying healthy cells. Tumors are classified into low-grade (benign, grades I–II) and high-grade (malignant, grades III–IV), where benign tumors grow slowly and are non-cancerous, while malignant tumors are aggressive, cancerous, and rapidly expanding [1]. Brain-related cancers are a leading cause of death, with nearly 200 tumor types identified and a reported 300% increase in mortality over the past three decades [2]. Magnetic Resonance Imaging (MRI) is the most widely used modality for tumor diagnosis due to its superior visualization of brain anatomy compared with imaging techniques such as PET, CT, MEG, CEST, SPECT, EEG, and X-ray [3]. However, manual diagnosis is time-consuming, expert-dependent, and prone to inconsistencies caused by contrast variations and observer subjectivity [4]. These challenges motivate the development of automated approaches that can support image-based analysis. Traditional machine learning (ML) depends substantially on handcrafted features [5], whereas Deep Learning (DL) enables automatic hierarchical feature extraction and has been widely investigated for classification and segmentation tasks [6]. DL models can capture spatial representations from MRI scans that are useful for tumor analysis [7]. Supervised DL models, particularly CNNs, generally require annotated datasets, while

*Correspondence to: E. Aravind (Email: aravind@chaitanya.edu.in), Chaitanya–Deemed to be University, Himayatnagar, Hyderabad, Telangana 500075, India.

ISSN 2310-5070 (online) ISSN 2311-004X (print)
Copyright © 2026 International Academic Press

generative and unsupervised approaches such as GANs, VAEs, and diffusion models provide alternative strategies for representation learning and anomaly analysis [8]. Transfer Learning (TL) using pretrained architectures such as ResNet50, VGG16, DenseNet21, and VGG19 has also been explored for brain tumor diagnosis [9]. Vision Transformer (ViT) architectures provide an alternative mechanism for modeling long-range image dependencies through self-attention [10]. Although CNN–Transformer combinations have increasingly been investigated, the contribution of TransMedNet lies in the coordinated integration of local convolutional representations, hierarchical transformer features, channel–spatial attention, shifted-window processing, and multiscale feature fusion within a unified MRI classification framework. The novelty is therefore associated with the organization and information flow among these complementary components rather than with the general concept of combining CNNs and transformers. The principal technical characteristics of TransMedNet are:

1. **Hybrid Architecture:** Combines convolutional layers for extracting fine-grained local features with a Vision Transformer backbone for capturing global context and long-range dependencies in MRI images.
2. **HATNet Encoder:** Integrates spatial feature extraction and hierarchical abstraction to preserve tumor-related structures while reducing irrelevant background information.
3. **Patch Tokenization with Positional Encoding:** Organizes image representations into fixed-size patch tokens and incorporates positional information to preserve spatial relationships during transformer processing.
4. **Multi-Head Self-Attention:** Models complementary relationships among multiple image regions and supports representation of spatially distributed tumor characteristics.
5. **Multiscale Feature Integration:** Combines representations obtained at different processing depths to retain complementary local texture and higher-level semantic information.

The remainder of this paper is structured as follows. Section 2 reviews related work on brain tumor classification using machine learning, convolutional neural networks (CNNs), and vision transformers (ViTs). Section 3 presents the proposed TransMedNet architecture and its principal components. Section 4 describes the experimental setting and discusses the reported comparative results. Section 5 concludes the study and outlines future research directions.

2. Related Work

Shinde et al. [11] applied Logistic Regression, MLP, Decision Tree, Naïve Bayes, and SVM for brain tumor classification, achieving 90% accuracy with Logistic Regression and MLP using DWT for feature extraction. Wang et al. [12] introduced RanMerFormer, a ViT-based CAD method with merging mechanisms, achieving 99% accuracy on DS1 and DS2 MRI datasets. Sharma et al. [13] used HOG with a modified ResNet50 and ensemble classifier, obtaining 88% accuracy. Zhou et al. [14] proposed a re-attention mechanism improving ViT accuracy by 1.6% on ImageNet, while Han et al. [15] and Arnab et al. [16] explored transformer architectures for high-, mid-, and low-level vision and video tasks. Liu et al. [17] developed the Swin Transformer with shifted windows for computationally efficient hierarchical representation learning. Tariq et al. [18] combined EfficientNetV2 with ViT, reporting 96% accuracy. Poornam and Angelina [19] proposed VITALT, reporting approximately 99% accuracy across four datasets. Krishnan et al. [20] introduced RViT and reported 98.6% accuracy on Kaggle. Asiri et al. [21, 22] proposed FT-ViT and evaluated five pretrained ViT architectures, reporting accuracies up to 98.24%. Simon and Briassouli [23] reported 96.5% classification and 98% detection accuracy, while Chauhan et al. [24] developed PBViT with 95.8% accuracy. These published performance values were obtained using different datasets, partitions, preprocessing pipelines, and evaluation protocols. Consequently, they provide contextual evidence regarding developments in brain MRI classification but should not be interpreted as direct head-to-head comparisons with TransMedNet. The present study therefore limits quantitative comparative conclusions to models evaluated within its reported experimental setting. MRI remains a principal modality for brain tumor analysis [25, 26], while CNN architectures continue to provide effective local feature representations [27]. Recent studies published in *Statistics,*

