IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 26 November 2024, accepted 22 December 2024, date of publication 3 January 2025, date of current version 13 January 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3525518

RESEARCH ARTICLE

GGLA-NeXtE2NET: A Dual-Branch Ensemble Network With Gated Global-Local Attention for Enhanced Brain Tumor Recognition

ADNAN SAEED[1], KHURRAM SHEHZAD[2], SHAHZAD SARWAR BHATTI[3], SAIM AHMED[4,5], AND AHMAD TAHER AZAR[4,5,6]

[1]Department of Computer Science and Information Technology, Lahore Leads University, Lahore 54792, Pakistan
[2]FG Public School Boys, Multan Cantt., Multan, Punjab, Pakistan
[3]Faculty of Computing and Emerging Technologies, Emerson University Multan, Multan 60000, Pakistan
[4]College of Computer and Information Sciences, Prince Sultan University, Riyadh 12435, Saudi Arabia
[5]Automated Systems and Soft Computing Laboratory (ASSCL), Prince Sultan University, Riyadh 11586, Saudi Arabia
[6]Faculty of Computers and Artificial Intelligence, Benha University, Banha 13511, Egypt

Corresponding author: Adnan Saeed (adnan6939188@gmail.com)

This work was supported by the Automated Systems and Soft Computing Laboratory (ASSCL), Prince Sultan University, Riyadh, Saudi Arabia.

ABSTRACT Due to the limited availability of training data, the diverse shapes of brain tumors among different patients, inter-class similarity, and intra-class variation, achieving high recognition accuracy and speed in deep learning-based brain tumor recognition remains challenging. To address these issues, we propose a Dual-Branch Ensemble and Gated Global-Local Attention network based on EfficientNetV2S and ConvNeXt (GGLA-NeXtE2NET) to improve identification accuracy and model interpretability. For inter-class and intra-class problems, we designed a Gated Global-Local Attention (GGLA) mechanism that captures dependency information of query points in both horizontal and vertical directions, thereby obtaining global information indirectly. Simultaneously, local information is captured through multiple convolutions with a gating layer. The gating mechanism within the GGLA dynamically balances the contributions of global and local information, enabling the model to adaptively focus on the most relevant features for accurate classification. Furthermore, we introduce a dual-branch ensemble network to address the issue of image variety. This network uses two branches to extract image features at different resolutions for fusion, thereby expanding the network receptive field. Additionally, we utilized an Enhanced Super-Resolution Generative Adversarial Network (ESRGAN) to generate images that balance MRI data and implemented multiple preprocessing techniques to tackle inherent noise in MRI images. These techniques enhance the clarity of MRI images while preserving essential details. This results in a clear improvement in the identification of tumor boundaries, crucial for accurate surgical planning and treatment strategies. We evaluated GGLA-NeXtE2NET on 3-class and 4-class brain tumor datasets and achieved 99.06%, and 99.62% overall accuracy on both datasets respectively.

INDEX TERMS MRI brain tumor, deep learning, image processing, medical image analysis, attention mechanism.

I. INTRODUCTION

Magnetic Resonance Imaging (MRI) is a non-invasive medical imaging technique that utilizes a powerful magnetic field and radio waves to produce intricate images of internal body structures, including the brain [1]. This versatile modality has become integral in modern medicine, providing detailed information about the size, location, type, and grade of tumors [2]. Additionally, MRI offers insights into the impact of tumors on surrounding brain

The associate editor coordinating the review of this manuscript and approving it for publication was Abedalrhman Alkhateeb.

tissue and blood vessels, making it a fundamental tool for precise diagnosis and treatment planning in clinical practice [3].

An essential component of MRI technology is the MRI scanner, a sizable cylindrical tube housing a potent magnet. During an MRI procedure, the patient reclines on a mobile bed that smoothly moves into the tube. Positioned around the patient's head is a specialized coil, responsible for transmitting and receiving radio waves. This coil generates a magnetic field, momentarily aligning the hydrogen atoms within the body. After the radio waves cease, the atoms revert to their original positions and emit signals, which the coil captures. These signals are then meticulously processed by a computer, culminating in the formation of a detailed image representing a cross-section of the brain [4]. By capturing multiple slices from diverse angles, the scanner constructs a comprehensive three-dimensional depiction of the brain's internal structures [5].

MRI stands as a primary diagnostic tool for identifying brain tumors, facilitating the differentiation of various brain lesions, including primary tumors such as gliomas, meningiomas, and pituitary tumors. Meningiomas, benign growths originating from the brain and spinal cord membranes, manifest as rounded masses with a distinctive bright rim, visible on the right side of the image [6], [7]. These tumors compress and displace nearby brain tissue. Gliomas, malignant tumors arising from glial cells that support and shield brain neurons, present as irregular masses with darker regions, evident on the left side of the image. They infiltrate and damage the surrounding brain tissue [6], [8]. Additionally, pituitary tumors, benign growths emerging from the pituitary gland at the brain's base, appear as rounded masses with bright centers, located in the lower part of the image. These tumors can disrupt hormonal balance and lead to vision problems [6], [7].

In recent years, Deep learning [9] has emerged as a solution in the field of medical imaging [10], [11], [12] with the examination of brain tumor MRI images [13], [14]. Deep learning methods like convolutional neural networks (CNN) have shown promise in tasks such as tumor detection [15], segmentation [16], [17], classification [18], and treatment response assessment in MRI images of the brain. Although CNN architecture has improved the accuracy of diagnosis of medical images but still improvement required in the field of brain tumor detection. One primary issue is the difficulty in obtaining abundant training samples. CNN architectures require millions of images for training, but most datasets in medical imaging consist of only thousands of images, which is a common problem [19]. While pre-trained models can address the issue of small datasets, they are less effective for brain tumor images due to the structural differences between brain images and the images used to train the pre-trained models. Consequently, the performance of brain tumor identification needs to be improved to handle the complexities of clinical situations. Another significant challenge in brain tumor recognition arises from the considerable intra-class variation and inter-class similarity within MRI images. Tumors within the same class, such as gliomas, frequently exhibit significant diversity in shape, size, and intensity, and may occupy different regions in an MRI image, sometimes appearing as a small section of the overall scan. These characteristics make it difficult to consistently locate and accurately identify tumors. Additionally, high inter-class similarity between certain types of tumors, such as glioma and meningioma, further complicates the ability to effectively distinguish between tumor types. The combined issues of high variability within tumor classes and subtle distinctions between classes make accurate classification particularly complex. Traditional feature extraction techniques often fail to capture these nuanced differences effectively. This situation highlights the necessity for advanced classification models and attention mechanisms that can adapt to the diverse presentations within each tumor type while discerning the subtle features that differentiate various tumor classes.

To address the complex challenges in brain tumor recognition, we propose the GGLA-NeXtE2NET model, which combines EfficientNetV2S and ConvNeXt architectures through a Dual-Branch Ensemble (DBE) and incorporates a novel Gated Global-Local Attention (GGLA) mechanism. EfficientNetV2S is known for its computational efficiency and scalability, which allow it to achieve high accuracy with fewer parameters. This is essential in medical applications where training data can be limited, as EfficientNetV2S optimizes feature extraction without requiring large-scale data. On the other hand, ConvNeXt complements this by providing strong representation learning with its advanced convolutional operations, which capture spatial details crucial for MRI image analysis. Additionally, the GGLA mechanism introduces an innovative approach to processing local and global patterns in brain tumor images. While standard attention methods typically emphasize either global or local feature extraction but struggle to balance the two effectively, especially in cases of high intra-class variation and inter-class similarity. Our GGLA mechanism introduces a novel gating strategy that dynamically balances global and local dependencies, allowing the network to adaptively focus on the most relevant features based on tumor-specific characteristics. This is critical for brain MRI analysis, where tumor features can vary widely across patients and even within the same tumor class. By capturing these differences more precisely, the GGLA module improves both classification accuracy and model interpretability, offering significant advancements over traditional attention mechanisms. This combination of dual-branch ensemble learning with adaptive gating distinguishes our approach, marking it as a novel and necessary contribution to the field of medical imaging. The summarized contribution of our research can be seen below:

- Considering the problem of brain tumor image variety, we propose the Dual-Branch Ensemble (DBE) network that leverages complementary receptive fields to capture

both fine-grained details and broader contextual patterns simultaneously compared to traditional models.

- The GGLA utilizes dedicated convolutional layers to generate query, key, and value matrices, followed by an innovative scaled dot-product attention that dynamically adjusts based on input characteristics to identify relevant features and tumor regions. This is complemented by a Local Attention module that further refines the attention weights through multi-scale processing, while an adaptive gating mechanism dynamically balances global and local information, ensuring optimal feature selection for accurate tumor detection despite limited training data.
- Our new denoising strategy combines multiple techniques, including patch-based denoising, wavelet decomposition, and selective noise reduction. This Strategy adeptly addresses challenges such as blurring, noise, and subtle contrasts, resulting in images that highlight abnormalities and structural details. The enhanced images are invaluable for both radiologists and deep learning models, as they facilitate more accurate diagnosis, treatment planning, and research.
- Finally, the Super-Resolution Generative Adversarial Network (ESRGAN) based data augmented technique further boosts up the model performance to avoid overfitting problems.

II. RELATED WORK

A. DEEP CONVOLUTIONAL NEURAL NETWORKS (DCNN) FOR BRAIN TUMOR CLASSIFICATION

Recently, CNN like deep learning models has been very helpful in classifying brain tumor images and provides excellent results than traditional machine learning methods. Huang et al. [20] presented the CNNBCN, an innovative approach utilizing randomly generated graph algorithms to optimize network structure with unique activation functions. This novel approach demonstrated impressive performance, with a classification accuracy of 95.49%, surpassing other models, and boasts lower test loss than established models like ResNet, DenseNet, and MobileNet. The authors, Rizwan et al. [21] developed a customized Gaussian Convolutional Neural Network (GCNN) approach for classifying brain tumors from non-invasive MRI scans. They applied various filters for image pre-processing and find the Gaussian filter provided the best performance. Using two MRI datasets, they achieve classification accuracies of 97.14% for tumor type and 99.8% for tumor grade. Wang et al. [22] introduced a modified version of the Vision transformer model and achieved 98.86% accuracy on the brain tumor dataset. Mehmood et al. [23] developed advanced techniques for automatic detection, colorization, and classification of tumor regions in MRI images to identify abnormalities early. Using T1-CE MRI datasets, methods like Pix2Pix Conditional GAN Networks (Pix2Pix-cGANs) were shown promising results and improved classification accuracy of 92.4% post-colorization with NASNet-Large. Both Anaya-Isaza et al. [24] and Reyes and Sánchez [25] used multiple pre-trained CNN models to examine the model's performance on the MRI dataset. İncir and Bozkurt [26] combined pretrained EfficientNetV2M and Inception-V3 networks to enhance the performance of individual models and achieved 98.41% overall accuracy. Furthermore, Khan et al. [27] and Patil and Kirange [28] combined pretrained VGG16 with the proposed CNN model to improve the decision-making power of both models. Moreover, integrating CNN architecture with machine learning models can enhance the performance of brain tumor image detection. For example, Senan et al. [29] and Haq et al. [30] implemented CNN models for feature extraction and SVM models for classification, demonstrating remarkable results on multi-class brain tumor datasets. According to Islam et al. [31] and Shanthi et al. [32], adding the LSTM layer at the end of the CNN model provided better results.

B. ATTENTION MECHANISMS

Recently, incorporating attention mechanisms into CNN architectures has yielded improved results in classifying brain tumor datasets and more efficiently localizing tumor areas compared to standalone CNN models. Alzahrani et al. [33], [34] and Pandi et al. [35] added the self-attention [36] mechanism in its models to classify the brain tumor images. Self-Attention mechanism is very popular due to its ability to capture global information and long-range dependencies in the image. On the other hand, CBAM [37] and BAM [38] like attention mechanisms focus on local information in the image. Tang et al. [39] used positional attention block with global attention mechanisms in the Spinal Convolutional Attention Network and achieved 99.28% overall accuracy on the MRI brain tumor dataset. Pacal et al. [40] integrated the channel and global attention mechanisms in the efficientnetV2 model to further enhance the model's abilities. According to Apostolopoulos et al. [41], spatial attention, and global average pooling-based attention mechanisms can provide better performance on MRI brain tumor datasets. Methods for obtaining global information, such as simple Global Average Pooling (GAP), often result in the loss of valuable information. This makes it challenging for the network to accurately identify images with high similarity and small tumor areas. Rasheed et al. [42] integrated both spatial and channel attention mechanisms in its models to classify the brain tumor images on the MRI dataset.

C. ENHANCED SUPER-RESOLUTION GENERATIVE ADVERSARIAL NETWORK (ESRGAN)

ESRGAN [43] is a deep architecture that builds upon the concept of Generative Adversarial Networks (GANs) and employs a generator-discriminator architecture to achieve its goal. It takes low-resolution images and produces high-resolution images. Both Alalwan et al. [44] and Yaqub et al. [45] used GAN architecture to balance the MRI dataset to further enhance the model ability. The mathematical presentation of GAN architecture is as follows:

The generator first takes a low-resolution image X_{LR} as input and aims to generate a high-resolution image X_{HR}. It can be represented as a deep neural network.

$$X_{HR} = G(X_{LR}; \theta_G) \tag{1}$$

where G represents the generator function, and θ_G are the parameters of the generator neural network. The discriminator's role is to distinguish between real high-resolution images X_{HR} and generated high-resolution images X_{HR}^{fake}. It provides feedback to the generator by indicating how realistic its generated images are.

$$D(X_{HR}; \theta_D) \rightarrow p_{real} \tag{2}$$

$$D(X_{HR}^{fake}; \theta_D) \rightarrow p_{fake} \tag{3}$$

Here D represents the discriminator function, θ_D are the parameters of the discriminator neural network, p_{real} is the probability that X_{HR} is a real high-resolution image, and p_{fake} is the probability that X_{HR}^{fake} is a generated high-resolution image. Perceptual loss measures the difference between the feature representations of the generated image and the ground truth image using a pre-trained feature extraction network F.

$$L_{perceptual} = \frac{1}{N} \sum_{i=1}^{N} \left|\left| F\left(X_{HR}\right) i - F\left(X_{HR}^{fake}\right) i \right|\right|_2^2 \tag{4}$$

Here N is the number of layers in the feature extraction network, and F represents the feature extraction network. The adversarial loss is based on the GAN framework, encouraging the generator to produce images that are difficult to distinguish from real images according to the discriminator.

$$L_{adversarial} = -log(D(X_{HR}^{fake}; \theta_D) \tag{5}$$

Content loss measures the pixel-wise difference between the generated high-resolution image and the ground truth high-resolution image.

$$Lcontent = \left|\left| X_{HR} - X_{HR}^{fake} \right|\right|_1 \tag{6}$$

The overall loss for training the generator is a combination of perceptual loss, adversarial loss, and content loss.

$$L_{total} = L_{perceptual} + \lambda_{adv} * L_{adversarial} + \lambda_{content} \tag{7}$$

where λ_{adv} and$\lambda_{content}$ are hyperparameters controlling the importance of adversarial loss and content loss. By iteratively updating the generator and discriminator using their respective loss functions, ESRGAN learns to generate high-resolution images with enhanced details and textures.

D. SCOPE OF DEVELOPMENT

The classification of brain tumor images using Deep Convolutional Neural Networks (DCNNs) has shown significant advancements. However, despite the impressive progress, several areas still require further exploration to improve the accuracy, efficiency, and generalizability of these models.

