Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 4, Issue 1, 2026

MULTI-PLANE ATTENTION-AIDED CNN FOR DETECTION OF BRAIN TUMOR IN MRI SCANS

Sami ul Haq[1], Saddam Hussain Khan[*2], Ayaz Khan[3], Syed Asghar Ali Shah[4]

[1,*2]*Artificial Intelligence Lab, Department of Computer Systems Engineering, University of Engineering and Applied Sciences (UEAS), Swat 19060, Pakistan*
[3]*Department of Artificial Intelligence, The University of Veterinary and Animal Science (UVAS), Swat, Pakistan*
[4]*Department of Computer Systems Engineering, University of Engineering and Applied Sciences (UEAS), Swat 19060, Pakistan*

[1]samiulhaq198@gmail.com, [*2]hengrshkhan822@gmail.com, [3]findayaz1@gmail.com, [4]syedasgharalishah@ueas.edu.pk

DOI: https://doi.org/10.5281/zenodo.18374595

Keywords
Brain Tumor Classification, MRI, Multi-plane CNN, Attention Fusion, Deep Learning, EfficientNet-B0, Automated Diagnosis.

Article History
Received: 01 December 2025
Accepted: 11 January 2026
Published: 26 January 2026

Corresponding Author: *
Hussain Khan

Abstract
The classification of brain tumors from MRI images plays an important role in assisting radiologists to get accurate and reliaable diagnostic decisions. Existing deep learning approaches often rely on single-plane MRI analysis, which limits the exploitation of complementary spatial information available across multiple anatomical views. To address this limitation, we have proposed Multi-Plane Attention-based Convolutional Neural Network (MPA-CNN) for automatic detection of brain tumor. The proposed framework independently processes axial, coronal, and sagittal MRI planes using customized EfficientNet-B0 backbones to extract discriminative deep features. An attention-based fusion mechanism is then employed to adaptively weight and integrate multi-plane features into a unified representation, which is subsequently classified using a fully connected prediction head with label smoothing. The proposed model is analyzed on BRISC2025 brain MRI dataset which contain four classes. Among these classes, one is healthy class named 'no_tumor' and remaining three classes are diseased one, which are: glioma, meningioma, and pituitary. The experimentatl findings shows that our proposed model has achieved precision of 99.31%. It is achieved 99.30% for precision, recall, and F1-score. Moreover, it has achieved highest class-wise performance, and ROC-AUC values for all types of tumors. On the basis of these results, it is concluded that multi-plan attention fusion for automated brain tumor classification is very effective and reliable in clinical decision-making. These results confirm the effectiveness and robustness of multi-plane attention fusion for automated brain tumor classification, highlighting its potential for reliable clinical decision support.

1. Introduction
Brain tumors are severe neurological conditions that have life threatening effects in case if they are not diagnosed at an initial stage. The correct and timely identification of brain tumor is very necessary for clinical decision making, planning treatment, and prognosis. MRI images have highest level of soft-tissue contrast and non-invasive visualization, so these are most commonly used for identification of

brain tumors. The manual analysis of these images is very tedious, prone to errors, consumes time, and needs expert radiologists.

Due to these shortcomings, it is very mandatory to have automated diagnostic systems [1], [2], [3]. Recently, machine learning algorithms especially CNNs have achieved very best result in analysis of medical images, like classification of brain tumors, their segmentation, and detection [3]–[5]. Previous studies shows that advanced CNN-based frameworks such as VGG, ResNet, DenseNet, and EfficientNet are highly effective for brain tumor classification, consistently delivering competitive and reliable accuracy across different datasets [3], [7] – [17]. Despite of these advancements, many of the existing approaches are reliable on single-plane MRI images, particularly on axial slices.

Recently, analysis of medical images is performed with the help of hybrid architectures which uses combinations of CNNs, attention mechanisms, and transformer models to extract more discriminative and robust feature representations. Attention frameworks have shown great abilities of selective highlighting of diagnostically salient characteristics and inhibiting noise, which results in better performance in more complex medical imaging tasks. As an example, models based on attention residual and transformer-integrated models have been effectively implemented to classification of noisy brain images and lesion analysis and have demonstrated greater discriminative capacity and generalization performance [12], [13]. In the same way, CNN-Transformer hybrids with channel-aware as well as spatial attention have shown encouraging outcomes in the medical image segmentation and disease diagnostics [14], [15], [18].

The scans of the brain MRI are usually obtained in three orthogonal directions, which are axial, coronal and sagittal. Each of the planes is providing anatomically complementary views of the tumor, capturing the tumor characteristics (shape, location, and boundary, etc.) in a different manner. Failure to pay attention to such multi-plane information may lead to underrepresentation of features and loss of generalization capability [8], [19] - [21]. In spite of the fact that the idea of multi-view or multi-slice learning has been studied in some recent works, most of the methods use simple feature concatenation or majority voting approaches, which do not take into account the relative significance and dependencies among various planes [22], [23].

Attention mechanisms have become an effective way to improve feature representation by dynamically emphasizing the most informative part of input data. Attention-based models have demonstrated enhanced performance in medical imaging in organ segmentation, lesion detection and disease classification focusing on diagnostically useful regions/features [24]–[27]. Nevertheless, adaptive fusion of multi-plane MRI features in brain tumor classification using attention mechanisms is not studied to a significant extent so far.

In addition to the conventional CNN architectures, recent literature has examined transformer-based and mixture-of-experts models to enhance efficiency and flexibility in clinical decision-support systems. Advanced attention-guided routing and feature enhancement strategies have been shown to dynamically prioritize informative and convincing representations, improving accuracy as well as computational efficiency in applications of medical and healthcare applications [28], [29]. According to recent surveys and applied studies, models which are based on vision transformers and their hybrids have achieved greater popularity in medical research because they are powerful enough to identify global contextual information from medical images [14], [30]. These processes show a rising trend of adaptive and attention-oriented fusion approaches that are capable of managing heterogeneous and multi-view information.

To address the identified challenges, we have proposed Multi-Plane Attention-driven CNN which automatically classify MRI images of brain tumors. Our framework is based on EfficientNet-B0 and is able to extract complementary features from axial, coronal, and sagittal anatomical planes. A fusion module that is attention based is then utilized to dynamically weight and fuse multi-plane features into an integrated feature, so that the model can utilize complementary spatial information effectively. The fused features are then classified with a fully connected prediction head that has been optimized to use label smoothing to improve generalization.