Optimization & Information Computing (SOIC) further demonstrate the increasing use of transformer, attention-based, and interpretable deep learning methods in medical image classification. Aniq et al. [28] investigated Vision Transformers and attention mechanisms for mammographic image classification, illustrating the potential of global attention for medical image representation. Ali and Sadeeq [29] developed an interpretable multi-class deep learning framework employing stratified five-fold cross-validation and explainability analysis, emphasizing robust and transparent evaluation in medical image classification. Hassan and Kako [30] proposed an attention-augmented EfficientNetV2B0 framework for multi-class cardiac disease classification from cine MRI, further illustrating the role of attention mechanisms in MRI-based classification. Together, these studies provide relevant methodological context for hybrid architectures that combine local convolutional representations with attention-based global feature modeling.

3. TransMedNet: Hybrid Attention Transformer Network

TransMedNet is a hybrid deep learning model combining Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs) for brain tumor classification from MRI images. The model combines local convolutional feature extraction with global context modeling through transformer processing [31]. The hierarchical attention design is informed by the shifted-window strategy introduced in the Swin Transformer [17]. As shown in Figure 1(a), the architecture begins with preprocessing of the MRI image, including normalization and skull stripping, followed by feature extraction using the HATNet encoder. The encoder, illustrated in Figure 1(b), combines convolutional operations for spatial feature extraction with attention-based processing for broader contextual representation. The extracted features are tokenized and passed to the ViT backbone (Figure 1(c)), where image representations are organized into patches and processed through multi-head self-attention. Positional encodings are incorporated to retain spatial relationships, and the classification output corresponds to four categories: glioma, meningioma, pituitary tumor, and no tumor. To reduce the loss of local spatial continuity associated with direct patch embedding, TransMedNet incorporates a hybrid embedding layer (Figure 1(d)) that combines convolutional low-level representations with ViT patch tokens.

This fusion enhances representation of subtle tumor patterns. Feature refinement is performed using channel and spatial attention, enabling the network to emphasize tumor-relevant representations. Hierarchical transformer blocks, adapted from the Swin Transformer and shown in Figure 1(e), use window-based self-attention and shifted windows to balance computational efficiency and contextual modeling, while residual connections support information propagation through deeper stages. A Multiscale Feature Fusion Module (MFFM) integrates representations obtained at different transformer depths. Feature maps are aligned and combined through attention-guided weighting so that lower-level texture information and higher-level semantic representations contribute jointly to classification. The overall processing sequence consists of: (i) a CNN stem for low-level intensity, edge, and texture representation; (ii) a hierarchical convolutional encoder for intermediate anatomical and tumor-related features; (iii) a hybrid embedding layer integrating convolutional representations with patch tokens; (iv) a transformer encoder containing six attention blocks with eight attention heads; (v) Swin-inspired windowed and shifted-window attention stages; (vi) channel–spatial attention for feature refinement; (vii) multiscale feature fusion across transformer stages; and (viii) global pooling followed by dense classification layers. Self-supervised approaches such as SimCLR and MoCo represent potential extensions for exploiting unlabeled MRI data; they are not part of the supervised experiments used to obtain the numerical results reported in this study. For the reported supervised configuration, the input resolution is 256×256 pixels and the patch size is 16×16. The embedding dimension is 256, and the transformer encoder contains six layers with eight attention heads. Training uses a batch size of 16 and the AdamW optimizer with a learning rate of 1×10^{-4}. Data augmentation is applied during training, and five-fold cross-validation is used as the reported internal evaluation strategy. All reported experiments use the PyTorch 2.0 framework and an NVIDIA RTX-series GPU with 24 GB of memory. TransMedNet therefore integrates convolutional feature extraction, transformer-based contextual modeling, dual attention, and multiscale fusion within a unified classification framework.