The self-attention mechanism can be effective at capturing global information within an image. However, its implementation in models can be computationally expensive and may lead to information loss when used within hybrid models. The integration of channel and spatial attention mechanisms in the model may not consistently capture global dependencies effectively, potentially overlooking crucial contextual information in brain tumor data. Furthermore, the challenge of balancing attention across multiple dimensions could lead to overfitting, particularly when the dataset is limited in size or diversity. The reviewed studies offer the same kind of architectures in ensemble modeling, which can constrain the model's ability to effectively handle brain tumor datasets that contain high-level information and complex patterns. Moreover, reviewed studies overlook the quality of MRI images and apply them directly to the model, which can limit the model's capacity to accurately classify brain tumor images. Lastly, the reviewed studies predominantly concentrate on enhancing accuracy, with limited attention given to the explainability and interpretability of the models. A notable gap exists in developing approaches that improve the transparency and comprehensibility of these complex models for clinical application.

III. MATERIALS AND METHODS

In our study, we systematically outlined the research framework. In Section A, we provided an in-depth exploration of the dataset, presenting comprehensive details. In Section B, we delved into the intricacies of denoising strategies, elucidating their significance in our approach. Subsequently, Section C expounded upon the application of the ESRGAN method, elaborating on its implementation. Lastly, Section D with its subsections was dedicated to the meticulous construction of our model, detailing each component and technique utilized.

A. MRI BRAIN TUMOR DATASET

For experimentation, openly accessible MRI Brain Tumor datasets were utilized. The first dataset was sourced from the Kaggle repository [46] and is referred to as the ''4-class MRI Brain tumor dataset.'' This dataset encompasses four distinct classes, specifically ''glioma,'' ''meningioma,'' ''pituitary tumor,'' and ''no tumor.'' Each class within this dataset comprises a substantial number of samples, as detailed in Table 1.

Additionally, another dataset [47] known as the '' 3-class MRI Brain tumor dataset '' was comprised of 3064 T1-weighted contrast MR images. These images involve 233 patients with ''meningioma'', ''glioma'', and ''pituitary tumors'', along with 73 patients exhibiting different grades of glioma. The images in this dataset possess a resolution of 512×512 pixels and are captured with a voxel spacing size of 0.49×0.49 mm^2. These images are oriented in axial (transverse plane), coronal (frontal plane), or sagittal (lateral plane) planes. There was an imbalance in the sizes of certain classes within Dataset 2 compared to others. To address this disparity, we employed the Enhanced Super-Resolution Generative Adversarial Network (ESRGAN) technique to enhance the

TABLE 1. Publicly available multi-class MRI brain tumor dataset.

Dataset	MRI Type	Classification Labels	Number of Images and split	ESRGAN generated images
Brain Tumor MRI Dataset (4-class dataset)	T1-weighted	Glioma tumor	1321/300 train/test	356
		Meningioma tumor	1339/306 train/test	355
		Pituitary tumor	1457/300 train/test	196
		No tumor	1595/405 train/test	-
			7023 total	807 total
Figshare (3-class MRI Brain Tumor Dataset)	T1-weighted	Glioma tumor	1426	-
		Meningioma tumor	708	696
		Pituitary tumor	930	476
			3064 total	1072 total

balance of 4-class and 3-class brain tumor datasets. ESRGAN was leveraged to generate novel images by utilizing pre-existing images, as outlined in section C of this paper. Further information regarding the distribution of samples for each class can be found in Table 1. The cropping function [48] was also employed to eliminate extraneous regions within the images. This function selectively retains the pertinent and meaningful content, contributing to the refinement of the dataset.

B. DENOISING IMAGE

MRI images can be susceptible to a variety of noise sources, potentially impacting image quality. Notably, patient motion, even subtle movements, during the MRI scan can introduce artifacts and compromise image precision [49]. This challenge is particularly critical in brain imaging, as motion-induced blurring can obscure the fine details essential for accurate tumor assessment. Another noise-related phenomenon, aliasing, arises when the field of view is insufficient to fully capture the signal, leading to wraparound artifacts [50]. In the context of brain tumor imaging, this can result in the misplacement of brain regions within the image. Furthermore, chemical shift artifacts may manifest at tissue interfaces, as is often the case in brain imaging, contributing to errors in the precise localization and characterization of brain tumors [51]. Additionally, radiofrequency (RF) interference stemming from external sources can introduce noise into the MRI signal, impacting image quality [52]. In a broader context, the presence of noise can even disrupt the accurate classification of distinct tumor types, as the distinguishing features between different tumor categories can be obscured or distorted.

To enhance the image quality, we employed a comprehensive suite of noise reduction techniques on both brain tumor datasets. These techniques encompassed a range of methodologies, each strategically selected to address distinct aspects of noise while preserving the critical features essential for accurate diagnosis. "Equation 1" computes the denoised pixel value for each pixel location by considering the weighted average of corresponding pixel values from similar patches [53] in the input MRI image:

$$D(x,y) = \frac{1}{C(x,y)} \sum_{i=1}^{M} w(x,y,x_i,y_i) \cdot I(x_i,y_i) \quad (8)$$

where D represents the denoised MRI image, $C(x, y)$ is the normalization factor (total weight) at the position (x, y), which ensures that the output pixel value is properly scaled, M is the total number of patches in the image. The weight Where D represents the denoised MRI image, $C(x, y)$ is the normalization factor (total weight) at the position (x, y), which ensures that the output pixel value is properly scaled, M is the total number of patches in the image. The weight $w(x, y, x_i, y_i)$ quantifies the similarity between the patch centered at (x_i, y_i) and the patch centered at (x, y). The weight is based on a Gaussian function that takes into account the differences in pixel values within the patches [54]:

$$w(x,y,x_i,y_i) = exp\left(-\frac{\sum_{p,q} h(p,q)\cdot(I(x+p,y+q)-I(xi+p,yi+q))^2}{h^2}\right) \quad (9)$$

where h is a parameter that controls the decay of similarity with increasing pixel distance within the patches and (p, q) iterate over the pixels within the patch centered at (x, y).

The "Equation 10" decomposes the denoising image D into different scales (or levels) of detail. Each level of detail captures information about different frequency components of the image [55]. This is particularly useful in medical imaging, where certain structures (such as tumors) may appear at different scales.

$$W(a,b) = \int D(x) \cdot \psi_{a,b}(x)dx \quad (10)$$

Here, $W(a, b)$ represents the wavelet coefficients obtained by convolving the image D with the wavelet function $\psi_{a,b}$, which is scaled by a and translated by b. Thresholding [56] is a critical step in denoising. It involves setting small coefficients to zero to remove noise while retaining important features. This helps in reducing random noise present in the image, making the tumor boundaries and structures more distinguishable:

$$\tilde{W}(a,b) = \begin{cases} W(a,b), & if\ W(a,b) \geq \lambda\sigma \\ 0, & otherwise \end{cases} \quad (11)$$

Here, $\tilde{W}(a, b)$ are the thresholded wavelet coefficients, λ is the threshold parameter and σ is an estimate of the noise standard deviation. After thresholding the coefficients, the denoised image D′ is reconstructed from the denoised wavelet coefficients $\tilde{W}(a, b)$:

$$D' = \sum_{a,b} \tilde{W}(a, b) \cdot \psi^*_{a,b}(x) \tag{12}$$

The Equation 13 further refines the denoised image D′ by using statistical properties of the noise and the signal [57]:

$$S = F^{-1}\left\{\frac{H^*(f)}{|H(f)|^2} \cdot F\left\{D'\right\}(f)\right\} \tag{13}$$

where S represents the noise-free estimated original image, F denotes the Fourier transform, and F^{-1} is the inverse Fourier transform. The term H(f) is the frequency response of the system in the frequency domain and $H^*(f)$ is the complex conjugate of H(f). The term $F\{D'\}(f)$ is the Fourier transform of the wavelet-denoised image.

"Equation 14" encompasses the enhancement of local contrast [58] across various regions within the image, with a specific focus on accentuating the visibility of both the tumor structure and brain tissues. This enhancement takes into account the noise characteristics that were mitigated through the application of "Equation 13".

$$\acute{I}(x, y) = T\left[S_{input}(x, y)\right] \tag{14}$$

The operator $(T[\cdot])$ involves several essential steps that simulate the adaptive histogram equalization process. These steps are integrated into the unified equation as follows:

$$T[S_{input}(x, y)] = \frac{L-1}{\beta^2} \sum_{i=x-\frac{\beta}{2}}^{x+\frac{\beta}{2}} \sum_{j=y-\frac{\beta}{2}}^{y+\frac{\beta}{2}} H\left(S_{input}(i, j)\right) \cdot S_{input}(x, y) \tag{15}$$

where β refers to the size of the local neighborhood for histogram analysis, L is the number of intensity levels, and H is the histogram function. Finally, "Equation 16" further accentuates the enhanced contrast and fine details, potentially making features such as edges and textures [59] more pronounced in the brain tumor MRI image.

$$E = I + \gamma N\left(L\left(\acute{I}\right)\right) \tag{16}$$

where E is the final enhanced image, γ is the weighting parameter that controls the strength of the edge enhancement effect, and $N(L(X))$ is the normalized function that is computed as:

$$N\left(L\left(\acute{I}\right)\right) = \alpha \frac{L\left(\acute{I}\right) - \min\left(L\left(\acute{I}\right)\right)}{\max\left(L\left(\acute{I}\right)\right) - \min\left(L\left(\acute{I}\right)\right)} + \beta \tag{17}$$

where α controls the contrast of the normalized image, β controls the brightness of the normalized image.

According to Figure 1, the similarity between denoised curve 1 and the original curve at the beginning, indicates the "Equation 8" denoising effectively retained the overall structure of the image. The denoising curve 1 slightly dips below the original curve in the mid-section. This could be due to the denoising process removing some noise components, which might have inadvertently affected certain features in that region. After the mid-section denoising curve 1 rises slightly above the original curve. This could be attributed to the denoising algorithm enhancing certain details that were previously buried in noise. In the end, both denoising curve 1 and the original curve combined them. The initial segment of denoising curve 2 starts with denoising curve 1 and the original curve. This indicates that "Equation 8" retained the original image characteristics while reducing noise. Before starting the mid-section denoising curve 2 slightly higher than both denoised curve 1 and the original curve that indicates "Equation 12" further enhances fine details and structures that may have been subtly lost during previous denoising. The curve's increase indicates a more pronounced contrast enhancement. The initial segment of the denoising curve 3 closely resembles the previous denoising image intensity distribution. This is expected because "Equation 13" is designed to preserve the existing image features while removing noise and blurring artifacts. Hence, it aims to maintain the overall structure of the previous denoising image. After the Initial Segment the denoising curve 3 rises above the previous curves. This can be attributed to the "Equation 13 "amplification effect on certain pixel intensity values. The denoising curve 3 enhances areas with low signal-to-noise ratios, which could correspond to low-contrast regions in the image. As a result, these regions experience a relative intensity boost compared to the original. The initial segment of the denoising curve 4, which aligns with the previous image curves that indicate "Equation 13" successfully preserved the initial image structure, and "Equation 15" has not yet come into effect. The sharp decrease in the denoising curve 4 after the initial segment that shows "Equation 15" enhances the local regions, particularly those with relatively lower contrast, it results in a more pronounced contrast distribution, causing the curve to decrease sharply. Unlike previous curves, the denoising curve 5 starts not at the beginning but somewhere in the middle of the plot. This initial positioning is indicative of the "Equation 24" influence on the image. "Equation 16" is known for its emphasis on high-frequency components, including edges and fine details. Consequently, the curve commences at an intermediate point, reflecting the enhanced details and edges. A notable characteristic of this curve is its initial elevation above all previously observed curves. This elevation signifies denoising curve 5 impact in enhancing edges and fine features. By accentuating regions with high gradients, "Equation 16" elevates the pixel intensities, which is particularly evident in areas with pronounced edges. The result is an enhanced contrast for these edge-rich portions of the image. However, as the curve progresses, a marked decline becomes evident. This sharp descent is a result of a trade-off inherent to the "Equation 17" operation.

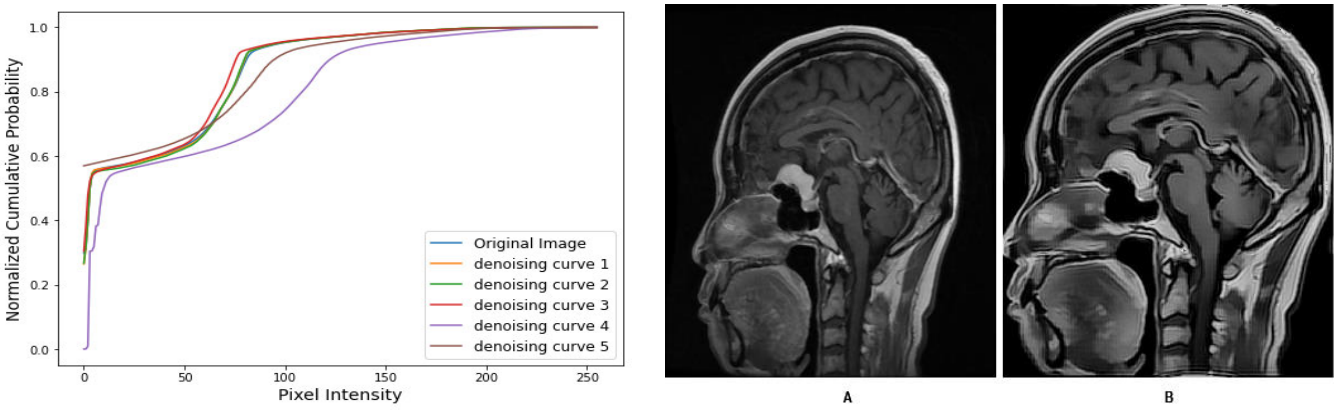

FIGURE 1. First image illustrates pixel intensity levels after applying the each denoising strategy, while second image depicts two parts: (A) the original image and (B) the denoised image obtained by applying equations 8 to 17 along with the cropping function.

C. DATA AUGMENTATION WITH GAN

Data augmentation is crucial in medical imaging tasks like MRI brain tumor classification because it helps mitigate the issue of limited datasets by generating additional training samples. The impact of data augmentation can be substantial, and using GANs (Generative Adversarial Networks) for this purpose offers several advantages over traditional augmentation methods. First, GAN can produce highly accurate and varied artificial images that closely mimic real MRI scans, potentially enhancing model generalization better than traditional augmentation techniques. Secondly, Using GAN-augmented data in training can make models more robust to variations and artifacts in real-world data, leading to improved performance on new, unseen data. Lastly, GAN can capture and replicate complex features and patterns in MRI scans that traditional methods might overlook.

According to Figure 2, the generator network starts by taking the input image $I \in R^{HXWXC}$. It first applies a convolutional layer with 64 filters and a kernel size of 9×9, followed by a ReLU activation function. This produces an intermediate tensor x_1. This x_1 is then fed into a residual dense block, which consists of multiple layers, each performing convolution, batch normalization, and ReLU activation. For each layer l, the output is concatenated with its input to form dense connections. After processing through these layers, the tensorx_{L+1} is obtained, where L is the number of layers in the residual dense block. Following the dense block, the tensor x_{L+1} goes through additional convolutional layers with 128 filters and a kernel size of 3×3, each followed by batch normalization and ReLU activation and produces the tenors x_2 and x_3. An upsampling operation with a factor of 2×2 is applied to x_3 to produce x_4, which is then passed through a final convolutional layer with 3 filters and a kernel size of 9×9, using a tanh activation function to generate the output image O.

On the other hand, the discriminator network processes an input image $I \in R^{HXWXC}$ and aims to output a probability p indicating whether the image is real or generated. It starts with a convolutional layer that uses 64 filters and a kernel size of 3×3 with LeakyReLU activation. This initial step produces tensor y_1. The y_1 is then passed through a series of additional convolutional layers with increasing filter sizes, where each layer is followed by batch normalization and LeakyReLU activation. Specifically, y_1 is downsampled through strides of 2×2 in the subsequent convolutional layers, resulting in tensors y_2, y_3 and y_4. Each of these tensors is processed similarly, with normalization and activation steps applied. After these layers, the y_4 is then fed into a dense layer with a sigmoid activation function, producing a final probability p that indicates the likelihood of the input image being real. We used VGG19 model to compute the perceptual loss on augmented images. Then, we modify the equation 4 as:

$$L_{perceptual} = \frac{1}{N}\sum_{i=1}^{N}\left|\left|F\left(Aug(X_{HR})\right)\_i - F\left(Aug(X_{HR}^{fake})\right)\_i\right|\right|_2^2 \tag{18}$$

Here N is the number of layers in the VGG19 network, and F represents the feature extraction network Aug denotes the data augmentation function.