The BRISC2025 brain MRI data is used to test the proposed MPA-CNN. The dataset contains four different classes. Three classes are diseased one, which are: glioma, meningioma, and pituitary. The one class is healthy which is no_tumor. We have performed large number of experiments on the basis of which we have proved that our proposed model is very powerful to achieve the great level of accuracy. These findings show that multi-plane attention fusion is effective to enhance the classification performance and reliability.

The final contributions of proposed methods are given below:

- An emergency multi-plane classification framework that jointly makes use of the axial, coronal, and sagittal MRI images for classification of brain tumor
- An attention-based feature fusion mechanism that adaptively weights multi-plane features, capturing inter-plane dependencies.
- Robust deep learning model based on customized EfficientNet-B0, optimized with label smoothing and modern training strategies.

2. Literature Review

MRI images have achieved strong research interest for classification of brain tumors because of the availability of huge amount of datasets and improvements in deep learning techniques. This section reviews existing studies related to brain tumor classification, focusing on single-plane CNN-based methods, multi-view and multi-plane approaches, and attention-based classification frameworks, highlighting their strengths and limitations.

Initially, deep learning algorithms was entirely based on simple CNNs that were trained on MRI slices of two-dimensional [31]. Cheng et al. [32] have used a CNN-based model with the axial MRI slices as inputs to classify multiple tumors, and have achieved promising accuracy but they were suffered from limited generalization because of ambiguity at slice level. In a similar way, CapsNet-based architectures have been suggested by Afshar et al. [33] to maintain spatial hierarchies, but their performance was poor with complicated tumor shapes.

Subsequently, other researchers have explored deeper architectures like VGG, ResNet and DenseNet. In [34], Sultan et al. employed deep CNNs with transfer learning to categorize brain tumors like glioma, meningioma, and pituitary tumors, and achieved notable accuracy on standard benchmark datasets. Despite these advancements, their methods was working on single-plan MRI views often axial slices, so they cannot be used to obtain complete spatial tumor images [35].

A number of researchers have investigated a multi-view or multi-slice learning approach to take advantage of more abundant spatial data. Pashaei et al. [22] suggested a multi-slice CNN model, which consolidates predictions of neighboring axial slices through majority voting. Even though this has enhanced robustness, inter-slice relationships were not explicitly modeled.

More enhanced multi-plane methods were implemented in order to have axial, coronal and sagittal views. Diaz-Pernas et al. [36] used independent CNN branches of various MRI planes and combined features by basic concatenation. Although multi-plane learning was found to be more accurate than single-plane based approaches, naive fusion techniques were ineffective at learning the relative significance of the planes, and gave redundant or noisy features.

Attention mechanisms have been extensively applied in medical imaging to provide greater discrimination on features by attending to salient regions or channels [3], [37] – [40]. Wang et al. [41] presented attention-based CNNs to classify medical images, demonstrating better interpretability and output. Attention modules have been used in analysis of brain tumors mostly in segmentation tasks [23], [42], and few studies have been done in pure classification problems.

More recent studies on classification utilizing attention mechanisms concentrating on spatial or channel attention in a single input stream [43]. Nevertheless, the area of adaptive attention-based multimodal fusion through several MRI planes has not been studied well. According to most of the current multiplane models, planes are all handled in the same manner regardless of their different levels of diagnosis with different types of tumors and their location.

From the above review, several key limitations can be identified:

- Most existing methods rely on single-plane MRI analysis, underutilizing 3D anatomical information.
- Multi-plane approaches often use simple feature concatenation or voting, lacking adaptive fusion mechanisms.
- Attention mechanisms are rarely applied for inter-plane feature weighting in classification tasks.
- Comprehensive evaluations using multiple performance metrics (ROC-AUC, class-wise robustness) are often missing.

To address these gaps, this work proposes a Multi-Plane Attention-based CNN (MPA-CNN) that adaptively fuses axial, coronal, and sagittal MRI features using an attention mechanism, enabling robust and discriminative brain tumor classification.

Table 1. Comparison of existing methods of classification.

Author (Year)	Dataset	Model	Accuracy
Cheng et al; (2015) [17]	Private MRI	CNN	91%
Afshar et al. (2018) [18]	BRATS	CapsNet	86%
Sultan et al. (2019) [19]	Figshare MRI	Deep CNN	98.7%
Pashaei et al. (2018) [20]	BRATS	CNN + Extreme Learning Machines	81.09%
Díaz-Pernas et al. (2021) [21]	Public MRI	Multi-branch CNN	97.3%
Pereira et al. (2016) [23]	BRATS	Patch-based CNN	88%
Havaei et al. (2017) [24]	BRATS	Multi-path CNN	87%
Li et al. (2019) [25]	ImageNet	Selective Kernel Network (SKNet)	81.5%
Proposed MPA-CNN (2025)	**BRISC2025**	**EfficientNet-B0 + Attention Fusion**	**99.30%**

3. Methodology

In this section, we will explain our proposed model in detail. The central motivation of the proposed methodology is to exploit complementary diagnostic information present in multiple anatomical planes of MRI while dynamically emphasizing the most informative views using an attention-driven fusion strategy. Unlike conventional single-plane CNN approaches, the proposed framework jointly learns discriminative representations from axial, coronal, and sagittal MRI views and integrates them in a principled manner for robust multi-class tumor classification.

The MPA-CNN framework proposed by us is shown in Figure 1. The entire framework is trained in such a way which allows the attention mechanism and classification layers to adaptively learn optimal representations directly from data.

3.1 Data Preparation and Preprocessing

The methodology is evaluated on the basis of BRISC2025 dataset which includes four classes, which are: glioma, meningioma, pituitary tumor, and no tumor. For each of the class, sample MRI images are shown in Figures 2, 3, 4, and 5. Let the MRI dataset be denoted as:

$$D = \{(X_i, y_i)\}_{i=1}^{N}$$

where Xi represents the i-th MRI sample and yi ∈ {1,2,3,4} denotes the related class labels. Every MRI sample Xi consists of three orthogonal anatomical planes:

$$X_i = \{X_i^{ax}, X_i^{co}, X_i^{sa}\}$$

where $X_i^{ax}, X_i^{co}, X_i^{sa}$ respectively represents the axial, coronal, and sagittal views. All slices of MRI are resized to resolution of 224 x 224 pixels. Images are transformed into the three-channel form and is normalized with values of standard deviation σ and ImageNet mean μ, which helps to easily transfer learning in the already-trained models. Mathematically,

$$\hat{X} = \frac{X - \mu}{\sigma}$$

The conventional augmentation method like horizontal flipping is used in training to minimize overfitting and to enhance the robustness. Each MRI sample is processed through the network simultaneously with the corresponding axial, coronal and sagittal images.