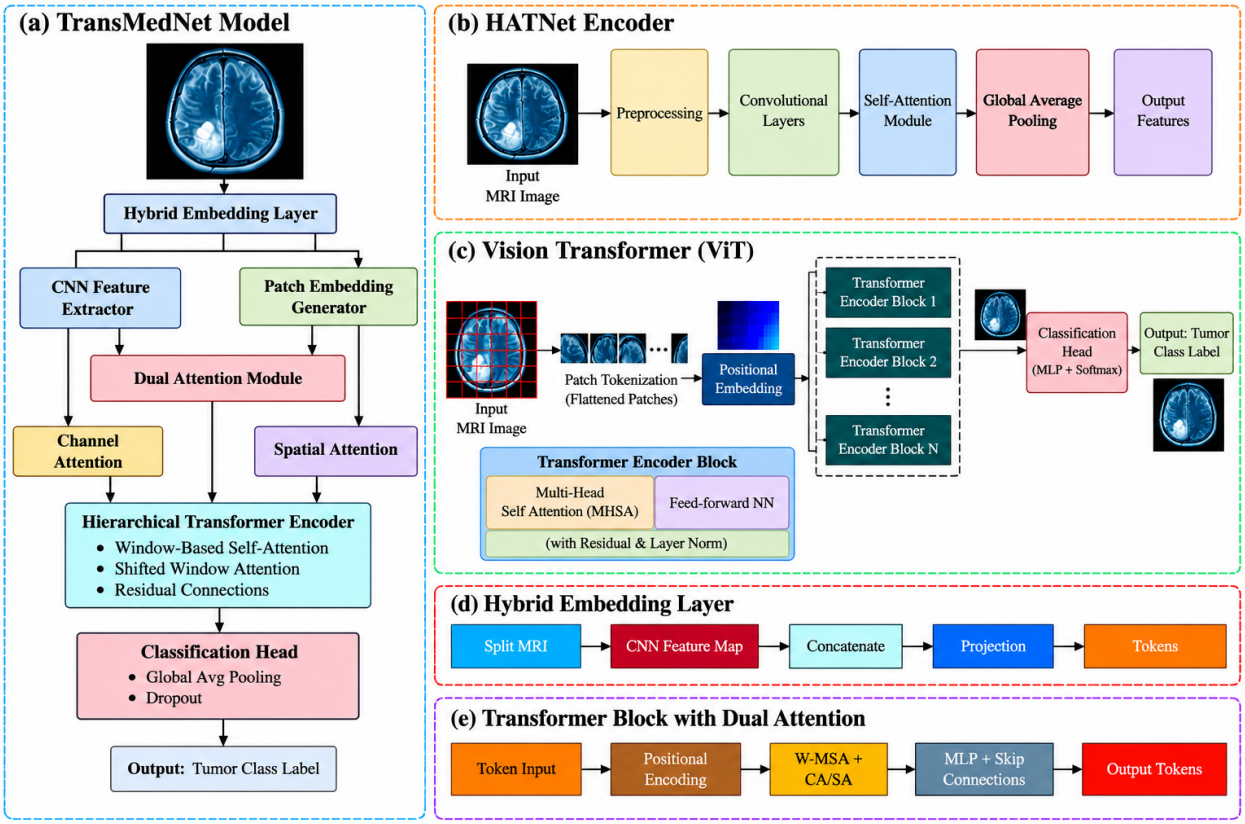


Figure 1. TransMedNet architecture: (a) TransMedNet model, (b) HATNet encoder, (c) Vision Transformer (ViT), (d) hybrid embedding layer, and (e) transformer block with dual attention.

4. Results and Discussion

4.1. Dataset and Experimental Setting

The experiments use a publicly available MRI brain tumor dataset from Kaggle [32], containing 3,264 images distributed among four categories: glioma, meningioma, pituitary tumor, and no tumor. The reported dataset partition contains 2,870 training images and 394 testing images. Images are resized to 256×256 pixels before model input. Labels represent four nominal tumor categories and are numerically encoded only for computational processing; the numerical codes do not imply an ordinal relationship among classes. Data augmentation is applied to the training data. Five-fold cross-validation is used for internal model assessment, while the reported 394-image test partition is retained as the stated test set. The comparative models are evaluated using the same reported preprocessing and training setting to support consistent within-study comparison. The available dataset description does not provide sufficient subject identifiers to independently establish whether all images are subject-disjoint. Consequently, the possibility of multiple images originating from the same subject cannot be excluded solely from the reported image counts and is considered a limitation of the present evaluation. All experiments are implemented in PyTorch 2.0 and executed using the NVIDIA RTX-series GPU configuration described in Section 3.