D. PROPOSED MODEL

We have proposed the GGLA-NeXtE2NET that has spatially designed for multi-class brain tumor datasets. In the proposed model, the GGLA block improves the model capability by using global and local attention. Furthermore, we have introduced an FE-GGLA block that refines the features in MRI images and enhances the classification accuracy. Lastly, our Dual-Branch Ensemble (DBE) network is designed to extract features from MRI images with different receptive fields. One branch is used fixed-size receptive field for feature extraction and fusion. On the other hand, the second branch uses different receptive fields to perform the same task and improve recognition accuracy with different conditions. Fig 3. Illustrates the architecture of the GGLA-NeXtE2NET. The individual detail of each module is described below.

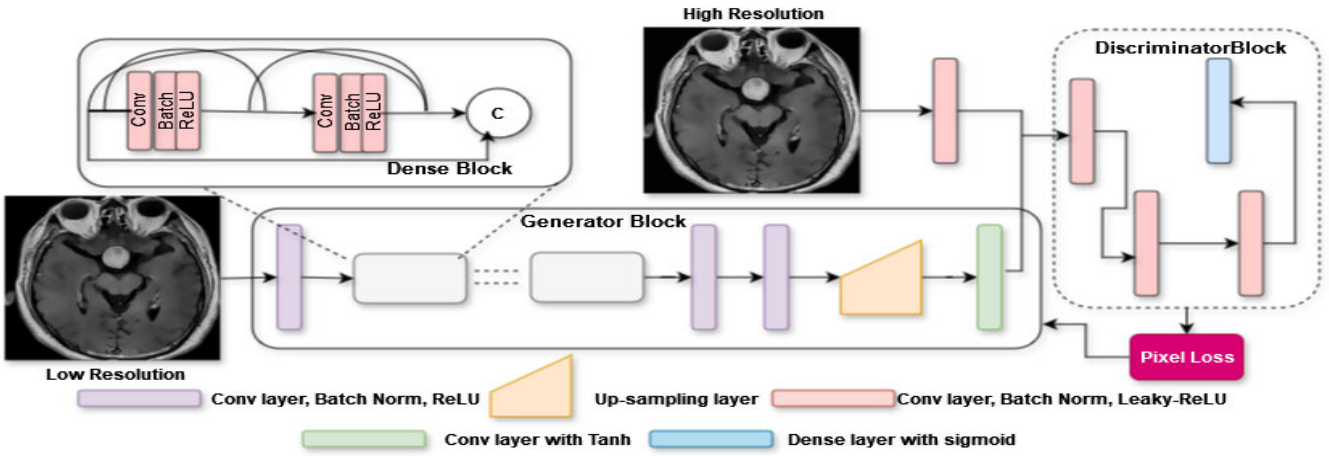

FIGURE 2. Complete overview of ESRGAN model.

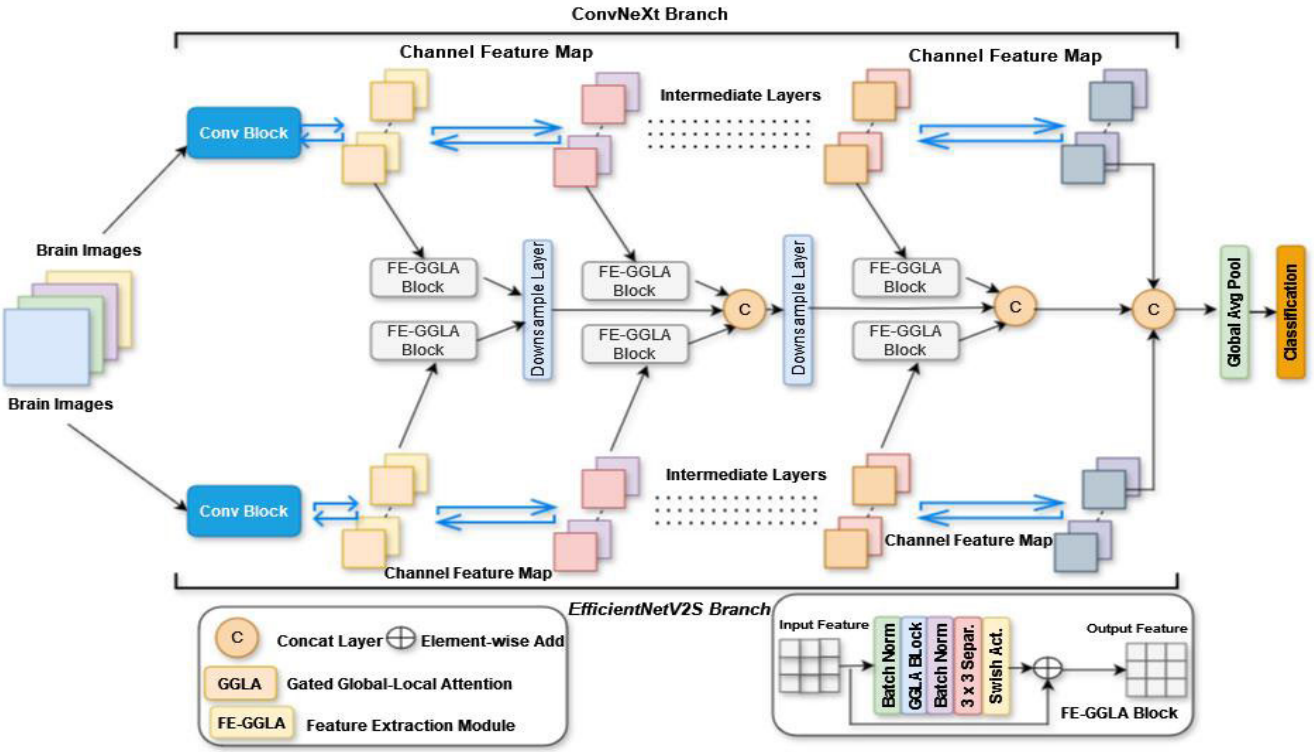

FIGURE 3. Complete overview of proposed GGLA-NeXtE2NET network.

1) DUAL-BRANCH ENSEMBLE NETWORK

A single CNN architecture might not be ideal for classifying brain tumor images because these images often have complex patterns and similar types of tumor lesions, making classification challenging. To address this challenge, we introduce a Dual-Branch Ensemble (DBE) network that consists of EfficientNetV2S [60] and ConvNeXt [61] models. In the DBE network, EfficientNetV2S utilizes a fixed 3×3 receptive field to extract features. Conversely, the ConvNeXt model employs receptive fields ranging from a minimum of 3×3 to a maximum of 27×27 to extract features in different ways. Both primary branches are trained simultaneously and in parallel. We then introduce a parameter-sharing scheme, wherein the extracted feature values from the two main branches at different levels are passed to the FE-GGLA block for further refinement. The specific values of parameters in the parameter-sharing scheme are determined by the chosen backbone network. This approach enhances the model's compatibility with a multi-class brain tumor dataset.

2) FE-GGLA MODULE FOR FEATURE EXTRACTION

The classification of brain tumor images is still difficult due to both intra-class differences and inter-class similarities in multi-class brain tumor images. Due to the slight differences in shape, texture, and color in various MRI images, it's crucial to thoroughly extract global features. Additionally, there's significant variation among patients within each type of brain tumor, often in specific areas of the lesions. Therefore, extracting local information is crucial.

Conventional attention localization methods often prioritize either global or local information, failing to leverage the synergies between the two. In response to these limitations, we draw upon the concept of the Global Local Attention (GLA) module [62] and enhance it to introduce the Gated Global Local Attention (GGLA) module. In the GGLA module, the global branch extracts broader contextual patterns, while the local branch focuses on fine-grained tumor details, such as texture and border irregularities that are essential for tumor type differentiation. For example, while gliomas and meningiomas may present overlapping intensities, subtle distinctions in texture and shape provide crucial classification cues. The GGLA module not only amalgamates global and local information for feature extraction but also incorporates novel gated layers that enable the network to selectively enhance or suppress specific features acquired at different processing stages. This dynamic balancing of global and local features allows the model to prioritize the most relevant information for each tumor class, effectively addressing the challenges posed by intra-class variability, inter-class similarities, and dynamic lesion localization in multi-class brain tumor images.

a: GATED-HORIZONTAL ATTENTION

The proposed Gated-Horizontal Attention (GHA) module serves as a core element within the GGLA framework, specifically engineered to capture horizontal dependencies pertinent to query points. To accommodate small-sample learning scenarios, the GHA module is meticulously crafted to prioritize computational efficiency across four key dimensions. First, the input feature maps undergo dimensionality reduction and subsequent expansion to capture horizontal dependency information efficiently. Specifically, horizontal dependence information is extracted following the initial dimension reduction. Second, the resulting feature map is downscaled to half of its original size, reducing the number of pixels associated with each query point to enhance computational efficiency. Next, the scaled dot-product mechanism is included to enhance the HA module abilities. Furthermore, information is only obtained from points sharing the same row index as the query point. Finally, to augment the module's capabilities, a gated layer is introduced at the conclusion of the GHA module. This gated layer acts as a regulatory mechanism, allowing for selective modulation of information flow within the network.

As shown in Fig 4., the GHA module begins by applying layer normalization to the input tensor X, denoted as $X_{norm} = LayerNorm(X)$. Here, both X and X_{norm} share identical dimensions (N, H, W, C). The normalization is succeeded by convolutional operations that produce the query Q, key K, and value V tensors. These tensors are then subjected to a series of operations to implement the attention mechanism. Both the query convolution Q and the key convolution K are generated with a reduced number of filters C/r, where r denotes the reduction ratio. The value convolution V maintains the original channel dimension C. Following the convolutional operations, Q undergoes mean pooling across the spatial dimensions, resulting in a tensor with dimensions $N \times \left(\frac{C}{r}\right)$. The mean pooling operation for Q is mathematically represented as:

$$Q \in R^{N\times(\frac{C}{r})} = \frac{1}{H \times W} \sum_{i=1}^{H-1} \sum_{j=1}^{W-1} QR^{N\times H\times W\times \frac{C}{r}} \tag{19}$$

Next, the scaled dot-product attention mechanism is applied. The affinity matrix A1 is computed by performing a dot product between Q and the transpose of K, followed by scaling with the square root of the dimensionality of K. The softmax function is then applied to obtain the attention weights.

$$A_1 \in R^{N\times\left(\frac{C}{r}\right)\times\left(\frac{C}{r}\right)} = \frac{\exp\left(affinity_{n,i,j}\right)}{\sum_{l=1}^{H}\sum_{m=1}^{W}\exp\left(affinity_{n,l,m}\right)} \tag{20}$$

Meanwhile, the value tensor V is subject to average pooling along the height dimension, followed by transposition to prepare for element-wise multiplication with the attention weights:

$$E \in R^{N\times H\times(\frac{C}{r})} = \frac{1}{H}\sum_{j=1}^{H-1} V_j \tag{21}$$

Subsequently, the attention feature tensor F is derived, where each feature is a weighted combination of values based on attention weights:

$$F \in R^{(\frac{C}{r})} = \sum_{j=1}^{\frac{C}{r}-1} A_1 * E_j \tag{22}$$

To reintegrate F into the original spatial dimensions of the input tensor x, F is replicated along the reduction dimension and then resized to match the height and width of x. The resulting tensor, denoted as Y_1, is generated by combining the original input tensor x with the adjusted attention features G_1 using a learnable parameter α. This operation is expressed as:

$$Y_1 = \alpha \odot \mathrm{G}_1 + X \tag{23}$$

Additionally, a gating mechanism is applied to modulate the information flow within Y_1, governed by a sigmoid activation function (σ). Finally, the gated output is obtained by element-wise multiplication between the gate tensor and Y_1, and the dimensions remain consistent with Y_1:

$$gate_1 = \sigma\left(\sum_{i=0}^{C-1} Y_1 * W_{gate_{c,i}}\right) \tag{24}$$

$$Output_{gated} = gae_1 \odot Y_1 \tag{25}$$

b: GATED-VERTICAL ATTENTION (GVA)

Our GVA module is most likely to GHA module but focuses on capturing vertical relationships within the input feature map that can be seen in Figure 5. After convolutional transformations, Q and K undergo mean pooling to reduce their spatial dimensions. Then, the attention mechanism is implemented by computing the affinity matrix A2 through matrix multiplication between Q and K using the dot product. Subsequently, softmax activation is applied along the last dimension to obtain normalized attention weights. The affinity operation is computed as:

$$affinity = \frac{1}{scaling\_factor} \sum_{k=1}^{\frac{C}{r}-1} Q_{n,k} * K \tag{26}$$

Simultaneously, the value tensor V undergoes average pooling along the width dimension, followed by transposition in preparation for element-wise multiplication with the attention weights:

$$E \in R^{N \times H \times (\frac{C}{r})} = \frac{1}{W} \sum_{j=1}^{W-1} V_{J,:} \tag{27}$$

This step yields the attention feature tensor F, where each feature represents a weighted combination of values based on attention weights. To reintegrate F into the original spatial dimensions of the input tensor x, F is replicated along the reduction dimension and then resized to match the height and width of x. The resulting tensor, denoted as Y_2, is generated by combining the original input tensor x with the adjusted attention features G_2 using a learnable parameter β. This operation is expressed as:

$$Y_2 \in R^{N \times H \times W \times C} = \beta \odot \mathrm{G}_2 + X \tag{28}$$

Furthermore, a second gating mechanism is applied to modulate the information flow within Y_2, governed by a sigmoid activation function σ:

$$gate_2 = \sigma(\sum_{i=0}^{C-1} Y_{2:} * W_{gate_{c,i}}) \tag{29}$$

$$output_{gated2} = gate_2 \odot Y_2 \tag{30}$$

c: GATED HORIZONTAL VERTICAL ATTENTION (GHVA)

Although the Gated-Horizontal Attention (GHA) module or Gated-Vertical Attention (GVA) module can effectively capture long-distance dependency information in various directions, they inherently focus on acquiring local information. This limitation arises from the inability to establish connections between pixel points along rows or columns. To address this issue, we propose the Hybrid Gated Horizontal-Vertical Attention (GHVA) module, which integrates both GHA and GVA modules sequentially as can be seen in Figure 6 and Algorithm 1.

d: GATED-LOCAL ATTENTION (GLA)

To extract local spatial information from brain images, we introduced the Gated-Local Attention Module. It attains this by applying spatial attention mechanisms, which focus on specific regions of interest within the brain tumor image.

According to Figure 7, the GLA module first applies two distinguish max pooling (M) and average pooling (A) operations over the channels of the input tensor $I \in R^{HXWXC}$. A and M are then stacked together along a new channel dimension and creating a combined tensor $S \in R^{HXWX2}$. Next, S undergoes a series of convolutional layers with a swish activation function to create a transformed tensor T. From T, the module calculates attention weights W using a softmax-based transformation. In parallel, the module computes gate values G from T using a sigmoid-based transformation. The final step involves calculating the gated attention weights W_g by performing element-wise multiplication of the attention weights W and the gate values G. Finally, the original input tensor I is multiplied element-wise with the gated attention weights W_g producing the output tensor $O \in R^{HXWXC}$.

e: FE-GGLA MODULE

We further embed the Gated Global-Local Attention (GGLA) within the Feature-Enhanced (FE) FE-GGLA module to better address inter-class and intra-class variability in the brain tumor dataset which can be seen in Figure 3. The FE-GGLA module initially takes inputs from two EfficientNetV2-S and ConvNeXt branches, independently. These inputs are passed through a batch normalization layer. The batch-normalized outputs are then fed into the GGLA module to extract global and local features. Subsequently, the FE-GGLA module processes the features through another batch normalization layer, followed by a separable convolution layer with a 3×3 kernel and a Swish activation function. Finally, a residual connection is added using a 1×1 convolution layer. The embedded GLA scheme can accurately locate brain tumors and focus on comprehensive information, such as tumor boundaries and contours. This results in more enriched and accurate extracted features.