3.2 Multi-Plane Feature Extraction Using CNN Backbones

The proposed framework is used to establish discriminative features in each anatomical plane separately, using three parallel branches of convolutional neural networks each focused on a given MRI plane. The backbone architecture of all the three branches is EfficientNet-B0, which has high representational capability and low computational time.

Let $\phi(\cdot\,;\theta)$ denote the convolutional feature extraction function parameterized by weights θ. The plane-specific feature representations are obtained as:

$$f_{ax} = \phi(X^{ax};\theta_{ax})$$

$$f_{co} = \phi(X^{co};\theta_{co})$$

$$f_{sa} = \phi(X^{sa};\theta_{sa})$$

where $f_{ax}, f_{co}, f_{sa} \in R^d$ represent the high-level semantic features which respectively extracted from axial, coronal, and sagittal planes. EfficientNet-B0 final classification layers are eliminated and convolutional feature extractor is retained. These attributes encode complementary tumor properties like boundary abnormalities, differences in texture and spatial distributions of tumor which cannot be fully determined from a single view.

3.3 Attention-Based Multi-Plane Feature Fusion

One of the key contributions of the suggested methodology is the attention-based fusion mechanism, which adaptively models the relative significance of each of the MRI planes in classification. Instead of concatenation or averaging of the extracted plane specific features, the plane specific features are first concatenated to create a single representation, and then such combined feature vector is passed through lightweight attention network containing fully connected layers and non-linear activation functions. Mathematically,

$$f_{cat} = |f_{ax}||f_{co}||f_{sa}|$$

where $||$ denotes vector concatenation. The concatenated feature vector is passed through a lightweight attention network to compute plan-wise importance weighs. Mathematically,

$$\propto = Softmax\left(W_2\sigma(W_1 f_{cat})\right)$$

where W1 and W2 reresents learnable weight matrices, and $\sigma\,(.)$ represents the ReLU activation function. The following equation represents the weights which are assigned to each anatomical plane.

$$\propto = [\propto_{ax}, \propto_{co}, \propto_{sa}]$$

The final fused feature representation is computed as a weighted sum. Mathematically:

$$f_{fused} = \propto_{ax} f_{ax} + \propto_{co} f_{co} + \propto_{sa} f_{sa}$$

With the help of this strategy, the network is able to dynamically emphasize the most informative plane with respect to tumor type, size, and spatial characteristics.

3.4 Classification Head

The combined multi-plane feature vector is given as input into fully connected classification head for multi-class prediction. This head includes a dense layer that projects the fused features into lower-dimensional embedding. For introduction of non-linearity, it is followed by ReLU activation. Mathematically, it can be written as:

$$z = ReLU(W_c f_{fused} + b_c),$$

Where Wc and bc represents the weights and bias of the dense layer. Regularization of dropout is applied to z for mitigation of overfitting and improvement of generalization. Final dense layer is then used for mapping of the learned representation to four output neurons with respect to related tumor classes. Softmax activation function is used to produce probability of class estimates, as:

$$\hat{y}k = \frac{e^{ok}}{\sum_{j=1}^{4} e^{oj}},\ \mathrm{k} \in \{1,2,3,4\}$$

The class having highest posterior probability is selected to get the predicted class label.

3.5 Loss Function and Optimization Strategy

Proposed model is trained with the help of cross-entropy loss having label smoothing. It prevents overconfident predictions as well as improves the ability of the model to be generalized for unseen data. Label smoothing redistributes a small portion of the target probability mass across non-target classes, making the learning process more robust to label noise. The smoothed target distribution is defined as:

$$y_k^{LS} = (1-\varepsilon)\cdot y_k + \frac{\varepsilon}{K}$$

where ε is the smoothing factor and K=4 is the number of classes. The loss function is given by:

$$L = -\sum_{k=1}^{k} y_k^{LS}\log(\hat{y}k)$$

Optimization is performed with the help of AdamW optimizer, which decouples weight decay from gradient updates, and provides steady convergence behavior. Mathematically,

$$\theta_{t+1} = \theta_t - \eta \frac{\widehat{m}_t}{\sqrt{\widehat{v}_t}} + \epsilon - \lambda\theta_t$$

where η is showing the learning rate and λ is showing the weight decay factor. To improve efficiency of training, a ReduceLROnPlateau scheduler is applied to reduce the learning rate automatically in case if validation loss not improved. This approach helps the model to overcome local minima and improve the performance.

3.6 Training Protocol and Model Selection

The network is trained in supervised manner for specific number of epochs using mini-batches gradient descent. Overfitting is mitigated by implementing strategy of early stopping with the help of predefined patience value. The model which is achieving best accuracy for validation is chosen as the final version. Parameters of that model are saved for future evaluations. This approach ensures that the reported results are corresponding to the most generalizable network configuration.

3.7 Evaluation Metrics

To check the performance of our proposed classification model, we have used several evaluation metrics like accuracy, precision, recall, and F1-score. The overall accuracy is the proportion of correctly classified samples. Mathematically, it is calculated as:

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

Furthermore, precision, recall, and F1-score for each class are computed as:

$$Precision = \frac{TP}{TP + FP}$$

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

$$F1 = \frac{2. Precision. Recall}{Precision + Recall}$$

For discriminative assessment of our proposed model, Receiver Operating Characteristic (ROC) and Area Under the Curve (AUC) is used for each of the four classes.

3.8 Computational Efficiency

Although the proposed MPA-CNN is based on three parallel CNN branches, it is computationally efficient because EfficientNet-B0 is lightweight and the attention fusion module has a small overhead. This structure renders the framework to be applicable to real world clinical decision support systems where accuracy and efficiency is very important.

4. Results and Discussion

This section includes a multi-dimensional examination of the suggested Multi-Plane Attention-based Convolutional Neural Network (MPA-CNN) of BRISC2025 brain MRI classification dataset. The outcomes of the models are discussed through quantitative measures like precision, recall, accuracy, F1-score, and ROC-AUC. All these measurements prove the strength, consistency and clinical applicability of the suggested method.

4.1 Overall Classification Performance

The suggested MPA-CNN has achieved a total test accuracy of 99.30%, which implies that it is very discriminative among all the four types of tumors. The model is attained 99.31% precision, and 99.30% recall, and F1-score. These values are indicating a well-balanced performance with minimal false positive and false negatives. Comparison of proposed model with other models is shown in detail in Table 2.