Table 1. Overall metrics of all models

Model	Accuracy	Precision	Recall	F1 score	Kappa
Inception-ResNet v2 [33]	29.3	8.6	29.3	1.0	0
LeNet5 [34]	74	29	27.1	26.9	5.7
ResNet101-LSTM [35]	75.4	80.1	75.4	73.7	65.8
Inception v3 [36]	76.4	77.5	76.4	94	68
Xception [37]	76.6	76.4	76.7	92.3	68.5
ResNet101-LSTM-GRU [38]	77.4	84.8	76.7	74.6	61.3
ConvNeXtBase [39]	86	87.8	87.1	98.5	82.5
VGG19 [40]	86.8	86	86	95.4	82.2
ResNet101 [41]	88.6	89	88.9	96.1	84.6
TransMedNet	92.5	93.2	93.5	98.6	90.5

As summarized in Table 1, TransMedNet reports an accuracy of 92.5%, precision of 93.2%, recall of 93.5%, F1-score of 98.6%, and kappa value of 90.5 within the reported experimental results. These values are interpreted as within-study results rather than evidence of universal superiority over methods evaluated using different datasets or protocols. Macro-, micro-, and weighted-average measures are additionally reported to characterize performance from complementary aggregation perspectives. Because these measures use different averaging procedures, their values should be interpreted according to the corresponding aggregation definition.

Table 2. Macro-average performance

Model	Precision	Recall	F1 score	AUC
Inception-ResNet v2 [33]	25	11	–	50
LeNet5 [34]	24	24	22	52
ResNet101-LSTM [35]	80	75	74	82.2
Inception v3 [36]	78	76	76	84
Xception [37]	77	77	77	84.6
ResNet101-LSTM-GRU [38]	85	77	75	92.4
ConvNeXtBase [39]	89	88	88	91.7
VGG19 [40]	87	87	87	91.4
ResNet101 [41]	89	89	89	92
TransMedNet	93.21	94.89	93.58	95.15

Table 3. Micro-average performance

Model	Precision	Recall	F1 score	AUC
Inception-ResNet v2 [33]	29	29	29	50
LeNet5 [34]	29.5	29.5	29.5	51
ResNet101-LSTM [35]	58	58	58	82.7
Inception v3 [36]	76.4	76.4	76.4	93
Xception [37]	76.7	76.7	76.7	93.1
ResNet101-LSTM-GRU [38]	77.4	77.4	77.4	89.9
ConvNeXtBase [39]	87.1	87.1	87.1	97.1
VGG19 [40]	86.8	86.8	86.8	93.9
ResNet101 [41]	88.6	88.6	88.6	97.3
TransMedNet	92.8	92.5	92.6	98.7

Table 4. Weighted-average performance

Model	Precision	Recall	F1 score	AUC
Inception-ResNet v2 [33]	9	29	13	50
LeNet5 [34]	24	25	23	52
ResNet101-LSTM [35]	80	75	73	81.3
Inception v3 [36]	78	76	76	83.9
Xception [37]	76	77	76	84
ResNet101-LSTM-GRU [38]	83	77	74	92.1
ConvNeXtBase [39]	88	87	87	91.1
VGG19 [40]	87	87	87	91.1
ResNet101 [41]	89	89	89	92.3
TransMedNet	93.28	92.25	92.86	94.89

4.2. Discussion

Tables 1–4 summarize the reported overall, macro-, micro-, and weighted-average results. The different averaging strategies characterize performance from different perspectives and should therefore be interpreted together with the underlying class distribution and evaluation protocol. For TransMedNet, the macro-average results are 93.21% precision, 94.89% recall, 93.58% F1-score, and 95.15% AUC. The micro-average results are 92.8% precision, 92.5% recall, 92.6% F1-score, and 98.7% AUC, while the weighted-average values are 93.28% precision, 92.25% recall, 92.86% F1-score, and 94.89% AUC. These aggregation-specific values are reported separately rather than being treated as interchangeable measurements.

The comparative results characterize TransMedNet as a competitive architecture relative to the CNN and CNN–RNN baselines included in the reported experiments. The architectural rationale is that convolutional processing preserves localized texture and structural information, whereas transformer self-attention provides a mechanism for modeling longer-range spatial dependencies. Dual attention and multiscale fusion provide additional mechanisms for selectively integrating representations obtained at different feature levels. Because the study does not contain independent external validation, reader assessment, or prospective clinical evaluation, the reported classification results characterize technical performance within the evaluated public MRI dataset and should not be interpreted as evidence of clinical effectiveness.