3) PROPOSED GGLA-NeXtE2NET

In this paper, we present the GGLA-NeXtE2NET model tailored for the classification of multi-class brain tumor images. Our proposed model, as depicted in Fig. 3, offers a comprehensive design that not only enhances model accuracy but also maintains consistency with the parameters of both EfficientNetV2 and ConvNeXt models. At the heart of our architecture lies the Gated Global Local Attention (GGLA) mechanism, which facilitates the sequential extraction of global and local features. Within the GGLA module, we initiate the Gated Horizontal and Vertical Attention (GHVA) to capture global features inherent in brain tumor images. Subsequently, the Gated Spatial Attention (GSA) is employed to focus on tumor-specific information locally. Initially, we extract fixed layers from the blocks of the DBE network and feed them into our FE-GGLA module. Following this, we introduce a separable convolutional layer with a 1×1 kernel size and a stride rate of 2, through which the FE-GGLA layer is processed which enables it to capture spatial information effectively while down-sampling the feature maps. Similarly, we repeat this procedure with the second layer from

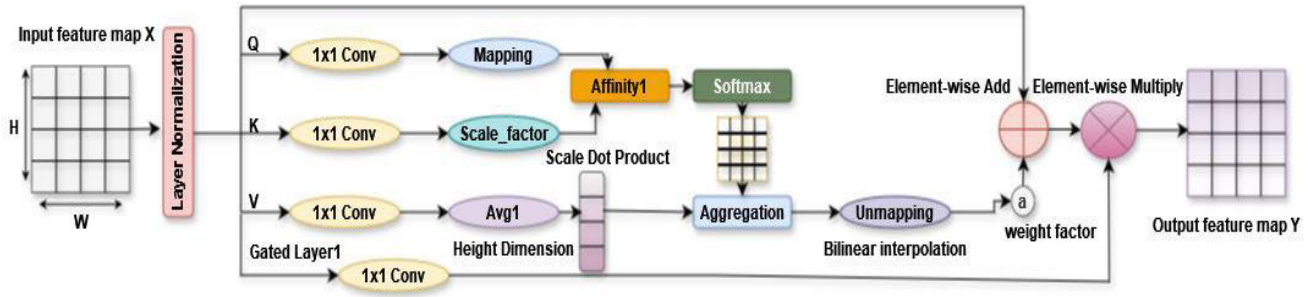

FIGURE 4. Structure of GHA module.

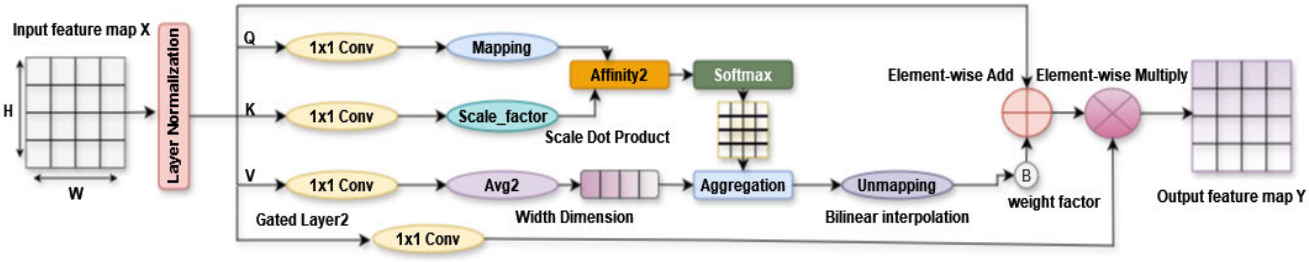

FIGURE 5. Structure of GVA module.

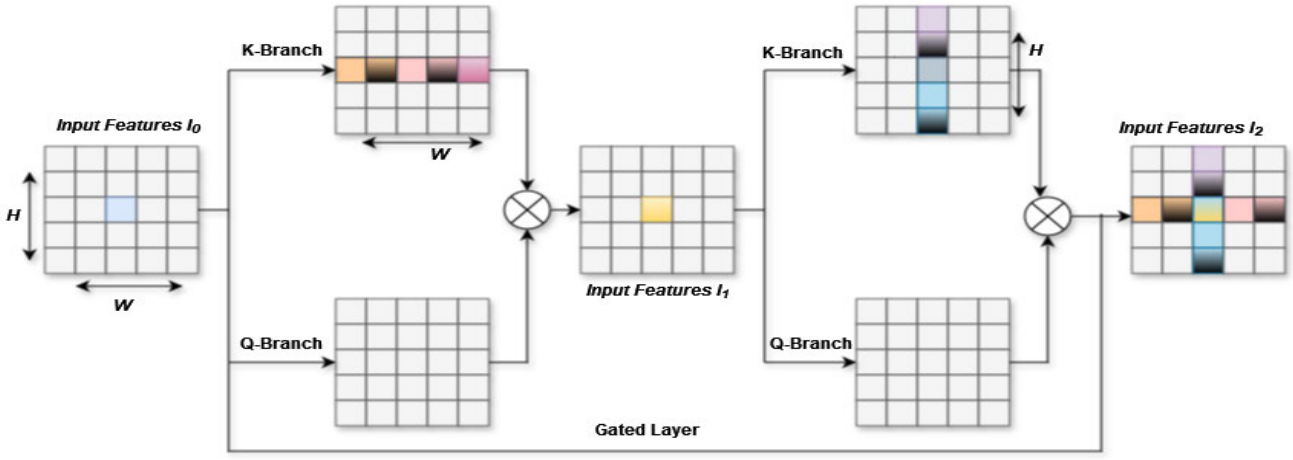

FIGURE 6. Structure of GHVA module. For each pixel in I_0 (for example, a blue pixel), the GHA module generates an attention map with WW weights. Similarly, the GVA module creates an attention map with H weights. By sequentially passing through the GHA and GVA modules, each pixel in I2 can indirectly acquire dependency information from all other pixels. To simplify the explanation, steps like the Value branches and other operations have been omitted.

the second block of the BBE network, passing it through the FE-GGLA module. Subsequently, we aggregate the information from the first and second layers using the concatenation function. This step ensures that the features extracted from different blocks of the DBE network, enhanced by the GGLA mechanism and refined by the separable convolutional layer, are effectively combined to enrich the overall feature representation. This approach is iterated for each block of the EfficientNetV2S and ConvConXt models, ensuring the comprehensive integration of the FE-GGLA mechanism across the architecture.

IV. EXPERIMENTS

A. EXPERIMENT SETTING

The Multi-Class Brain Tumor dataset was partitioned into three sets: a training set, a validation set, and a test set. The training set contained 80% of the images, the validation set contained 10% of the images, and the test set contained 10% of the images. The learning rate was set to 0.001, with 20 epochs and a batch size of 8. The learning rate was reduced by a factor of 0.5 with a patience value of 1. We used the categorical cross-entropy loss function, which is well-suited for multi-class classification tasks as it penalizes

Algorithm 1 Gated Horizontal-Vertical Attention (GHVA) Module

Input:
I0: Input feature map
I1, I2: Intermediate feature maps
Output:
I2: Final feature map with horizontal and vertical dependencies
1. Operation:
a. GHA Module:
- Input: I0
- Output: I1
- I1 = GatedLayer(HA(I0))

b. GVA Module:
- Input: I1
- Output: I2
- I2 = GatedLayer(VA(I1))

2. Information Propagation:
-Functions f_1 and f_2 define the information propagation between spatial positions of feature maps.
3. Definitions:
-Let $q = (q_x, q_y) \in R^{H/2\times W/2}$ be any spatial position of the feature map mapped by I_0.
-$A' \in R^{W\times\left(\frac{H}{2}\times\frac{W}{2}\right)}$ and $A'' \in R^{H\times\left(\frac{H}{2}\times\frac{W}{2}\right)}$ are the attention maps from the GHA and GVA modules, respectively.
4. Horizontal Dependency Information (GHA Module):
-Operation f_1:
For $\forall i \in R^W, A'_{i,q} = f_1(A', k_x^{GHA}, k_y^{GHA}, q_x, q_y)$,
where $k^{GHA}(k_x^{GHA}, k_y^{GHA}) \in R^W$ is a position in the same row as query point q in the horizontal structure, while Function f_1 multiplies query point q with the corresponding horizontal structure k^{GHA}, mapping the results to elements in A'.
5. Vertical Dependency Information (GVA Module):
-Operation f_2:
For $\forall j \in R^H, A''_{j,q} = f_2(A'', k_x^{GVA}, k_y^{GVA}, q_x, q_y)$,
where $k^{GVA}(k_x^{GVA}, k_y^{GVA}) \in \mathbb{R}^H$ is a position in the same column as query point q in the vertical structure. While, Function f_2 multiplies query point q with the corresponding vertical structure k^{GVA}, mapping the results to elements in A''.
6. Gated Layer:
-GatedLayer(I) applies a gating mechanism to selectively propagate information in the input tensor I.
7. Return I2 as the output of the GHVA module.

incorrect predictions based on their confidence scores. This helps ensure precise classification by encouraging the model to produce accurate probability distributions for each class. We train the model using the Adamax optimizer, which builds on the Adam algorithm by using an infinity norm, making it robust for handling sparse gradients and allowing for stable convergence in high-dimensional feature spaces. The best results were saved in the model using the validation set. The TensorFlow platform was used to perform all experiments with Python version 3.10. The multiple metrics were used to measure the model performance. These metrics are defined in equations 31 to 34.

$$\text{Accuracy} = \frac{1}{n}\sum_{i=1}^{n} \mathrm{I}(y_i = \hat{y}_i) \tag{31}$$

$$Precision = \frac{\sum_{i=1}^{n} I(y_i = 1 \; and \; \hat{y}_i = 1)}{\sum_{i=1}^{n} I(\hat{y}_i = 1)} \tag{32}$$

$$Recall = \frac{\sum_{i=1}^{n} I(y_i = 1 \; and \; \hat{y}_i = 1)}{\sum_{i=1}^{n} I(y_i = 1)} \tag{33}$$

$$\text{F1} - \text{score} = \frac{2}{\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}} \tag{34}$$

where I(x) is the indicator function that returns 1 if x is true and 0 otherwise. y_i be the actual label of the i-th observation, and $\hat{y}_i$ be the predicted label of the i-th observation, and N be the total number of observations.

B. RESULTS

Table 2 presents the results of the GGLA-NeXtE2NET model on two different brain tumor datasets. For the 4-class Brain tumor dataset, the model achieved an accuracy of 98.98% for Glioma, with a precision of 99.49%, recall of 98.99%, and an F1-Score of 99.24%. Meningioma was classified with perfect accuracy, recall, and F1-Score of 100%, and a precision of 99.50%. The No Tumor class also had perfect scores across all metrics. Pituitary tumors were classified with an accuracy of 99.48%, precision of 99.49%, recall of 99.49%, and an F1-Score of 99.49%. On the 3-class Figshare Brain tumor dataset, the model's

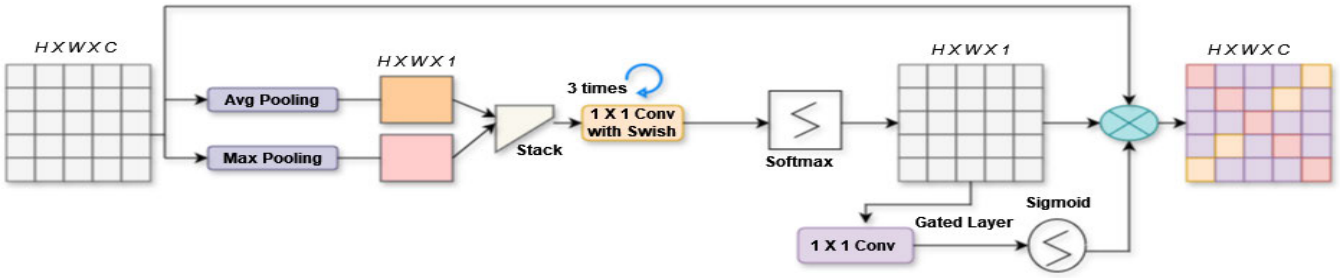

FIGURE 7. Structure of GLA module.

Algorithm 2 GGLA-NeXtE2NET

Input: Image X
Output: Predicted class probabilities
For each image X
Extract feature maps $F_{eff_1}, F_{eff_2}, F_{eff_3}, F_{eff_4}, F_{eff_5}, F_{eff_6}$, $F_{eff_{out}}$ using EfficientNetV2S.
Extract feature maps $F_{Conv_1}, F_{Conv_2}, F_{Conv_3}, F_{Conv_5}, F_{Conv_{out}}$ using ConvNeXt.
Apply attention mechanisms:
For i = 1 to 6:
Apply FE-GGLA (H_i) on F_{eff_i}.
Apply FE-GGLA(H_m) on F_{Conv_i}.
Apply separable convolution layers:
For i = 1 to 5:
If i = 1:
Apply sep_conv_1 on H_i.
Apply sep_conv_2 on H_m.
Else:
Let $H_{concat} = \sum_{j=1}^{i-1} \sum_{n=1}^{m-1} (S_j, S_n)$ // Concatenation of feature maps from FE-GGLA mechanisms.
Apply sep_conv_i on H_{concat}.
Let $F_{concat} = \sum_{i=1}^{5} \sum_{m=1}^{5} (H_i, H_m)$ // Concatenation of all H_i and H_m.
Concatenate $F_{Conv_{out}}$ and $F_{eff_{out}}$ using Concatenate Layer (β) :
Let $F_{final} = [F_{concat}; F_{Conv_{out}} + F_{eff_{out}}]$// Concatenation of F_{concat} and $F_{Conv_{out}}$ with $F_{eff_{out}}$.
Apply global average pooling (α) on F_{final}:
Let $\alpha = 1/(H * W)$ // Normalization factor where H is the height and W is the width of F_{final}.
Let $g(F_{final}) = \sum_{i=1}^{H} \sum_{j=1}^{W} F_{final}[i, j] * \alpha$ // Global average pooling operation.
Apply the classifier layer (α) for final classification.

performance was slightly different. Glioma was classified with an accuracy of 98.60%, precision of 99.30%, recall of 98.60%, and an F1-Score of 98.95%. Meningioma had an accuracy of 99.28%, precision of 97.89%, recall of 99.29%, and an F1-Score of 98.58%. Pituitary tumors were classified with an accuracy of 99.29%, a perfect precision of 100%, a recall of 99.29%, and an F1-Score of 99.64%. The GGLA-NeXtE2NET network impressive performance on key classification metrics like accuracy, precision, recall and F1-score highlight its ability to learn robust discriminative features from the MRI data. This enables highly accurate discrimination between complex tumor types as well as rejection of normal images. The model shows promise for real-world clinical usage for automated screening and diagnosis of brain tumors.

"Figures 8 and 9" illustrate the confusion matrix and ROC-AUC curve for the novel GGLA-NeXtE2NET network on 4-class and 3-class brain tumor Datasets respectively. The confusion matrix shows the counts of true positive, true negative, false positive, and false negative predictions. The matrix is divided into rows and columns, where each row represents the instances of an actual class, and each column represents the instances of a predicted class. The confusion matrix indicates exceptional performance on 4-class brain tumor dataset in figure 8(a), with only 2 misclassified cases for the glioma and 1 misclassification for the pituitary class. The meningioma and no tumor classes achieved perfect results, with no misclassifications. For the 3-class brain tumor dataset in Figure 9(a), our GGLA-NeXtE2NET network correctly predicted 141 cases and misclassified 2 cases as meningioma.

TABLE 2. Both brain tumor datasets result with GGLA-NeXtE2NET.