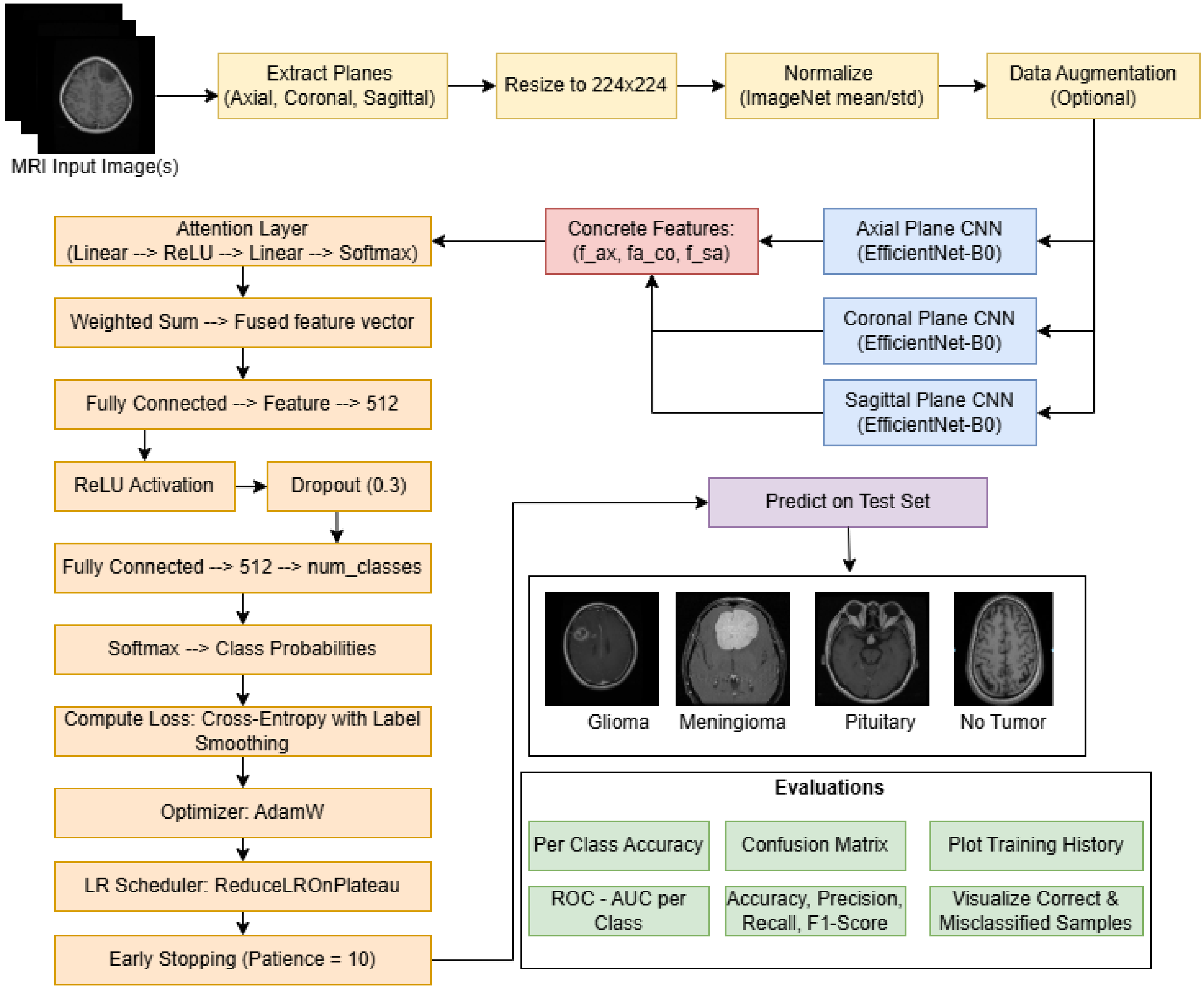

Figure 1: Proposed Methodology

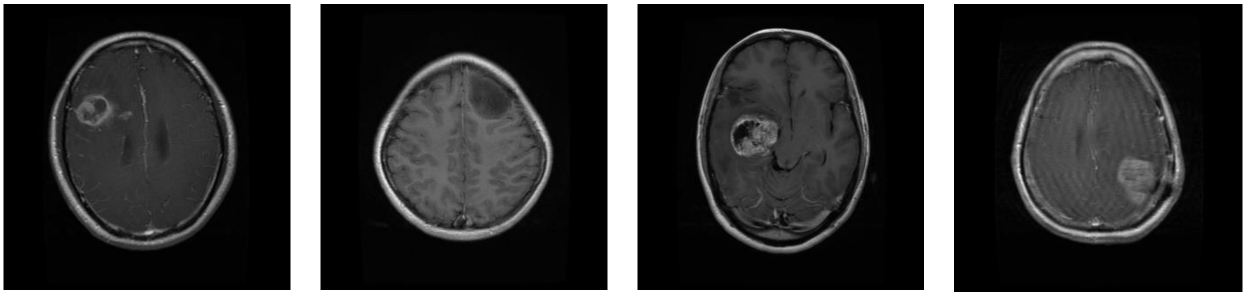

Figure 2: Sample Glioma MRI Images

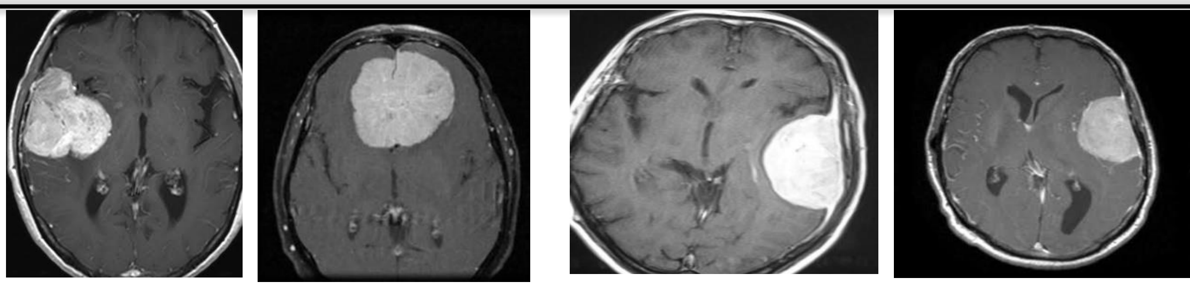

Figure 3: Sample Meningioma MRI Images

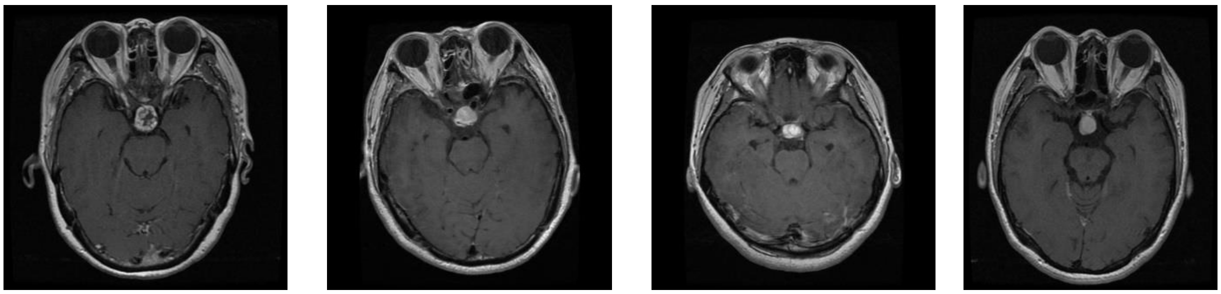

Figure 4: Sample Pituitary MRI Images

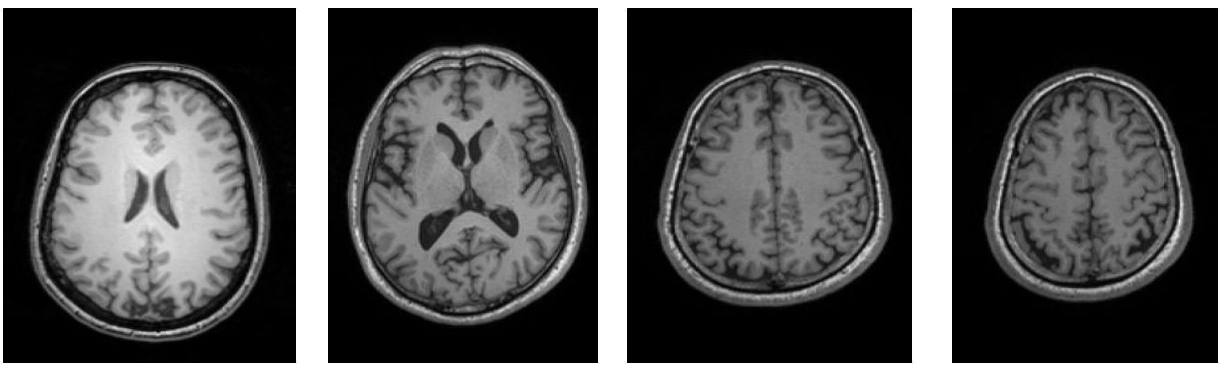

Figure 5: Sample No Tumor MRI Images

Table 2. Comparison of proposed model

Models	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
ResNet18	98.20	97.32	98.20	98.10
DenseNet121	97.00	97.28	97.35	89.09
MobileNetV2	96.40	95.36	96.30	96.10
VGG16	95.20	96.04	96.25	94.00
EfficientNet-B0	**99.30**	**99.31**	**99.30**	**99.30**

This type of high performance demonstrates the success of multiplane feature learning with attention-based fusion. The network is able to combine axial, coronal, and sagittal views to capture complimentary information in the anatomy that is usually overlooked by single view CNN models. The equal preciseness and recall rates also indicate that the model does not bias towards a specific class which is essential in the clinical decision making.

4.2 Confusion Matrix Analysis

The confusion matrices for other comparative models (ResNet18, DenseNet121, MobileNetV2, and VGG16) are shown in figure 6, 7, 8, and 9 respectively. The confusion matrix shown in Figure 10 indicates that our model has improved results as compared to other models. In the glioma class, the correct classification was done in 252 out of 254 samples, and the misclassification was only two. Confusion was slightly higher in the meningioma class with a few samples being wrongly classified as either pituitary or no-tumor. This action may be explained by the similarity of the visual features in the MRI slices in some cases of meningioma and pituitary tumors. It is notable that the no-tumor and the pituitary tumor were considerably classified with no misclassifications of no-tumor. This is especially critical in clinical screening cases, where false tumor diagnosis on healthy individuals should be reduced to a minimum. In general, it can be indicated that the proposed MPA-CNN can be proved to be effective based on the confusion matrix.