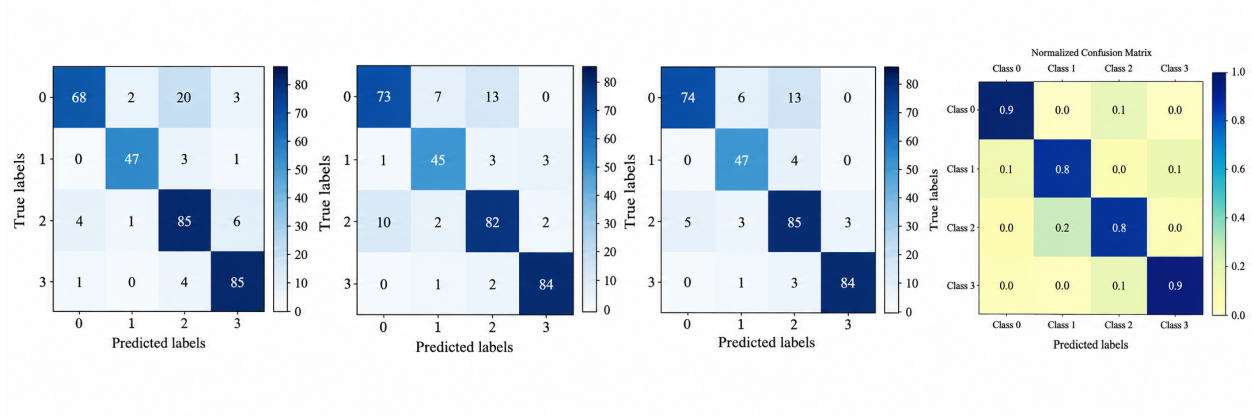


Figure 2. Confusion matrices for (a) ConvNeXtBase, (b) VGG19, (c) ResNet101, and (d) TransMedNet.

The confusion matrices in Figure 2 provide a class-wise representation of correct and incorrect predictions and permit examination of the direction of classification errors. Such class-level analysis is relevant when tumor categories exhibit overlapping MRI appearance.

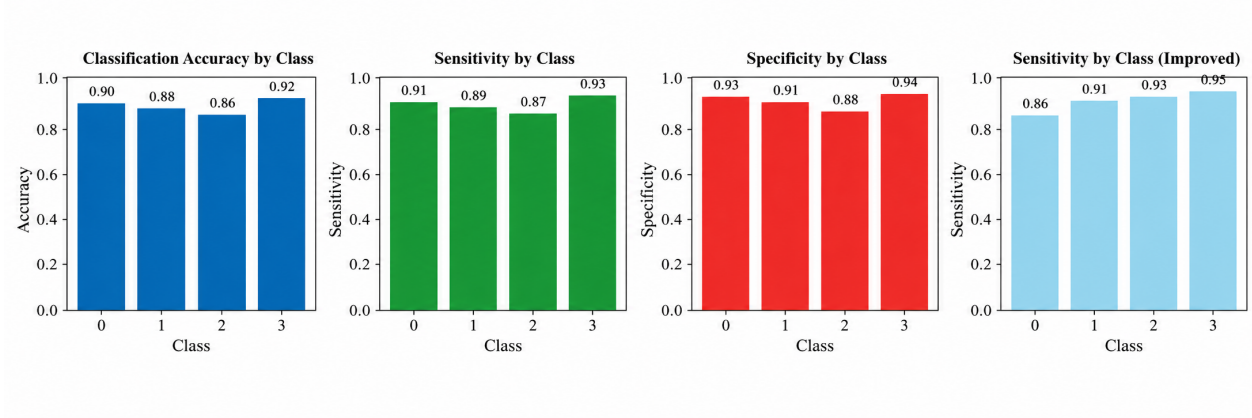


Figure 3. Sensitivity/recall by class for (a) ConvNeXtBase, (b) VGG19, (c) ResNet101, and (d) TransMedNet.

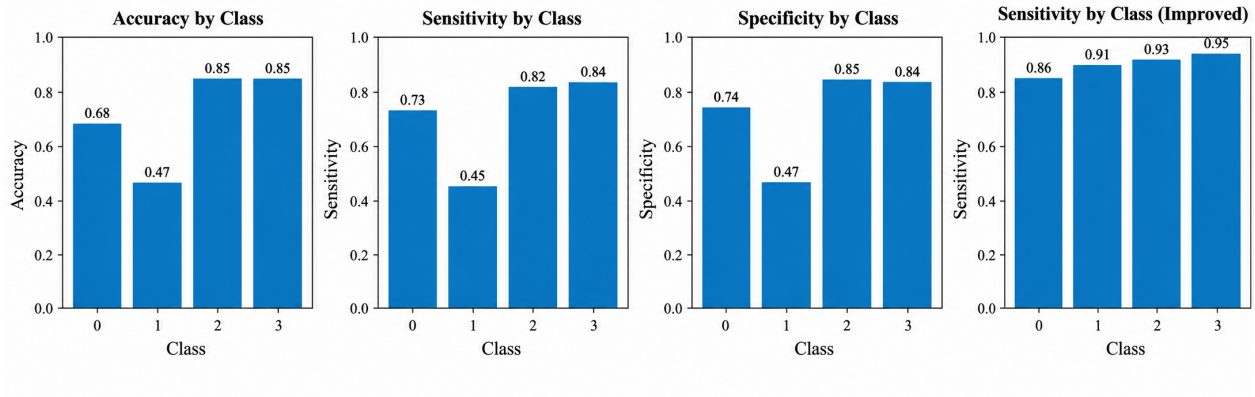


Figure 4. Specificity/precision-oriented class-wise analysis for (a) ConvNeXtBase, (b) VGG19, (c) ResNet101, and (d) TransMedNet.