Dataset	Classes	Accuracy	Precision	Recall	F1-Score	Dataset	Classes	Accuracy	Precision	Recall	F1-Score
4-class Brain tumor dataset	Glioma	98.98%	99.49%	98.99%	99.24%	3-class Brain tumor dataset	Glioma	98.60%	99.30%	98.60%	98.95%
	Meningioma	100%	99.50%	100%	99.75%		Meningioma	99.28%	97.89%	99.29%	98.58%
	No Tumor	100%	100%	100%	100%		Pituitary	99.29%	100%	99.29%	99.64%
	Pituitary	99.48%	99.49%	99.49%	99.49%		-	-	-	-	-

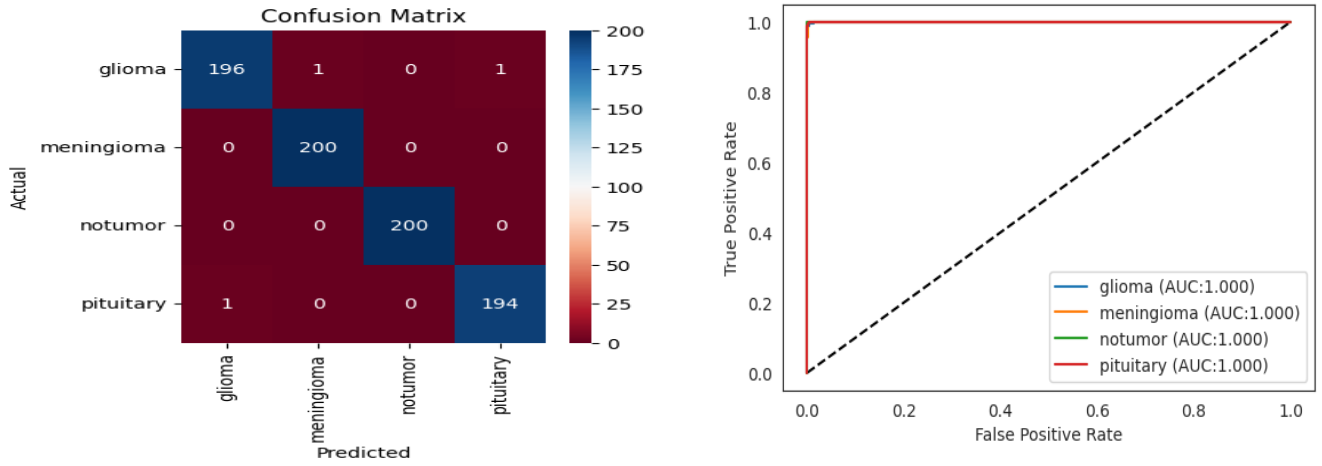

FIGURE 8. GGLA-NeXtE2NET results on 4-class brain tumor dataset (A) Confusion matrix (B) ROC-AUC curve.

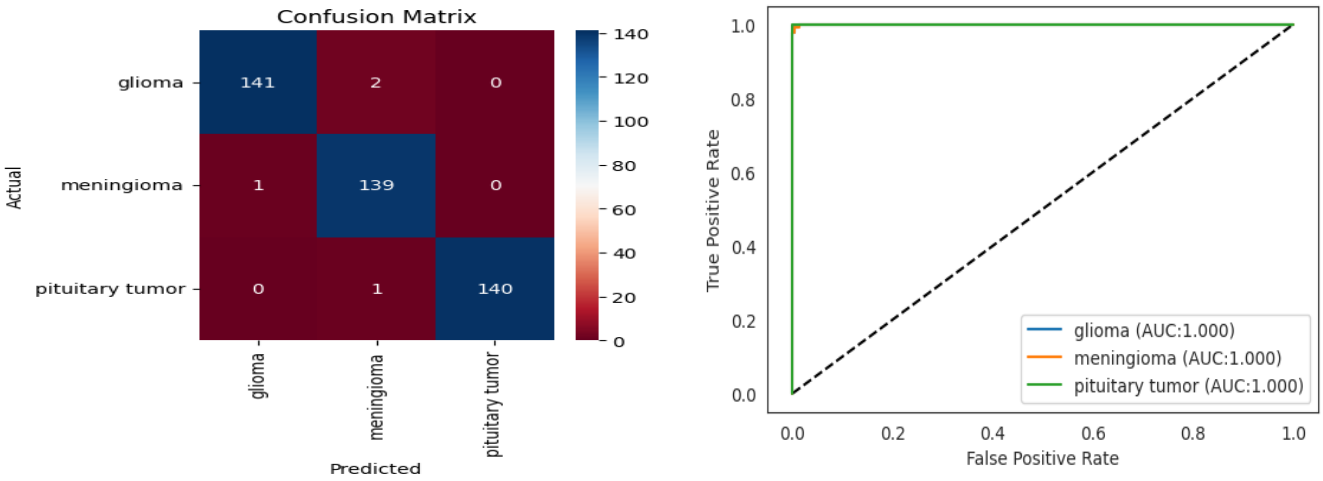

FIGURE 9. GGLA-NeXtE2NET results on 3-class brain tumor dataset (A) Confusion matrix (B) ROC-AUC curve.

For meningioma, it correctly identified 139 cases but misclassified 1 case as glioma. For pituitary tumors, the model correctly classified 140 cases, with 1 case misclassified as meningioma. Overall, the confusion matrix demonstrates the GGLA-NeXtE2NET model competency in accurately categorizing cases into the correct classes.

The ROC-AUC curve further verifies the GGLA-NeXtE2NET model effectiveness. The Figures 8(b) and 9(b) show the Receiver Operating Characteristic (ROC) curve along with the Area Under the Curve (AUC) values for each class. The ROC curve is a graphical representation that illustrates the diagnostic ability of a binary classifier system. For 4-class and 3-class brain tumor datasets, the curves for glioma, meningioma, and pituitary tumors all show perfect performance with an AUC of 100%, indicating that the GGLA-NeXtE2NET model can perfectly distinguish between the positive and negative classes for each type of tumor without any errors. The GGLA-NeXtE2NET model

TABLE 3. Performance comparison of different variations with GGLA-NeXtE2NET.

Model	Brain tumor dataset Data	Accuracy	Precision	Recall	F1-Score	Loss
EfficientNetV2S	4-class	97.73%	97.73%	97.74%	97.72%	0.078
	3-class	96.93%	96.96%	96.94%	96.93%	0.089
DBS-NeXtE2NET	4-class	98.23%	98.24%	98.25%	98.24%	0.064
	3-class	98.35%	98.38%	98.36%	98.35%	0.050
GGLA-E2NET	4-class	98.61%	98.61%	98.62%	98.62%	0.074
	3-class	98.11%	98.14%	98.11%	98.10%	0.048
GGLA-NeXtE2NET	4-class	99.62%	99.60%	99.62%	99.62%	0.014
	3-class	99.06%	99.05%	99.05%	99.06%	0.015

also attained an overall ROC-AUC score of 99.99% for both 4-class and 3-class brain tumor Datasets.

C. COMPARISON RESULTS OF DIFFERENT VARIATIONS WITH GGLA-NeXtE2NET

We embedded the DBS module and FE-GGLA module in the proposed GGLA-NeXtE2NET model with the backbone EfficientNetV2S and ConvNeXt models. Further, we conducted the ablation study to verify the each module effectiveness with original EfficientNetV2S, DBS model with original EfficientNetV2S and ConvNeXt(DBS-NeXtE2NET), GGLA module with EfficientNetV2S(GGLA-E2NET). This study highlights the impact of each module on overall performance by selectively removing individual modules from the GGLA-NeXtE2NET architecture. Table 3 provides a comprehensive comparison of the performance metrics for GGLA-NeXtE2NET with its variations on the 4-class and 3-class brain tumor datasets. EfficientNetV2S shows an accuracy of 97.73% on the 4-class dataset and 96.93% on the 3-class dataset, with corresponding precision, recall, and F1-scores being very similar to its accuracy. Its loss values are 0.078 for the 4-class and 0.089 for the 3-class dataset. DBS-NeXtE2NET improves upon these results, achieving an accuracy of 98.23% on the 4-class dataset and 98.35% on the 3-class dataset, with slightly better precision, recall, and F1-scores, and lower loss values of 0.064 and 0.050 respectively. GGLA-E2NET further enhances the performance, attaining an accuracy of 98.61% on the 4-class dataset and 98.11% on the 3-class dataset. Its precision, recall, and F1-scores are consistent with its accuracy, and the loss values are 0.074 for the 4-class and 0.048 for the 3-class dataset. The GGLA-NeXtE2NET model demonstrates the highest performance, achieving a remarkable accuracy of 99.62% on the 4-class dataset and 99.06% on the 3-class dataset. Its precision, recall, and F1-scores are nearly identical to its accuracy, and it exhibits the lowest loss values of 0.014 for the 4-class dataset and 0.015 for the 3-class dataset. This indicates a clear trend of performance improvement with each successive model variation.

The confusion matrix and ROC-AUC curve analysis of GGLA-NeXtE2NET variations can be seen in figure 10, figure 11, figure 12 and figure 13 respectively. Our GGLA-NeXtE2NET architecture clearly outperformed its variants and only misclassified 4 and 3 cases for 4-class and 3-class brain tumor datasets that can be seen in Figure 8(a), figure 9(a), figure 10 and Figure 12. Similarly, our proposed model maintained superior performance in Roc-AUC curve analysis compared to its variants.

D. COMPARISON RESULTS WITH AUGMENTED DATA AND WITHOUT AUGMENTED DATA

Figure 14 illustrates the comparative performance of the proposed GGLA-NeXtE2NET architecture when trained with GAN-based augmented, simple augmented pipeline and non-augmented data. The results indicate that the model performance is significantly enhanced when using data augmented with Generative Adversarial Networks (GANs) and a simple augmented pipeline. Without augmented data, the model accuracy is lower as can be seen in Figure 14. However, with GAN-based augmented data, the model achieves a 2.75% and 2.64% higher accuracy for the 4-class and 3-class brain tumor datasets against non-augmented data, respectively. On the other hand, a simple augmented pipeline also improves the accuracy of the model but is slightly shorter than GAN-based augmented data.

In addition to improved accuracy, the use of augmented data also enhances the GGLA-NeXtE2NET performance across other metrics. The precision, recall, and F1-scores are all significantly improved, indicating that the augmented technique helps the model recognize brain tumor images more effectively. Furthermore, the GAN-based augmentation introduces variability and diversity in the training data, which enhances the model's ability to generalize to new, unseen images. This leads to more robust and reliable classification performance, making the GGLA-NeXtE2NET model more effective in practical medical imaging applications.

E. VISUALIZATION WITH GRAD-CAM AND BOUNDING BOX

"Figure 15 and Figure 16" depict the complete overview of how well our model performs in the input image to highlight the most important part and also predict the particular

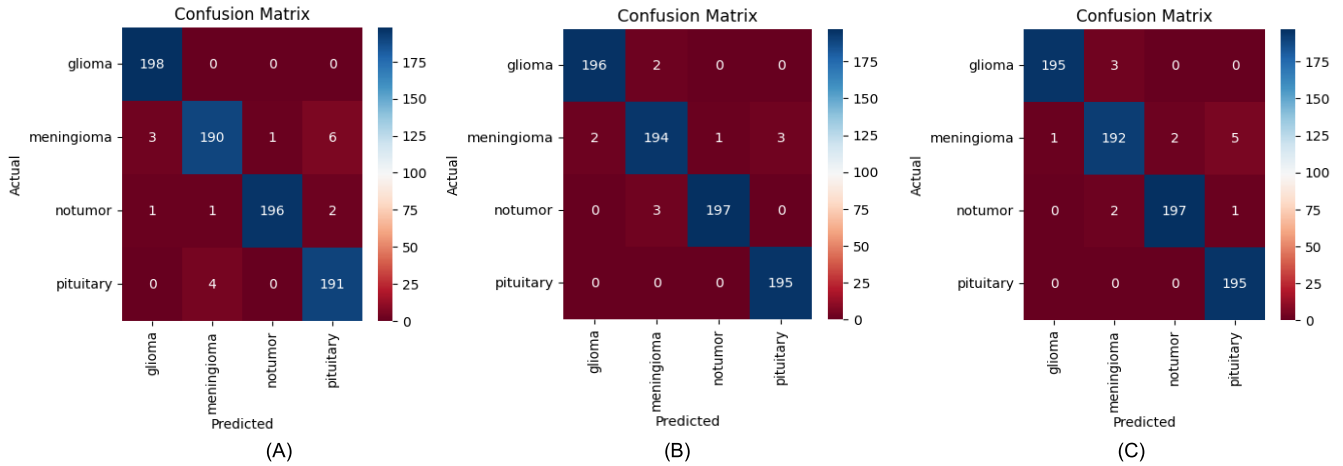

FIGURE 10. Confusion matrix results of (A) EfficientNetV2S (B) GGLA-E2NET (C) DBS-NeXtE2NET.

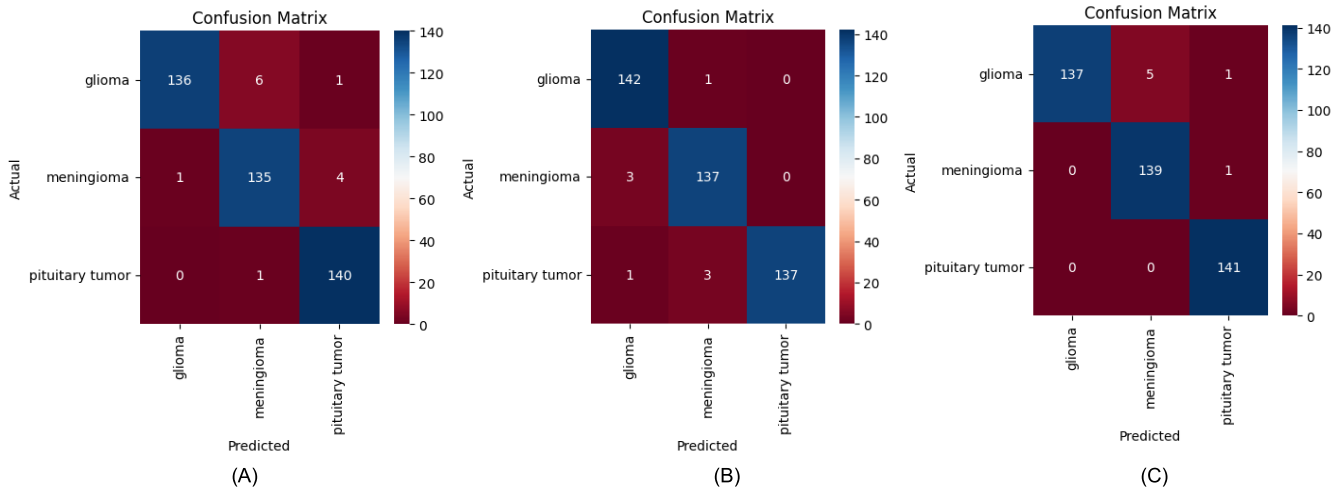

FIGURE 11. Confusion matrix of (A) EfficientNetV2S (B) GGLA-E2NET (C) DBS-NeXtE2NET.

class using this highlighted tumor region. The first step of our approach in the "figure15 and Figure 16" involves the application of Grad-CAM to the input image. Grad-CAM is a well-established method used for visualizing the regions in an image that exert the most significant influence on a neural network's prediction, particularly in classification tasks. To achieve this, we passed the input image through our proposed network and computed the gradients of the target class with respect to the final convolutional layer's feature maps. Following the Grad-CAM heatmap generation, we proceed to create a heatmap that vividly highlights the areas of the input image that bear the greatest relevance to the network's prediction for the target class. The heatmap is produced by weighting the feature maps in the final convolutional layer with the obtained gradients. Subsequently, higher values within the heatmap signify regions in the image that exerted a more substantial impact on the network's prediction, thereby pinpointing the areas of interest.