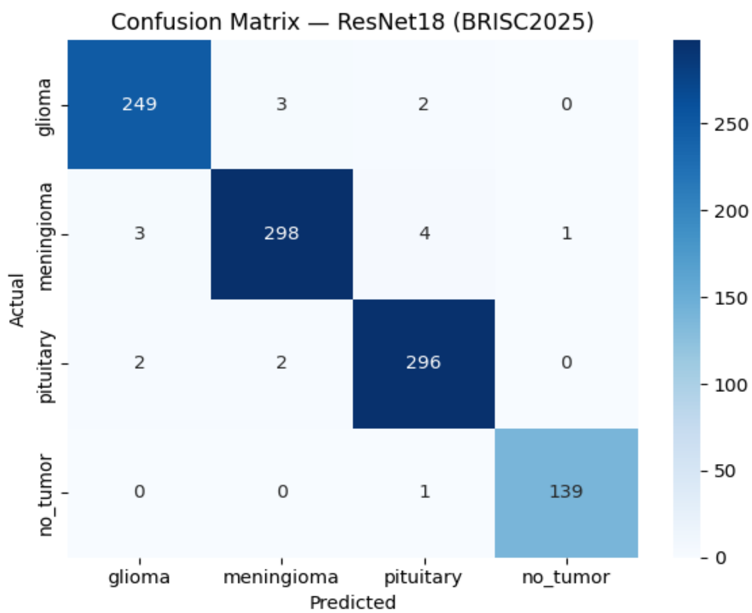

Figure 6. Confusion Matrix for ResNet18

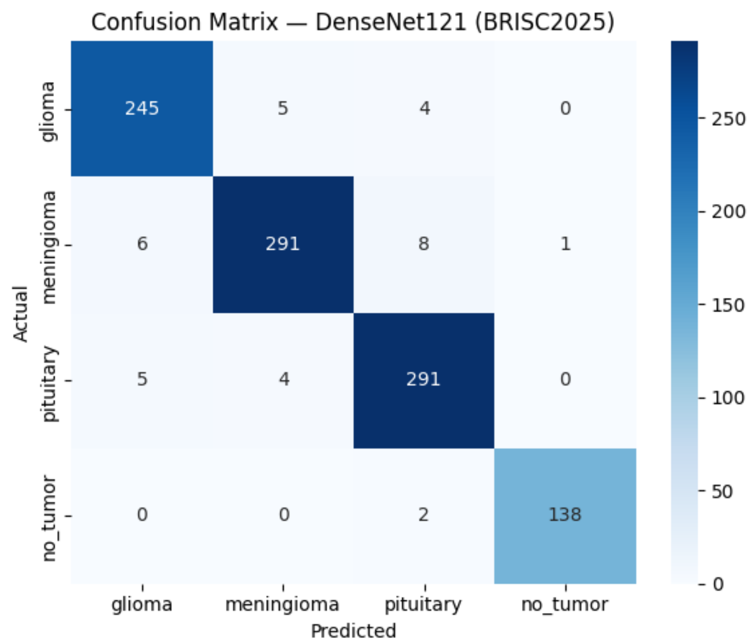

Figure 7. Confusion Matrix for DenseNet121

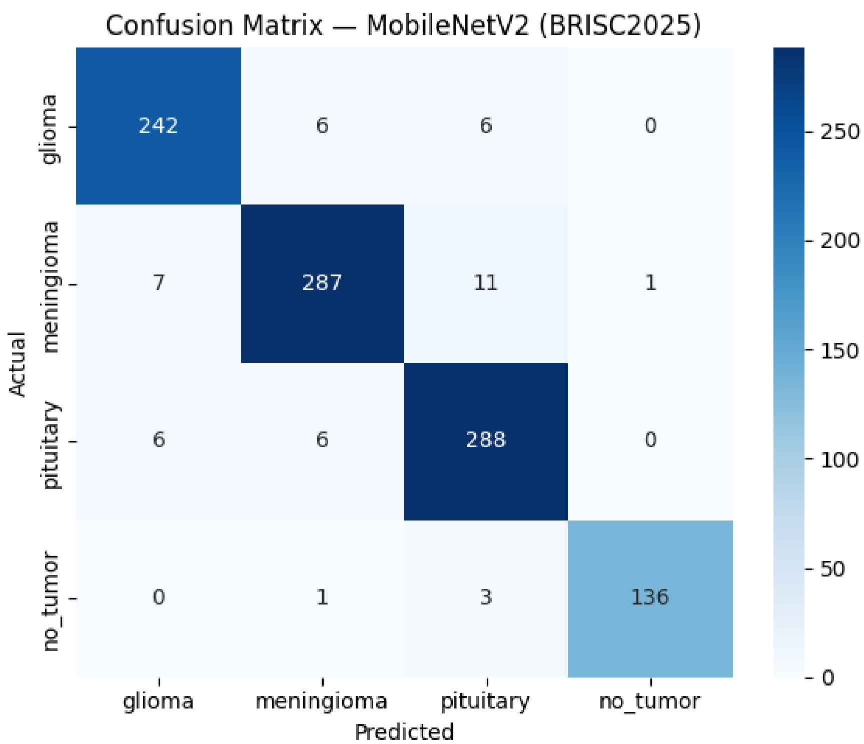

Figure 8. Confusion Matrix for MobileNetV2

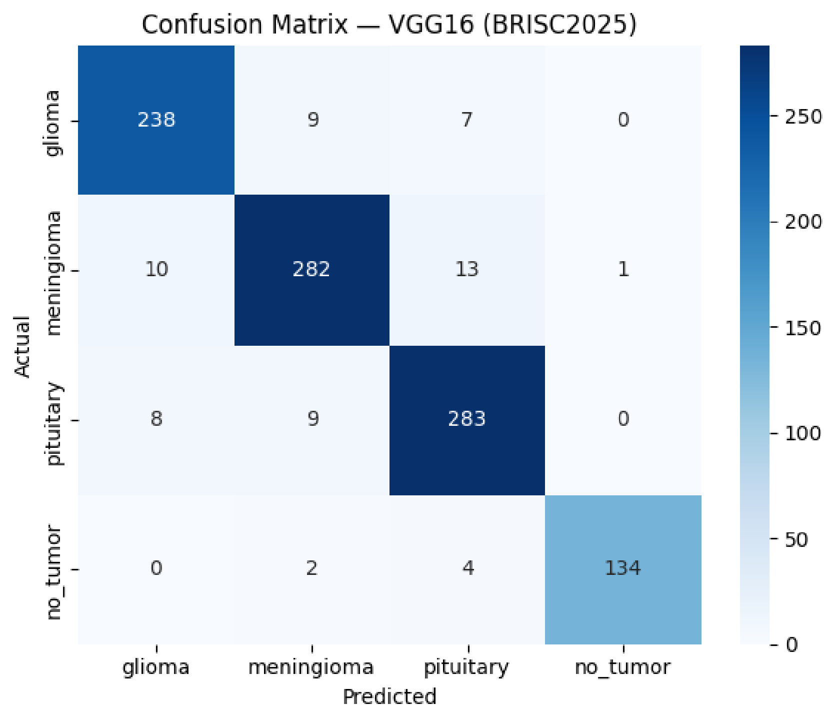

Figure 9. Confusion Matrix for VGG16

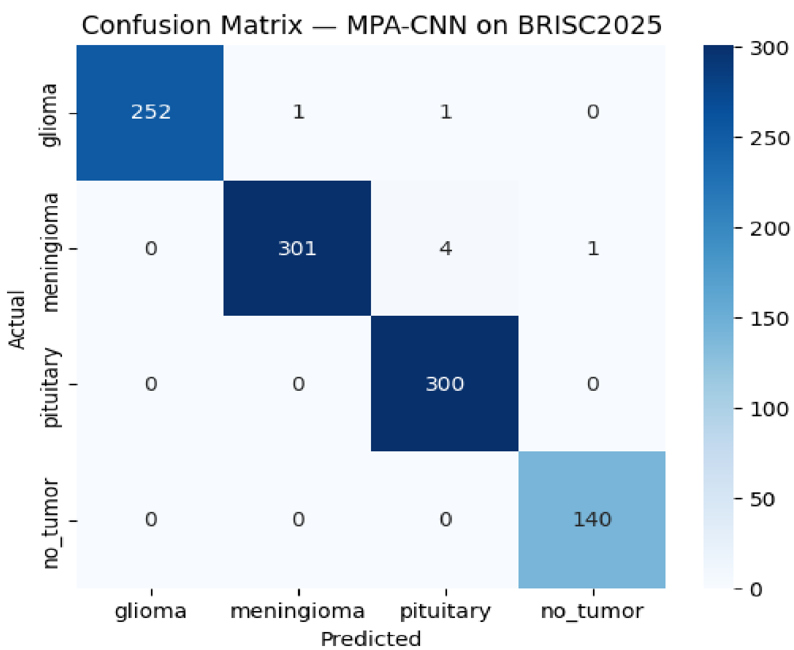

Figure 10. Confusion Matrix for MPA-CNN (Proposed Model)