Figures 3 and 4 provide complementary class-wise views of model behavior and expose variation among tumor categories rather than implying equivalent performance across all classes.

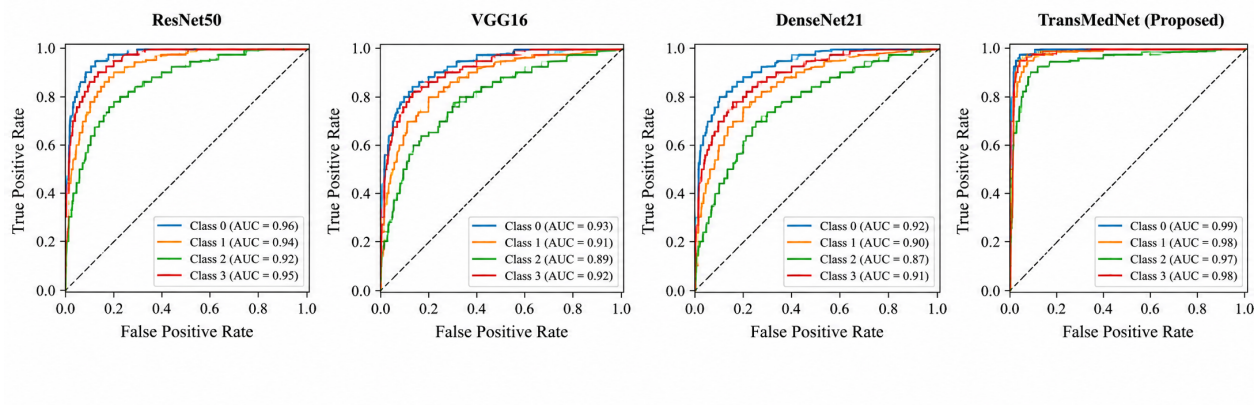


Figure 5. Comparative performance curves for the evaluated models.

Figure 5 provides an additional graphical representation of discrimination performance and should be interpreted according to the axis definitions and AUC formulation used to generate the curves. It complements the tabulated performance measures without being treated as an independent source of numerical estimates.

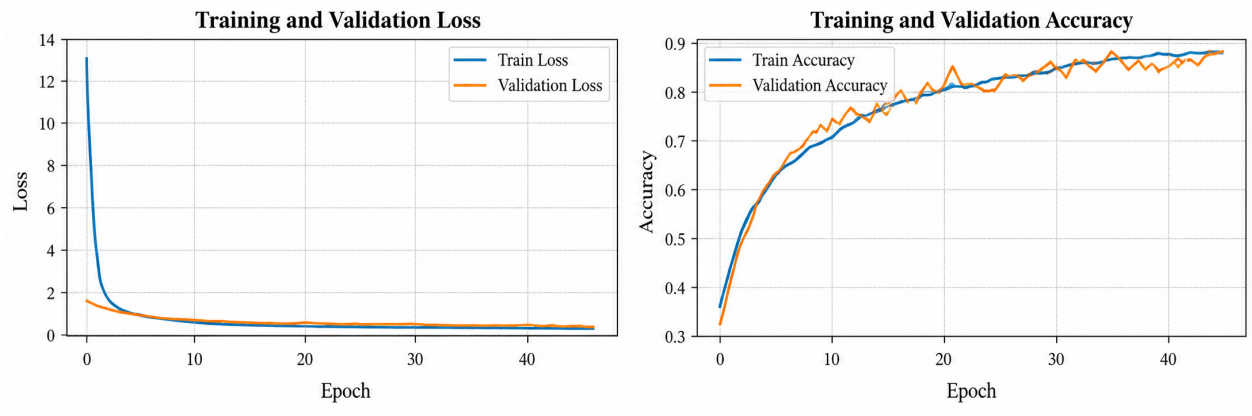


Figure 6. Training and validation loss and accuracy of the TransMedNet model.

Figure 6 illustrates the reported optimization behavior of TransMedNet through the training and validation loss and accuracy trajectories, providing a visual representation of convergence under the stated training configuration. Inspection of the reported classification behavior indicates that challenging cases are associated particularly with low-contrast slices, irregular tumor boundaries, small lesions with limited distinction from surrounding tissue, and residual preprocessing artifacts. These observations identify image quality and morphological overlap as important factors for future robustness analysis. The reported TransMedNet configuration contains 42.8M parameters and requires 9.6 GFLOPs per forward pass, with an inference time of 38 ms per slice on an RTX-3080 GPU. These measurements characterize the stated hardware configuration and should not be generalized directly to other devices or deployment environments.