To facilitate further analysis and interpretation, we convert the Grad-CAM heatmap into a binary mask in figure 16. This transformation is achieved by applying a predefined threshold to the heatmap. Pixels with values exceeding this threshold are set to 1, indicating their significance, while those falling below the threshold are set to 0, indicating their relative insignificance. The choice of the threshold value can be tailored to the specific requirements of the application, allowing for different levels of sensitivity in identifying important regions.

Finally, our strategy culminates in the creation of bounding boxes around the crucial regions identified in the original image. To achieve this, we analyze the binary mask and identify connected components within it. Each connected component represents a coherent, contiguous region of importance within the image. For each such component, we calculate a bounding box that encloses the region. These bounding boxes effectively serve as a visual representation

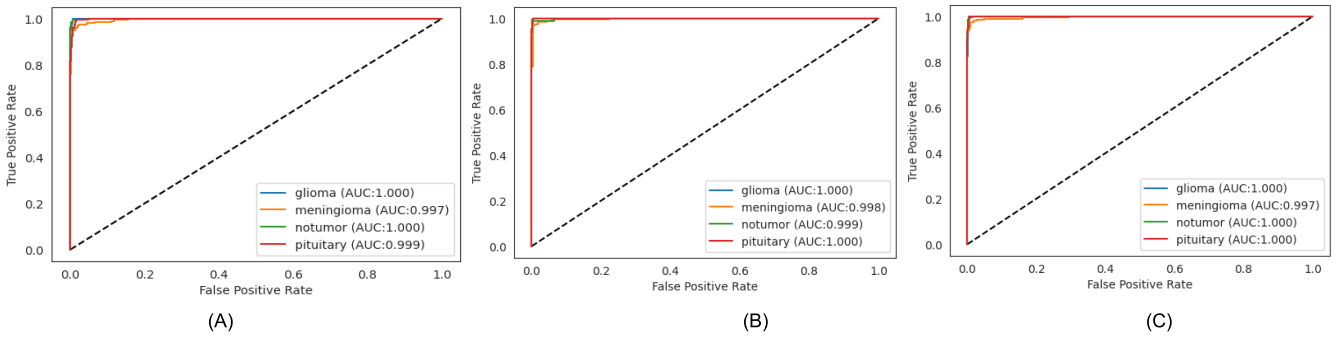

FIGURE 12. ROC-AUC curve of (A) EfficientNetV2S (B) GGLA-E2NET (C) DBS-NeXtE2NET using 4-class brain tumor dataset.

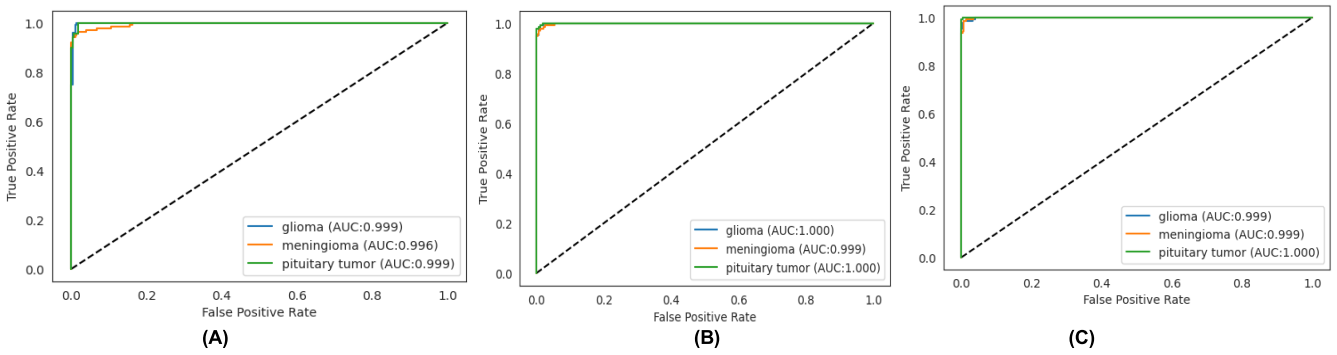

FIGURE 13. ROC-AUC curve of (A) EfficientNetV2S (B) GGLA-E2NET (C) DBS-NeXtE2NET 3-class brain tumor dataset.

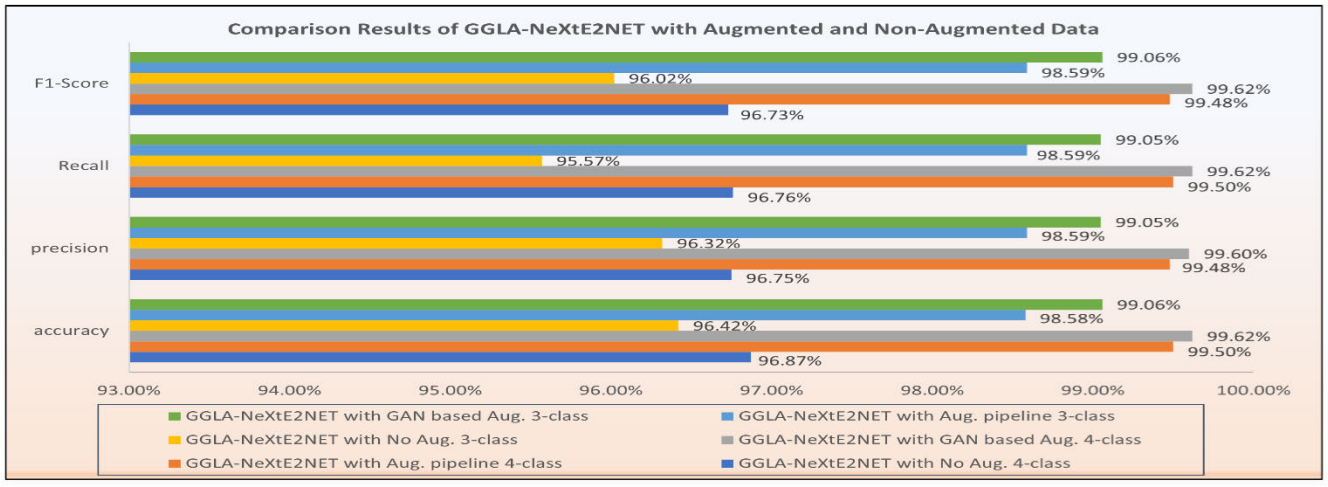

FIGURE 14. Comparison results of GGLA-NeXtE2NET with augmented and non-augmented data.

of the most significant regions within the input image, as determined by the Grad-CAM heatmap.

V. DISCUSSION

It is difficult and time-consuming to classify brain tumors into different types, because MRI images can vary so much. We have developed a new way to classify brain tumors that is more accurate than previous methods. Our approach starts by denoising the images to make them clearer. Then, we introduced the GGLA-NeXtE2NET network which makes it better at extracting information from the images. This allows the model to focus on the tumor areas and identify the type of tumor more accurately. In this section, we will discuss our approach in more detail and explain why it is better than previous methods.

A. COMPARISON RESULTS WITH EXISTING STATE-OF-THE-ART CNN MODELS

The table 4 compares various models for brain tumor classification with our proposed GGLA-NeXtE2NET model with preprocessing techniques, highlighting significant

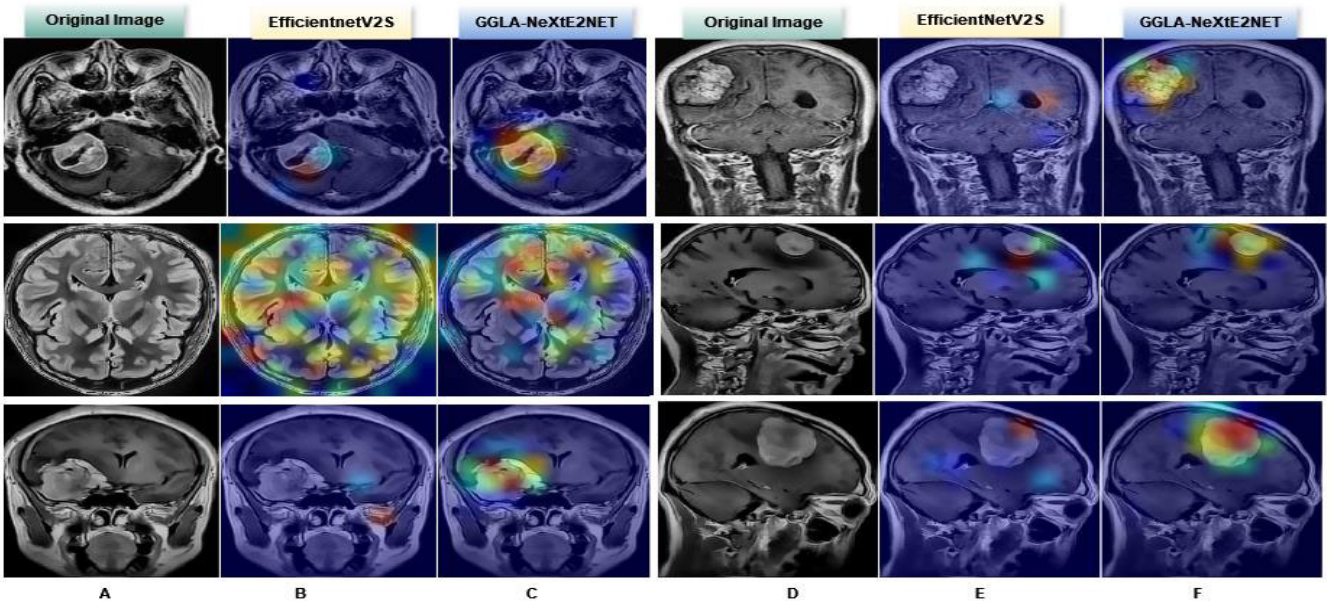

FIGURE 15. Grad-Cam analysis with (A-D) Ground Truth (B-E) EfficientNetV2S (C-F) GGLA-NeXtE2NET.

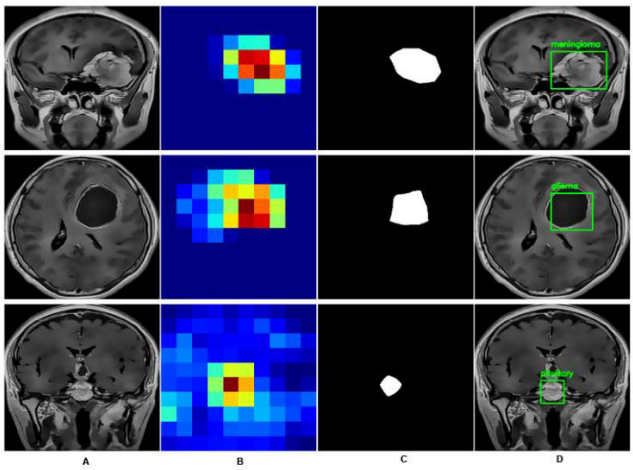

FIGURE 16. (A) Ground Truth image (B) predicted most important part in the image using heatmap (C) predicted binary mask of heatmap image using predefined threshold value (D) create bounding box with predicted accurate class name using binary mask in the original image.

performance differences. Wang et al. [22] used a modified version of the Vision Transformer model, achieving 80.78% accuracy and 80.00% recall for 4-class and 71.08% and 64.23% for 3-class classification. If we compare our technique to the vision transformer, our proposed model captures both local and global information simultaneously by using efficientNetV2S and ConvNeXt models and improves the accuracy by almost 19%, and 28% higher on 4-class and 3-class datasets. Mehmood et al. [23] used Pix2Pix-cGANs and NasNetLarge models for 3-class brain tumor datasets. Pix2Pix-cGANs is a novel approach but our models individually gave better results for image classification tasks on the imagenet database compared to NasNetLarge traditional model. Anaya-Isaza et al. [24] and Reyes and Sánchez [25] only used single models InceptionResNetV2 and EfficientNetB3 respectively for brain tumor classification. These models traditionally perform well for image classification tasks but brain tumor images contain complex, high-level information that these models struggle to capture due to the localized nature of convolutional layers. In contrast, our advanced model addresses this limitation by effectively capturing both local and global information, resulting in superior handling and classification of brain tumor images. Patil and Kirange [28], İncir and Bozkurt [26], Khan et al. [27] and Rizwan et al., [21] utilized traditional CNN models within ensemble networks that tend to collect similar types of information. In contrast, our model gathers richer information and more precisely highlights tumor areas, resulting in more accurate and detailed brain tumor classification. Senan et al. [29], and Haq et al. [30] utilized hybrid models (CNN + SVM) that struggle with feature representation and scalability, as it require separate training stages and may not fully leverage complex patterns in brain images. In contrast, our model integrates advanced feature extraction and classification in a unified framework, and achieves almost 4.50%, and 1.50% higher accuracy than both models respectively. CNN + LSTM [31], [32] models may struggle with capturing spatial and temporal dependencies effectively in brain tumor images, as LSTMs are typically designed to capture temporal dependencies in sequential data rather than complex spatial patterns. In these models, convolutional layers can extract local spatial features from brain tumor images, but they are limited in their ability to model long-range dependencies across the image.

While our GGLA-NeXtE2NET model focuses on brain tumor classification, recent advancements in semi-supervised medical image segmentation, such as the EVIL framework

TABLE 4. Comparison results of state-of-art-models with our proposed model.

Ref.	Classes	Preprocessing techniques	Model	Accuracy	Recall
Wang et al.[22]	4 classes	Data augmentation	Modified Vision transformer	80.78%	80.00%
	3 classes			71.08%	64.23%
Mehmood et al.[23]	3 classes	Tumor Region Colorization, image resizing	Pix2Pix-cGANs, NasNetLarge	92.4%	92.3%
Andres et al.[24]	3 Classes	Data Augmentation	InceptionResNetV2	97.22%	98.15%
Reyes et al.[25]	3 Classes	Data Augmentation	EfficientNetB3	98.72%	98.71%
	4 classes			97.50%	97.20%
Patil et al.[28]	3 Classes	Normalized Image, Augmentation	EDCNN (VGG + CNN)	97.77%	96.66%
. Incir et al.[26]	4 Classes	Data Augmentation, croping	InceptionV3 + DenseNet+EfficientNetV2	98.41%	98.5%
Khan et al.[27]	3 Classes	Data Augmentation	VGG16 + CNN	97.8%	96.4%
Rizwan et al.,[21]	3 Classes	Gaussian Filter, Augmentation	GCNN	97.14%	97.50%
Huang et al.,[20]	3 Classes	Resize Image	CNNBCN (CNN + Graph)	95.49%	-
Senan et al.[29]	4 Classes	Normalize image, Mean Filter, Laplacian Filter, Augmentation	Hybrid Model (Alex-Net + SVM)	95.10%	95.25%
Haq et al.[30]	3 Classes	Image Resizing, Image Registration	Deep CNN+SVM-RBF	98.3%	98.1%
Islam et al.[31]	2 classes	Data augmentation	Ensemble Model + LSTM	98.82%	98%
Shanthi et al.[32]	2 Classes	Data Augmentation	CNN +LSTM	97.50%	97.30%
Our Proposed	4 Classes	patch-based denoising, wavelet decomposition, Fourier transform, CLAHE, Laplacian Edge Enhancement, Crop Image, GAN based Data Augmentation	GGLA-NeXtE2NET(Dual Branch Ensemble Network (EfficientNetV2+ ConvNeXt) + FE-GGLA module)	99.62%	99.62%
	3 classes			99.06%	99.05%

TABLE 5. Comparison results of attention-based network with proposed attention-based network.