4.3 Class-Wise Precision, Recall, and F1-Score

The soundness of the proposed model is also confirmed through performance analysis of brain tumor classes. Glioma category had a 100% precision, and a recall of 99% that represent very strong and representative tumor identification with low levels of false negativity. The meningioma group had a precision of 99.67% and a recall of 98.37%, which is considered to be very good given its comparatively complex visual shapes. In the case of the pituitary tumor classification, the model had a perfect recall of 100%, meaning that the cases of pituitary tumors were not missed. The no-tumor group also showed a perfect recall and F1 -score of 99.64%, which confirms the model as effective in differentiating healthy scans and pathological cases. All these findings confirm that the proposed architecture has a consistent and reliable performance in the two categories of tumors and non-tumors.

4.4 ROC-AUC Analysis

ROC analysis also gives more insight into the discriminative power of the proposed MPA-CNN. This model showed very high ROC-AUC values of all classes, glioma 99.98%, meningioma 99.63%, pituitary tumor 99.92%, and no-tumor 100%. These near-perfect values of AUC demonstrate that there is an excellent separability between classes and that the learned feature representations are most discriminative. The high ROC-AUC level is also a good indication that the proposed model would be useful in the clinical context, where the important criterion is the correct separation of classes.

4.5 Discussion and Comparative Insights

Our proposed model has three superior key factors. First, multi-plane MRI inputs allow the network to obtain the complementary anatomical data that cannot be used by single-plane models. Second, the attention-based fusion system dynamically selects the most informative MRI plane to each sample and enhances the adaptability of the fusion system to tumors. Lastly, the customized EfficientNet-B0 backbones provide efficient feature extractions with an insignificant computational cost. The proposed model addresses competitive or better performance with architectural simplicity compared to traditional CNN-based classification.

4.6 Clinical Relevance

Clinically, the proposed MPA-CNN has high potential in the role of a decision-support system to radiologists. The almost flawless classification of pituitary and no-tumor cases helps to minimize the possibilities of missed diagnoses and false alarms. Besides, the attention-based fusion strategy offers a measure of interpretability, by revealing implicitly, which MRI plane contributes most to the final decision.

5. Conclusion and Future Work

The model proposed by us for classification of brain tumor MRI image is customized MPA-CNN. The framework is automatically classifying the MRI images into four classes. The suggested framework is capable of merging complementary anatomical data in the axial, coronal and sagittal MRI planes via an attention-based fusion that allows strong and highly precise multi-classes classification. Unlike conventional single view or naïve multi-view approaches, the proposed method dynamically learns the relative significance of each anatomical plane, so increasing diagnostic reliability. The effectiveness of the suggested approach can be proved with the help of extensive experimental analysis of the BRISC2025 dataset. The MPA-CNN had a total classification accuracy of 99.30%. Other factors like precision, recall, and F1-score are also very high for all types of tumors. The proposed framework performs better than the existing ones: (i) it uses the multi-plane MRI data to learn a variety of tumor features, (ii) it uses an attention-based fusion mechanism that selectively focuses on views that are diagnostically meaningful, and (iii) it uses customized EfficientNet-B0 backbones that balance classification and computational efficiency. Collectively, these elements allow the proposed MPA-CNN to perform better than the conventional CNN-based brain tumor classification frameworks and still have a lightweight and scalable architecture that can fit well in clinical settings.

Although, the proposed model has very strong performance, there are still several directions for future researchers. First, future work may incorporate explicit interpretability mechanisms, such as attention visualization or gradient-based explainability methods, to further enhance clinical trust and transparency. Second, extending the framework to a joint multi-task setting, where classification is integrated with tumor localization or segmentation could provide richer clinical insights. Third, evaluation on multi-institutional and heterogeneous datasets would further validate the generalization of the proposed model. Finally, integrating transformer-based or state-space models, such as Mamba-style architectures, with the proposed attention framework represents a promising avenue for future exploration.

Acknowledgment:
We have completed this work with the support of University of Engineering and Applied Sciences. We are thankful to Artificial Intelligence Lab, Department of Computer Systems Engineering, for giving us access to computational resources and a collaborative research environment.

REFERENCES

[1] Gordillo, N., Montseny, E., & Sobrevilla, P. (2013). State of the art survey on MRI brain tumor segmentation. *Magnetic resonance imaging*, *31*(8), 1426-1438.

[2] Ortiz, A., Munilla, J., Gorriz, J. M., & Ramirez, J. (2016). Ensembles of deep learning architectures for the early diagnosis of the Alzheimer's disease. *International journal of neural systems*, *26*(07), 1650025.

[3] Zahoor, M. M., Qureshi, S. A., Bibi, S., Khan, S. H., Khan, A., Ghafoor, U., & Bhutta, M. R. (2022). A new deep hybrid boosted and ensemble learning-based brain tumor analysis using MRI. Sensors, 22(7), 2726.

[4] Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., ... & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical image analysis*, *42*, 60-88.

[5] Shen, D., Wu, G., & Suk, H. I. (2017). Deep learning in medical image analysis. *Annual review of biomedical engineering*, *19*(1), 221-248.

[6] Suzuki, K. (2017). Overview of deep learning in medical imaging. *Radiological physics and technology*, *10*(3), 257-273.

[7] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

[8] Zahoor, M. M., Khan, S. H., Alahmadi, T. J., Alsahfi, T., Mazroa, A. S. A., Sakr, H. A., ... & Alshemaimri, B. K. (2024). Brain tumor MRI classification using a novel deep residual and regional CNN. Biomedicines, 12(7), 1395.

[9] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

[10] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4700-4708).

[11] Tan, M., & Le, Q. (2019, May). Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning* (pp. 6105-6114). PMLR.

[12] Hameed, M., Zameer, A., Khan, S. H., & Raja, M. A. Z. (2024). ARiViT: attention-based residual-integrated vision transformer for noisy brain medical image classification. The European Physical Journal Plus, 139(5), 440.

[13] Khan, S. H. (2024). A novel Feature Map Enhancement technique integrating residual CNN and Transformer for Alzheimer diseases diagnosis. arXiv preprint arXiv:2405.12986.