4.3. Limitations and Reproducibility Considerations

The present study has several limitations that define the scope of its conclusions. First, evaluation is based on a single publicly available Kaggle MRI dataset; independent multi-institutional evaluation across different scanners, acquisition protocols, and patient populations is therefore required before broad generalizability can be established. Second, the available dataset information does not provide sufficient subject-level identifiers to independently verify complete patient-level separation solely from the reported image counts; potential subject overlap therefore remains an important consideration when interpreting the results. Third, a complete component-wise ablation of the hybrid embedding, dual attention, multiscale feature fusion, windowed attention, shifted-window attention, residual connections, and CNN stem is not available in the present experimental results, and consequently the individual quantitative contribution of each architectural component cannot be established from the current analysis. Fourth, fold-level prediction records and corresponding confidence intervals are not available for independent uncertainty analysis; future evaluation should report fold-level results together with confidence intervals and effect-size measures. Fifth, the present study does not include independent radiologist assessment or localization annotations for quantitative evaluation of interpretability, and attention-based feature refinement should therefore be understood as an architectural mechanism rather than validated clinical explainability. Finally, future reproducibility studies should provide fixed data splits, subject identifiers where permitted, random seeds, environment specifications, training configurations, model checkpoints, prediction records, and metric-computation scripts to facilitate independent verification and comparison of the framework.

5. Conclusion and Future Scope

The proposed TransMedNet framework combines convolutional neural networks with Vision Transformers for brain tumor classification from MRI images. Convolutional processing provides localized spatial representations, while transformer self-attention models longer-range contextual dependencies. The framework additionally incorporates hierarchical attention and multiscale feature integration. Within the comparisons reported in this study, TransMedNet records competitive performance relative to the listed CNN, CNN–RNN, ConvNeXt, and ResNet101 baselines across the presented evaluation summaries. These findings are specific to the dataset and protocol considered here and should not be interpreted as universal state-of-the-art or clinical superiority. Future work should prioritize independent multi-center and scanner-diverse validation, strict subject-level partitioning, component-wise ablation studies, fold-level uncertainty analysis, and expert-based assessment of model interpretability. Evaluation on additional brain MRI datasets will also be important for determining robustness under domain shift. Further architectural investigation can include 3D MRI processing and multimodal imaging to capture volumetric and complementary diagnostic information. Lightweight transformer designs may additionally reduce computational requirements for resource-constrained inference. Semi-supervised and federated learning strategies also provide potential directions for exploiting larger distributed datasets while addressing annotation scarcity and privacy constraints. These extensions require independent experimental validation before conclusions regarding clinical applicability can be established.

REFERENCES

1. A. Farzamnia, S. H. Hazaveh, S. S. Siadat, and E. G. Moung. Mri brain tumor detection methods using contourlet transform based on time adaptive self-organizing map. *IEEE Access*, 11:113480–113492, 2023.
2. A. A. Asiri, T. A. Soomro, A. A. Shah, G. Pogrebna, and M. Irfan. Optimized brain tumor detection: A dual-module approach for mri image enhancement and tumor classification. *IEEE Access*, 12:42868–42887, 2024.
3. L. Sun, Y. Zhang, and H. Wang. Star-yolo: A high-accuracy and ultra-lightweight method for brain tumor detection. *IEEE Access*, 2025.
4. N. M. H. Hassan and W. Boulila. Efficient approach for brain tumor detection and classification using fuzzy thresholding and deep learning algorithms. *IEEE Access*, 2025.
5. M. Wageh and K. Amin. Brain tumor detection based on deep features concatenation and machine learning classifiers with genetic selection. *IEEE Access*, 2024.
6. L. A. Abraham, G. Palanisamy, and V. Goutham. Dilated convolution and yolov8 feature extraction network: An improved method for mri-based brain tumor detection. *IEEE Access*, 2025.
7. K. Neamah, F. Mohamed, M. M. Adnan, T. Saba, S. A. Bahaj, and K. A. Kadhim. Brain tumor classification and detection based dl models: A systematic review. *IEEE Access*, 12:2517–2542, 2023.
8. W. Wang, X. Zhang, Y. Li, and J. Liu. A two-stage generative model with cyclegan and joint diffusion for mri-based brain tumor detection. *IEEE Journal of Biomedical and Health Informatics*, 28(6):3534–3544, 2024.
9. M. Z. Khaliki and M. S. Başarslan. Brain tumor detection from images and comparison with transfer learning methods and 3-layer cnn. *Scientific Reports*, 14:2664, 2024.
10. B. Srinivas and P. S. Reddy. A fine-tuned transformer model for brain tumor detection and classification. *Multimedia Tools and Applications*, pages 1–25, 2024.
11. A. S. Shinde and S. Patil. Performance analysis of machine learning algorithm of detection and classification of brain tumor using computer vision. *Advances in Engineering Software*, 173:103221, 2022.
12. J. Wang and X. Zhang. Ranmerformer: Randomized vision transformer with token merging for brain tumor classification. *Neurocomputing*, 573:127216, 2024.
13. A. K. Sharma and A. Kumar. Hog transformation based feature extraction framework in modified resnet50 model for brain tumor detection. *Biomedical Signal Processing and Control*, 84:104737, 2023.
14. D. Zhou and Y. Zhang. Deepvit: Towards deeper vision transformer. *arXiv preprint arXiv:2103.11886*, 2021.
15. K. Han and Y. Zhang. A survey on vision transformer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1):87–110, 2022.
16. A. Arnab and W. Matusik. Vivit: A video vision transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021.
17. Z. Liu and Y. Lin. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021.
18. A. Tariq and M. Khan. Transforming brain tumor detection empowering multi-class classification with vision transformers and efficientnetv2. *IEEE Access*, 2025.
19. S. Poornam and J. J. R. Angelina. Vitalt: A robust and efficient brain tumor detection system using vision transformer with attention and linear transformation. *Neural Computing and Applications*, 36(12):6403–6419, 2024.