Ref.	Classes	Preprocessing techniques	Model	Accuracy	Recall
Alzahrani et al. [33]	4 classes	Data Augmentation, rescaling	ConvAttenMixer (self-attention & external-attention)	97.94%	96.65%
Alzahrani et al. [34]	3 Classes	Data Augmentation	Student-Teacher model with multi-head self-attention module	80.34%	-
pandi et al.[35]	3 Classes	Data Augmentation	Self-attention based GAN network	99.29%	-
. Tang et al.[39]	2 Classes	Data Augmentation	SpCaNet (Spinal Convolution Attention Network)	99.18%	98.95%
Ioannis et al.[41]	3 Classes	Image resizing	VGG16 + Gobal Avg pooling + Spatial attention module	95.65%	95.4%
Rasheed et al.[42]	4 Classes	Image resizing	CNN + Channel and Spatial Attention	98.33%	98.30%
Our Proposed	4 Classes	patch-based denoising, wavelet decomposition, Fourier transform, CLAHE, Laplacian Edge Enhancement, Crop Image, GAN based Data Augmentation	GGLA-NeXtE2NET(Dual Branch Ensemble Network (EfficientNetV2+ ConvNeXt) + FE-GGLA module)	99.62%	99.62%
	3 classes			99.06%	99.05%

[19], emphasize uncertainty-aware techniques to improve segmentation reliability under limited supervision. The EVIL framework utilizes the Dempster-Shafer Theory of Evidence to generate reliable pseudo-labels by quantifying uncertainty, which is highly effective in segmentation tasks but primarily targets label-scarce environments. In contrast, our model addresses challenges specific to fully supervised brain tumor classification, including high intra-class variation and inter-class similarity in MRI images, by employing the Gated Global-Local Attention (GGLA) mechanism and dual-branch ensemble design to enhance feature extraction and adaptability. Unlike EVIL, which emphasizes uncertainty estimation

TABLE 6. Training efficiency comparison of previously used hyperparameters with our hyperparameters.

Ref.	Optimizer	Activation Function	Loss Function	Classification Function	Training Epochs
Wang et al.[22]	SGD	GELU	Categorical Cross-entropy	-	50
Andres et al.[24]	Adadelta	ReLU	Categorical Cross-entropy	SoftMax	50
Patil et al.[28]	Adam	ReLU	-	-	30
Reyes et al.[25]	Adam	ReLU	Categorical Cross-entropy	SoftMax	30
Shanthi et al.[32]	-	ReLU	Categorical Cross-entropy	SoftMax	80
Khan et al.[27]	Adam	ReLU	Categorical Cross-entropy	SoftMax	80
Rizwan et al.,[21]	SGD	ReLU	Categorical Cross-entropy	SoftMax	300
Huang et al.,[20]	-	GeLU + ReLU		-	100
Haq et al.[30]	Adam	ReLU	Categorical Cross-entropy	SoftMax	100
Mehmood et al.[23]	Adam	ReLU, leaky ReLU	Categorical Cross-entropy	SoftMax	100
Islam et al.[31]	-	ReLU		-	30
Alzahrani et al. [33]	Adam	GELU	Categorical Cross-entropy	SoftMax	30
Rasheed et al.[42]	Adam	ReLU	Categorical Cross-entropy	SoftMax	50
Our Proposed	Adamax	Swish	Categorical Cross-entropy	SoftMax	20

and pseudo-labeling, GGLA-NeXtE2NET prioritizes feature refinement for more precise classification.

B. COMPARISON RESULTS WITH EXISTING ATTENTION-BASED MODEL

Further, table 5 compares various attention-based models for brain tumor classification with our proposed GGLA-NeXtE2NET model. The proposed model stands out in several aspects. In terms of architecture, the GGLA-NeXtE2NET utilizes a dual-branch ensemble network combining EfficientNetV2 and ConvNeXt, along with an FE-GGLA module. This sophisticated design appears to leverage the strengths of multiple advanced architectures.

The performance of the proposed model is exceptional, achieving 99.62% accuracy and recall for 4-class classification, surpassing all other models in the comparison. For 3-class classification, it maintains high performance with 99.06% accuracy and 99.05% recall. These results represent a significant improvement over other recent models, including those using attention mechanisms and specialized architectures like ConvAttenMixer [33] and SpCaNet [39], CNN + CBAM [42], Self-attention-based GAN network [35], and Student-Teacher model with a multi-head self-attention module [34].

C. TRAINING EFFICIENCY COMPARISON WITH PREVIOUSLY USED HYPERPARAMETERS

Table 6 compares our proposed model with others, emphasizing training efficiency components like optimizer, activation function, loss function and training epochs. While many prior studies rely on the Adam optimizer, our model uses Adamax, which offers enhanced stability and convergence speed in the presence of large gradients. Most previous models also utilize ReLU or GELU activation functions; however, we opted for Swish, which facilitates better gradient flow and faster convergence than ReLU and GELU. Notably, our model achieves convergence in just 20 epochs, highlighting a significant improvement in training efficiency compared to others, which often require up to 100 epochs or more. This reduction in training time indicates substantial gains in computational efficiency without sacrificing accuracy.

D. DENOISING STRATEGY

Original MRI images often exhibit noise, making it challenging to precisely delineate tumor boundaries within the brain. We employed a comprehensive set of preprocessing techniques that contribute to the model's high performance by enhancing the quality and diversity of input data.

"Figure 17" showcases the results and analysis of denoising a brain MRI image. The first set of images focuses on the frequency spectrum of both the noisy and denoised images, while the second set illustrates the visual comparison between the noisy, denoised, and different images, along with their corresponding histograms.

Starting with the frequency spectrum analysis, the left image represents the magnitude spectrum of the Fourier Transform of the noisy image. In this spectrum, bright spots indicate the presence of high-frequency components, which are typically associated with noise. The central bright spot represents the low-frequency components, corresponding to different images. The noisy image, displayed on the left, clearly shows the presence of noise as random speckles or variations, particularly noticeable in the homogeneous regions of the brain MRI. This noise obscures fine details and can complicate the interpretation of the image. The middle image in the top row is the denoised version of the original image. Here, the noise is significantly reduced, resulting in a smoother and clearer image. The fine details of the brain tumor and surrounding tissue are better preserved, making the image more suitable for diagnostic purposes. The rightmost

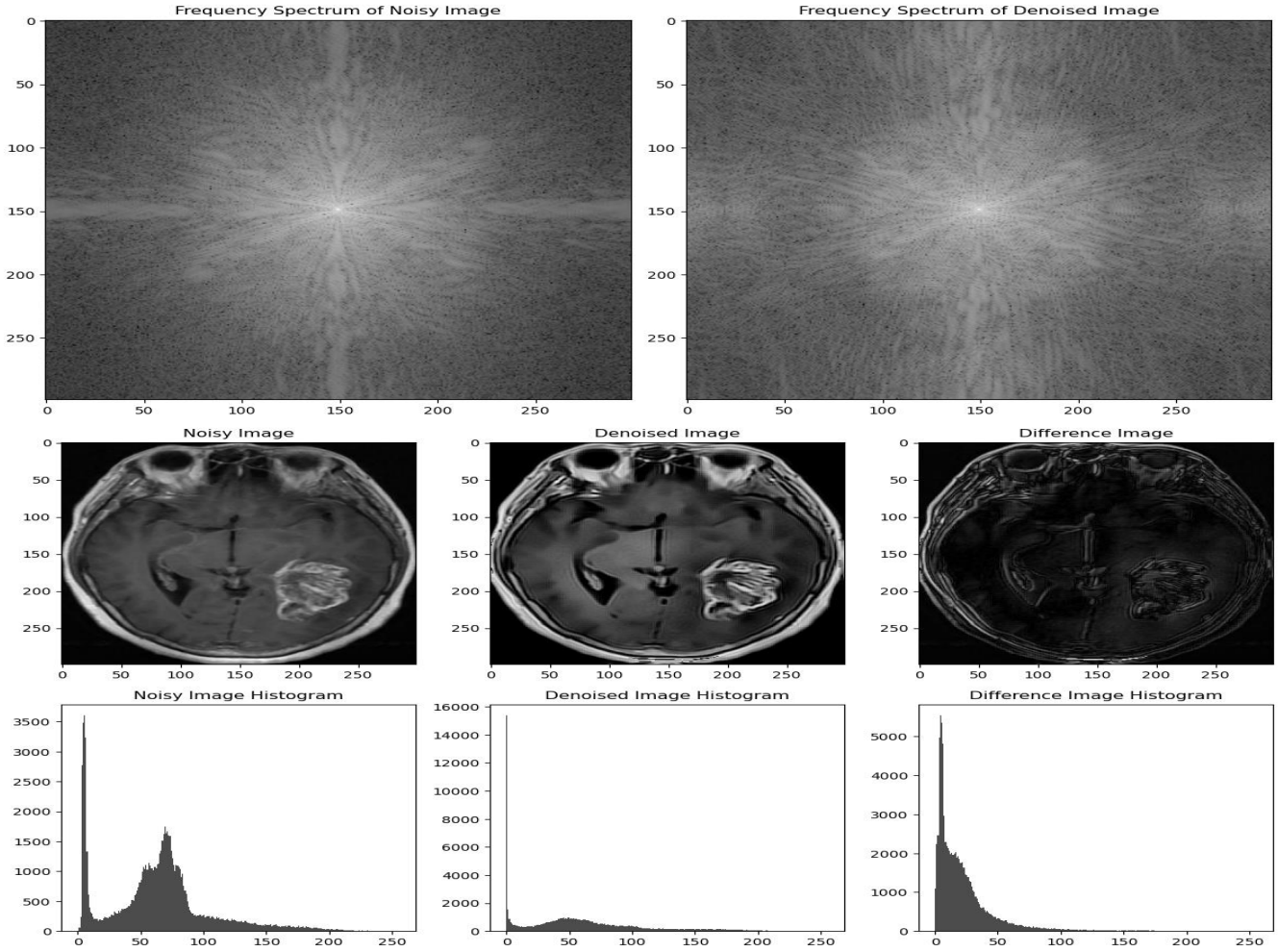

FIGURE 17. In first row, frequency spectra of the noisy (left) and denoised (right) brain MRI images show reduced high-frequency noise after denoising, while preserving essential image features. In second and third row, Visual and histogram analysis of noisy (left), denoised (middle), and difference (right) brain MRI images highlights effective noise reduction and preservation of critical image details.

image in the top row is the difference image, which represents the absolute difference between the noisy and denoised images. This difference image highlights the areas where the most significant noise reduction has occurred. Bright areas in this image indicate regions where the denoising algorithm made substantial changes, effectively removing noise. Darker areas show where the denoising algorithm made minimal changes, preserving the original image details.

The bottom row of the second set of images provides the histograms of the noisy, denoised, and different images. The histogram of the noisy image, shown on the left, reveals a wide distribution of pixel values, indicative of high variance due to the presence of noise. Peaks within this histogram correspond to both noise and image features. In contrast, the histogram of the denoised image, displayed in the middle, shows a more concentrated distribution of pixel values. This narrower distribution indicates reduced variance, suggesting that much of the noise has been removed. The peak around lower pixel values is particularly noticeable, signifying that the denoising process effectively targeted and reduced the noise. The histogram of the difference image, on the right, shows the distribution of pixel intensity differences between the noisy and denoised images. A high peak at low values indicates that most of the differences are small, meaning that the denoising process primarily removed noise without drastically altering the underlying image structure.

VI. LIMITATIONS AND FUTURE WORK

Although the model can process a range of image qualities, it performs optimally when both low- and high-resolution images are available. In practice, high-quality images may not always be accessible, potentially limiting the model's accuracy. Currently, the GGLA-NeXtE2NET model is designed for centralized learning setups, which may restrict its application in collaborative environments across multiple institutions. This limitation also introduces potential privacy concerns, as sensitive patient data would need to be transferred to a central location for model training.

The proposed GGLA-NeXtE2NET model could benefit from integration with distributed learning frameworks to support privacy-preserving, collaborative learning across

institutions without sharing sensitive data. For example, federated learning paradigms such as HyperFed [63], which combines hypernetwork-based personalization with physics-driven global sharing, could allow each participating institution to personalize the model based on its unique data distribution, improving model accuracy across varied settings. Additionally, a hybrid approach like Dynamic Corrected Split Federated Learning [64] would allow us to incorporate homomorphic encryption for enhanced data privacy while supporting a split-learning structure to distribute computational demands. DC-SFL's dynamic weight correction could further stabilize training across heterogeneous medical datasets, which is often the case with MRI data Future exploration of these integrations could make the GGLA-NeXtE2NET model more robust in privacy-sensitive environments, improving both scalability and accessibility in multi-institutional studies.

In the future, the YOLO (You Only Look Once) model can also be effectively applied for brain tumor detection. YOLO is a popular object detection framework that excels at real-time identification and localization of objects within images. In the context of brain tumor recognition, the YOLO model can be utilized to not only classify the tumor type, but also accurately delineate the tumor boundaries within the MRI scans. This dual capability of classification and segmentation is crucial for clinical applications, as it provides valuable information to assist with surgical planning and treatment strategies. The ability of YOLO to rapidly process images also makes it a viable option for deployment in time-sensitive medical workflows.

VII. CONCLUSION

We present the GGLA-NeXtE2NET model in this paper that is based on the Dual-Branch Ensemble (DBE) network and Gated Global-Local attention (GGLA) module to classify the brain tumor images. Our GGLA module successfully captures both global and local information in the brain tumor images simultaneously. Further, our DBE network allows the model to capture the information of brain tumor images by using different receptive fields that improve the model performance at different levels. We also embedded the GGLA module in the FE-GGLA block to extract features more accurately. Moreover, we have introduced a powerful denoising strategy that effectively reduces noise in MRI images, leading to clearer and more precise tumor boundary identification. We used the 3-class and 4-class brain tumor datasets to assess the proposed model performance and attained more remarkable results than previous studies. Furthermore, our work has addressed critical challenges in medical imaging, including the need for precise tumor boundary delineation, accurate classification of tumor types, and improved interpretability of deep learning models. These advancements hold the potential to revolutionize clinical decision-making, enabling more accurate diagnoses, treatment planning, and patient monitoring.

ACKNOWLEDGMENT

The authors would like to thank Prince Sultan University, Riyadh, Saudi Arabia, for support with the article processing charges (APC) of this publication. They specially acknowledge the Automated Systems and Soft Computing Laboratory (ASSCL), Prince Sultan University. They also acknowledge the editor and reviewers for their insightful comments, which have improved the quality of this publication.

REFERENCES

[1] C. A. Nelson, "Incidental findings in magnetic resonance imaging (MRI) brain research," *J. Law, Med. Ethics*, vol. 36, no. 2, pp. 315–319, 2008, doi: 10.1111/j.1748-720x.2008.00275.x.

[2] J. Wu, D. Guo, L. Wang, S. Yang, Y. Zheng, J. Shapey, T. Vercauteren, S. Bisdas, R. Bradford, S. Saeed, N. Kitchen, S. Ourselin, S. Zhang, and G. Wang, "TISS-Net: Brain tumor image synthesis and segmentation using cascaded dual-task networks and error-prediction consistency," *Neurocomputing*, vol. 544, Aug. 2023, Art. no. 126295, doi: 10.1016/j.neucom.2023.126295.

[3] M. G. Abbas Malik, A. Saeed, K. Shehzad, and M. Iqbal, "DEF-SwinE2NET: Dual enhanced features guided with multi-model fusion for brain tumor classification using preprocessing optimization," *Biomed. Signal Process. Control*, vol. 100, Feb. 2025, Art. no. 107079, doi: 10.1016/j.bspc.2024.107079.

[4] V. P. B. Grover, J. M. Tognarelli, M. M. E. Crossey, I. J. Cox, S. D. Taylor-Robinson, and M. J. W. McPhail, "Magnetic resonance imaging: Principles and techniques: Lessons for clinicians," *J. Clin. Exp. Hepatol.*, vol. 5, no. 3, pp. 246–255, Sep. 2015, doi: 10.1016/j.jceh.2015.08.001.

[5] P. Kunju, "Basics of MRI and CT," in *IAP Textbook of Pediatric Neurology*. India: IAP, 2014, p. 198, doi: 10.5005/jp/books/12166_22.