[14] Naz, S., & Khan, S. H. (2025). Residual-SwinCA-Net: A Channel-Aware Integrated Residual CNN-Swin Transformer for Malignant Lesion Segmentation in BUSI. arXiv preprint arXiv:2512.08243.

[15] Iqbal, R., & Khan, S. H. (2025). RS-CA-HSICT: A Residual and Spatial Channel Augmented CNN Transformer Framework for Monkeypox Detection. arXiv preprint arXiv:2511.15476.

[16] Iqbal, R., & Khan, S. H. (2026). RSwinV2-MD: An Enhanced Residual SwinV2 Transformer for Monkeypox Detection from Skin Images. arXiv preprint arXiv:2601.01835.

[17] Zahoor, M. M., & Khan, S. H. (2025). CE-RS-SBCIT A Novel Channel Enhanced Hybrid CNN Transformer with Residual, Spatial, and Boundary-Aware Learning for Brain Tumor MRI Analysis. arXiv preprint arXiv:2508.17128.

[18] Khan, S. H., & Iqbal, R. (2024). RS-FME-SwinT: A Novel Feature Map Enhancement Framework Integrating Customized SwinT with Residual and Spatial CNN for Monkeypox Diagnosis. arXiv preprint arXiv:2410.01216.

[19] Mohsen, H., El-Dahshan, E. S. A., El-Horbaty, E. S. M., & Salem, A. B. M. (2018). Classification using deep learning neural networks for brain tumors. *Future Computing and Informatics Journal*, *3*(1), 68-71.

[20] Rehman, N. U., Zia, M. S., Meraj, T., Rauf, H. T., Damaševičius, R., El-Sherbeeny, A. M., & El-Meligy, M. A. (2021). A self-activated cnn approach for multi-class chest-related COVID-19 detection. *Applied Sciences*, *11*(19), 9023.

[21] Khan, S. H., Khan, A., Lee, Y. S., Hassan, M., & Jeong, W. K. (2023). Segmentation of shoulder muscle MRI using a new Region and Edge based Deep Auto-Encoder. Multimedia Tools and Applications, 82(10), 14963-14984.

[22] Pashaei, A., Sajedi, H., & Jazayeri, N. (2018, October). Brain tumor classification via convolutional neural network and extreme learning machines. In *2018 8th International conference on computer and knowledge engineering (ICCKE)* (pp. 314-319). IEEE.

[23] Pereira, S., Pinto, A., Alves, V., & Silva, C. A. (2016). Brain tumor segmentation using convolutional neural networks in MRI images. *IEEE transactions on medical imaging*, *35*(5), 1240-1251.

[24] Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., ... & Rueckert, D. (2018). Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*.

[25] Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7132-7141).

[26] Wang, X., Girshick, R., Gupta, A., & He, K. (2018). Non-local neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7794-7803).

[27] Rauf, Z., Sohail, A., Khan, S. H., Khan, A., Gwak, J., & Maqbool, M. (2023). Attention-guided multi-scale deep object detection framework for lymphocyte analysis in IHC histological images. Microscopy, 72(1), 27-42.

[28] Khan, H., & Khan, S. H. (2026). MambaFormer: Token-Level Guided Routing Mixture-of-Experts for Accurate and Efficient Clinical Assistance. arXiv preprint arXiv:2601.01260.

[29] Ullah, W., Khalid, Y. N., & Khan, S. H. (2025). A Novel Deep Hybrid Framework with Ensemble-Based Feature Optimization for Robust Real-Time Human Activity Recognition. arXiv preprint arXiv:2508.18695.

[30] Khan, A., Rauf, Z., Khan, A. R., Rathore, S., Khan, S. H., Shah, N., ... & Gwak, J. (2025). A recent survey of vision transformers for medical image segmentation. IEEE Access.

[31] Ahmed, U., Khan, A., Khan, S. H., Basit, A., Haq, I. U., & Lee, Y. S. (2019). Transfer learning and meta classification based deep churn prediction system for telecom industry. arXiv preprint arXiv:1901.06091.

[32] Cheng, J., Huang, W., Cao, S., Yang, R., Yang, W., Yun, Z., ... & Feng, Q. (2015). Enhanced performance of brain tumor classification via tumor region augmentation and partition. *PloS one*, *10*(10), e0140381.

[33] Afshar, P., Mohammadi, A., & Plataniotis, K. N. (2020). Bayescap: A bayesian approach to brain tumor classification using capsule networks. *IEEE Signal Processing Letters*, *27*, 2024-2028.

[34] Sultan, H. H., Salem, N. M., & Al-Atabany, W. (2019). Multi-classification of brain tumor images using deep neural network. *IEEE access*, *7*, 69215-69225.

[35] Khan, S. H., & Iqbal, R. (2025). A Comprehensive Survey on Architectural Advances in Deep CNNs: Challenges, Applications, and Emerging Research Directions. arXiv preprint arXiv:2503.16546.

[36] Díaz-Pernas, F. J., Martínez-Zarzuela, M., Antón-Rodríguez, M., & González-Ortega, D. (2021, February). A deep learning approach for brain tumor classification and segmentation using a multiscale convolutional neural network. In *Healthcare* (Vol. 9, No. 2, p. 153). MDPI.

[37] Khan, S. H., Sohail, A., Khan, A., Hassan, M., Lee, Y. S., Alam, J., ... & Zubair, S. (2021). COVID-19 detection in chest X-ray images using deep boosted hybrid learning. Computers in Biology and Medicine, 137, 104816.

[38] Khan, S. H., Sohail, A., Khan, A., & Lee, Y. S. (2022). COVID-19 detection in chest X-ray images using a new channel boosted CNN. Diagnostics, 12(2), 267.

[39] Hussain Khan, S. (2022). COVID-19 Detection and Analysis From Lung CT Images using Novel Channel Boosted CNNs. arXiv e-prints, arXiv-2209.

[40] Khan, S. H., Shah, N. S., Nuzhat, R., Majid, A., Alquhayz, H., & Khan, A. (2022). Malaria parasite classification framework using a novel channel squeezed and boosted CNN. Microscopy, 71(5), 271-282.

[41] Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., & Summers, R. M. (2017). Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2097-2106).

[42] Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., ... & Larochelle, H. (2017). Brain tumor segmentation with deep neural networks. *Medical image analysis*, *35*, 18-31.

[43] Li, X., Wang, W., Hu, X., & Yang, J. (2019). Selective kernel networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 510-519).