20. P. T. Krishnan and S. Kumar. Enhancing brain tumor detection in mri with a rotation invariant vision transformer. *Frontiers in Neuroinformatics*, 18:1414925, 2024.
21. A. A. Asiri and T. A. Soomro. Exploring the power of deep learning: Fine-tuned vision transformer for accurate and efficient brain tumor detection in mri scans. *Diagnostics*, 13(12):2094, 2023.
22. A. A. Asiri and A. A. Shah. Advancing brain tumor classification through fine-tuned vision transformers: A comparative study of pre-trained models. *Sensors*, 23(18):7913, 2023.
23. E. Simon and A. Briassouli. Vision transformers for brain tumor classification. In *BIOIMAGING*, 2022.
24. P. Chauhan and D. Patel. Pbvit: A patch-based vision transformer for enhanced brain tumor detection. *IEEE Access*, 2024.
25. A. Drevelegas. *Imaging of Brain Tumors with Histological Correlations*. Springer, 2002.
26. G. Litjens et al. A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42:60–88, 2017.
27. A. Dosovitskiy and T. Brox. Discriminative unsupervised feature learning with exemplar convolutional neural networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(9):1734–1747, 2020.
28. Elmehdi Aniq, Faouzi-Ayoub El Ghanaoui, and Mohamed Chakraoui. Vision transformers for breast cancer mammographic image classification. *Statistics, Optimization & Information Computing*, 15(2):1087–1098, 2025.
29. Dawlat Abdulkarim Ali and Haval Tariq Sadeeq. An interpretable deep learning framework for multi-class dental disease classification from intraoral rgb images. *Statistics, Optimization & Information Computing*, 14(6):3380–3397, 2025.
30. Shivan Hussein Hassan and Najdavan Abduljawad Kako. Attention-augmented efficientnetv2b0 for multi-class cardiac disease classification from cine mri. *Statistics, Optimization & Information Computing*, 14(5):2605–2623, 2025.
31. J. Chen and Y. Zhang. Transunet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306*, 2021.
32. S. Bhuvaji. Brain tumor classification (mri). Kaggle, 2020. Online dataset.
33. C. Peng and Y. Zhang. Research of image recognition method based on enhanced inception-resnet-v2. *Multimedia Tools and Applications*, 81(24):34345–34365, 2022.
34. Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.
35. R. Vankdothu, M. A. Hameed, and H. Fatima. A brain tumor identification and classification using deep learning based on cnn-lstm method. *Computers & Electrical Engineering*, 101:107960, 2022.
36. R. Pillai and S. Kumar. Brain tumor classification using vgg16, resnet50, and inception v3 transfer learning models. In *Proceedings of INOCON*, 2023.
37. A. Saleh, R. Sukaik, and S. S. Abu-Naser. Brain tumor classification using deep learning. In *Proceedings of iCareTech*, 2020.
38. Y. Yang and N. Razmjooy. Early detection of brain tumors: Harnessing the power of gru networks and hybrid dwarf mongoose optimization algorithm. *Biomedical Signal Processing and Control*, 91:106093, 2024.
39. D. Reyes and J. Sánchez. Performance of convolutional neural networks for the classification of brain tumors using magnetic resonance imaging. *Heliyon*, 10(3):e12493, 2024.
40. P. Podder and S. Roy. Transfer learning for classification of brain tumor. In *Deep Learning for Biomedical Applications*. CRC Press, 2021.
41. P. Ghosal and S. Roy. Brain tumor classification using resnet-101 based squeeze and excitation deep neural network. In *Proceedings of the Second International Conference on Advanced Computational*, 2019.