[6] D. N. Louis, A. Perry, P. Wesseling, D. J. Brat, I. A. Cree, D. Figarella-Branger, C. Hawkins, H. K. Ng, S. M. Pfister, G. Reifenberger, R. Soffietti, A. von Deimling, and D. W. Ellison, "The 2021 WHO classification of tumors of the central nervous system: A summary," *Neuro-Oncology*, vol. 23, no. 8, pp. 1231–1251, Aug. 2021, doi: 10.1093/neuonc/noab106.

[7] J. Benzakoun, C. Robert, L. Legrand, J. Pallud, J.-F. Meder, C. Oppenheim, F. Dhermain, and M. Edjlali, "Anatomical and functional MR imaging to define tumoral boundaries and characterize lesions in neuro-oncology," *Cancer/Radiothérapie*, vol. 24, no. 5, pp. 453–462, Aug. 2020, doi: 10.1016/j.canrad.2020.03.005.

[8] L. S. Cardona and E. M. Mata, "Imaging of brain tumors: MR spectroscopy and metabolic imaging," *Neuroimaging Clinics North Amer.*, vol. 83, no. 1, pp. 1–11, 2013.

[9] A. Koubaa, A. Ammar, M. Alahdab, A. Kanhouch, and A. T. Azar, "DeepBrain: Experimental evaluation of cloud-based computation offloading and edge computing in the Internet-of-Drones for deep learning applications," *Sensors*, vol. 20, no. 18, p. 5240, Sep. 2020, doi: 10.3390/s20185240.

[10] A. Saeed, F. Rehman, A. Ashfaq, H. Sharif, R. Zeeshan, and A. Zahid, "Acute lymphoblastic leukemia prediction from microscopic blood smear images using advanced machine learning," in *Proc. 14th Int. Conf. Math., Actuarial Sci., Comput. Sci. Statist. (MACS)*, Nov. 2022, pp. 1–6, doi: 10.1109/MACS56771.2022.10022378.

[11] K. Shehzad, T. Zhenhua, S. Shoukat, A. Saeed, I. Ahmad, S. S. Bhatti, and S. A. Chelloug, "A deep-ensemble-learning-based approach for skin cancer diagnosis," *Electronics*, vol. 12, no. 6, p. 1342, Mar. 2023, doi: 10.3390/electronics12061342.

[12] A. Saeed, S. Shoukat, K. Shehzad, I. Ahmad, A. A. Eshmawi, A. H. Amin, and E. Tag-Eldin, "A deep learning-based approach for the diagnosis of acute lymphoblastic leukemia," *Electronics*, vol. 11, no. 19, p. 3168, Oct. 2022, doi: 10.3390/electronics11193168.

[13] K. Shah, Z. A. Khan, A. Ali, R. Amin, H. Khan, and A. Khan, "Haar wavelet collocation approach for the solution of fractional order COVID-19 model using caputo derivative," *Alexandria Eng. J.*, vol. 59, no. 5, pp. 3221–3231, Oct. 2020, doi: 10.1016/j.aej.2020.08.028.

[14] M. Assam, H. Kanwal, U. Farooq, S. K. Shah, A. Mehmood, and G. S. Choi, "An efficient classification of MRI brain images," *IEEE Access*, vol. 9, pp. 33313–33322, 2021, doi: 10.1109/ACCESS.2021.3061487.

[15] M. Alnowami, E. Taha, S. Alsebaeai, S. Muhammad Anwar, and A. Alhawsawi, "MR image normalization dilemma and the accuracy of brain tumor classification model," *J. Radiat. Res. Appl. Sci.*, vol. 15, no. 3, pp. 33–39, Sep. 2022, doi: 10.1016/j.jrras.2022.05.014.

[16] E. Emary, H. M. Zawbaa, A. E. Hassanien, G. Schaefer, and A. T. Azar, "Retinal blood vessel segmentation using bee colony optimisation and pattern search," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2014, pp. 1001–1006, doi: 10.1109/IJCNN.2014.6889856.

[17] H. H. Inbarani, A. T. Azar, and G. Jothi, "Leukemia image segmentation using a hybrid histogram-based soft covering rough K-means clustering algorithm," *Electronics*, vol. 9, no. 1, p. 188, Jan. 2020, doi: 10.3390/electronics9010188.

[18] A. Chattopadhyay and M. Maitra, "MRI-based brain tumour image detection using CNN based deep learning method," *Neurosci. Informat.*, vol. 2, no. 4, Dec. 2022, Art. no. 100060, doi: 10.1016/j.neuri.2022.100060.

[19] Y. Chen, Z. Yang, C. Shen, Z. Wang, Z. Zhang, Y. Qin, X. Wei, J. Lu, Y. Liu, and Y. Zhang, "Evidence-based uncertainty-aware semi-supervised medical image segmentation," *Comput. Biol. Med.*, vol. 170, Mar. 2024, Art. no. 108004, doi: 10.1016/j.compbiomed.2024.108004.

[20] Z. Huang, X. Du, L. Chen, Y. Li, M. Liu, Y. Chou, and L. Jin, "Convolutional neural network based on complex networks for brain tumor image classification with a modified activation function," *IEEE Access*, vol. 8, pp. 89281–89290, 2020, doi: 10.1109/ACCESS.2020.2993618.

[21] M. Rizwan, A. Shabbir, A. R. Javed, M. Shabbir, T. Baker, and D. A.-J. Obe, "Brain tumor and glioma grade classification using Gaussian convolutional neural network," *IEEE Access*, vol. 10, pp. 29731–29740, 2022, doi: 10.1109/ACCESS.2022.3153108.

[22] J. Wang, S.-Y. Lu, S.-H. Wang, and Y.-D. Zhang, "RanMerFormer: Randomized vision transformer with token merging for brain tumor classification," *Neurocomputing*, vol. 573, Mar. 2024, Art. no. 127216, doi: 10.1016/j.neucom.2023.127216.

[23] M. Mehmood, N. Alshammari, S. A. Alanazi, A. Basharat, F. Ahmad, M. Sajjad, and K. Junaid, "Improved colorization and classification of intracranial tumor expanse in MRI images via hybrid scheme of Pix2Pix-cGANs and NASNet-large," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 34, no. 7, pp. 4358–4374, Jul. 2022, doi: 10.1016/j.jksuci.2022.05.015.

[24] A. Anaya-Isaza, L. Mera-Jiménez, L. Verdugo-Alejo, and L. Sarasti, "Optimizing MRI-based brain tumor classification and detection using AI: A comparative analysis of neural networks, transfer learning, data augmentation, and the cross-transformer network," *Eur. J. Radiol. Open*, vol. 10, Jan. 2023, Art. no. 100484, doi: 10.1016/j.ejro.2023.100484.

[25] D. Reyes and J. Sánchez, "Performance of convolutional neural networks for the classification of brain tumors using magnetic resonance imaging," *Heliyon*, vol. 10, no. 3, Feb. 2024, Art. no. e25468, doi: 10.1016/j.heliyon.2024.e25468.

[26] R. Incir and F. Bozkurt, "Improving brain tumor classification with combined convolutional neural networks and transfer learning," *Knowl.-Based Syst.*, vol. 299, Sep. 2024, Art. no. 111981, doi: 10.1016/j.knosys.2024.111981.

[27] M. S. I. Khan, A. Rahman, T. Debnath, M. R. Karim, M. K. Nasir, S. S. Band, A. Mosavi, and I. Dehzangi, "Accurate brain tumor detection using deep convolutional neural network," *Comput. Struct. Biotechnol. J.*, vol. 20, pp. 4733–4745, Sep. 2022, doi: 10.1016/j.csbj.2022.08.039.

[28] S. Patil and D. Kirange, "Ensemble of deep learning models for brain tumor detection," *Proc. Comput. Sci.*, vol. 218, pp. 2468–2479, Nov. 2023, doi: 10.1016/j.procs.2023.01.222.

[29] E. M. Senan, M. E. Jadhav, T. H. Rassem, A. S. Aljaloud, B. A. Mohammed, and Z. G. Al-Mekhlafi, "Early diagnosis of brain tumour MRI images using hybrid techniques between deep and machine learning," *Comput. Math. Methods Med.*, vol. 2022, pp. 1–17, May 2022, doi: 10.1155/2022/8330833.

[30] E. U. Haq, H. Jianjun, X. Huarong, K. Li, and L. Weng, "A hybrid approach based on deep CNN and machine learning classifiers for the tumor segmentation and classification in brain MRI," *Comput. Math. Methods Med.*, vol. 2022, pp. 1–18, Aug. 2022, doi: 10.1155/2022/6446680.

[31] M. N. Islam, M. S. Azam, M. S. Islam, M. H. Kanchan, A. H. M. S. Parvez, and M. M. Islam, "An improved deep learning-based hybrid model with ensemble techniques for brain tumor detection from MRI image," *Informat. Med. Unlocked*, vol. 47, Jan. 2024, Art. no. 101483, doi: 10.1016/j.imu.2024.101483.

[32] S. Shanthi, S. Saradha, J. A. Smitha, N. Prasath, and H. Anandakumar, "An efficient automatic brain tumor classification using optimized hybrid deep neural network," *Int. J. Intell. Netw.*, vol. 3, pp. 188–196, Jan. 2022, doi: 10.1016/j.ijin.2022.11.003.

[33] S. M. Alzahrani, "ConvAttenMixer: Brain tumor detection and type classification using convolutional mixer with external and self-attention mechanisms," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 35, no. 10, Dec. 2023, Art. no. 101810, doi: 10.1016/j.jksuci.2023.101810.

[34] S. M. Alzahrani and A. M. Qahtani, "Knowledge distillation in transformers with tripartite attention: Multiclass brain tumor detection in highly augmented MRIs," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 36, no. 1, Jan. 2024, Art. no. 101907, doi: 10.1016/j.jksuci.2023.101907.

[35] S. S. Pandi, A. Senthilselvi, T. Kumaragurubaran, and S. Dhanasekaran, "Self-attention-based generative adversarial network optimized with color harmony algorithm for brain tumor classification," *Electromagn. Biol. Med.*, vol. 43, nos. 1–2, pp. 31–45, Apr. 2024, doi: 10.1080/15368378.2024.2312363.

[36] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," 2017, *arXiv:1706.03762*.

[37] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis.*, in Lecture Notes in Computer Science: Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics, vol. 11211, Sep. 2018, pp. 3–19, doi: 10.1007/978-3-030-01234-2_1.

[38] J. Park, S. Woo, J. Y. Lee, and I. S. Kweon, "BAM: Bottleneck attention module," in *Proc. Brit. Mach. Vis. Conf. (BMVC)*, 2019, pp. 1–11.

[39] C. Tang, B. Li, J. Sun, S.-H. Wang, and Y.-D. Zhang, "GAM-SpCaNet: Gradient awareness minimization-based spinal convolution attention network for brain tumor classification," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 35, no. 2, pp. 560–575, Feb. 2023, doi: 10.1016/j.jksuci.2023.01.002.

[40] I. Pacal, O. Celik, B. Bayram, and A. Cunha, "Enhancing EfficientNetv2 with global and efficient channel attention mechanisms for accurate MRI-based brain tumor classification," *Cluster Comput.*, vol. 27, no. 8, pp. 11187–11212, Nov. 2024, doi: 10.1007/s10586-024-04532-1.

[41] I. D. Apostolopoulos, S. Aznaouridis, and M. Tzani, "An attention-based deep convolutional neural network for brain tumor and disorder classification and grading in magnetic resonance imaging," *Information*, vol. 14, no. 3, p. 174, Mar. 2023, doi: 10.3390/info14030174.

[42] Z. Rasheed, Y.-K. Ma, I. Ullah, M. Al-Khasawneh, S. S. Almutairi, and M. Abohashrh, "Integrating convolutional neural networks with attention mechanisms for magnetic resonance imaging-based classification of brain tumors," *Bioengineering*, vol. 11, no. 7, p. 701, Jul. 2024, doi: 10.3390/bioengineering11070701.

[43] X. T. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. C. Loy, "ESRGAN: Enhanced super-resolution generative adversarial networks," in *Proc. Comput. Vis. ECCV Workshops*, in Lecture Notes in Computer Science: Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics, vol. 11133, 2019, pp. 63–79, doi: 10.1007/978-3-030-11021-5_5.

[44] N. Alalwan, A. Alwadain, A. I. Alzahrani, A. H. Al-Bayatti, A. Abozeid, and R. M. A. El-Aziz, "Advancements in brain tumor identification: Integrating synthetic GANs with federated-CNNs in medical imaging analysis," *Alexandria Eng. J.*, vol. 105, pp. 105–119, Oct. 2024, doi: 10.1016/j.aej.2024.06.080.

[45] M. Yaqub, F. Jinchao, S. Ahmed, A. Mehmood, I. S. Chuhan, M. A. Manan, and M. S. Pathan, "DeepLabV3, IBCO-based ALCResNet: A fully automated classification, and grading system for brain tumor," *Alexandria Eng. J.*, vol. 76, pp. 609–627, Aug. 2023, doi: 10.1016/j.aej.2023.06.062.

[46] M. Nickparvar. *Brain Tumor MRI Dataset*. Accessed: May 20, 2024. [Online]. Available: https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset/data

[47] J. Cheng. (Jun. 2017). *Brain Tumor Dataset. Figshare. Dataset*. Accessed: May 20, 2024. [Online]. Available: https://doi.org/10.6084/m9.figshare.1512427.v5

[48] H. M. Rai and K. Chatterjee, "Detection of brain abnormality by a novel Lu-net deep neural CNN model from MR images," *Mach. Learn. Appl.*, vol. 2, Dec. 2020, Art. no. 100004, doi: 10.1016/j.mlwa.2020.100004.

[49] P. Jezzard and R. S. Balaban, "Correction for geometric distortion in echo planar images from B0 field variations," *Magn. Reson. Med.*, vol. 34, no. 1, pp. 65–73, Jul. 1995, doi: 10.1002/mrm.1910340111.

[50] M. Munowitz and A. Pines, "Multiple-quantum nuclear magnetic resonance spectroscopy," *Science*, vol. 233, no. 4763, pp. 525–531, Aug. 1986, doi: 10.1126/science.233.4763.525.
[51] D. K. Sodickson and W. J. Manning, "Simultaneous acquisition of spatial harmonics (SMASH): Fast imaging with radiofrequency coil arrays," *Magn. Reson. Med.*, vol. 38, no. 4, pp. 591–603, Oct. 1997, doi: 10.1002/mrm.1910380414.
[52] V. L. Yarnykh, "Actual flip-angle imaging in the pulsed steady state: A method for rapid three-dimensional mapping of the transmitted radiofrequency field," *Magn. Reson. Med.*, vol. 57, no. 1, pp. 192–200, Jan. 2007, doi: 10.1002/mrm.21120.
[53] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 2, Jun. 2017, pp. 60–65, doi: 10.1109/CVPR.2005.38.
[54] K. Leng, "An improved non-local means algorithm for image denoising," in *Proc. IEEE 2nd Int. Conf. Signal Image Process. (ICSIP)*, Aug. 2017, pp. 149–153, doi: 10.1109/SIPROCESS.2017.8124523.
[55] Y. Si, X. Li, L. Kong, J. Zhen, and Y. Li, "Improved empirical wavelet denoising algorithm with application to whirling detection in deep hole drilling process," *Proc. CIRP*, vol. 104, pp. 1924–1929, Aug. 2021, doi: 10.1016/j.procir.2021.11.325.
[56] G. Baldazzi, E. Sulas, M. Urru, R. Tumbarello, L. Raffo, and D. Pani, "Wavelet denoising as a post-processing enhancement method for non-invasive foetal electrocardiography," *Comput. Methods Programs Biomed.*, vol. 195, Oct. 2020, Art. no. 105558, doi: 10.1016/j.cmpb.2020.105558.
[57] M. Fontaine, A. Liutkus, L. Girin, and R. Badeau, "Explaining the parameterized Wiener filter with alpha-stable processes," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust. (WASPAA)*, Oct. 2017, pp. 51–55, doi: 10.1109/WASPAA.2017.8169993.
[58] S. M. Pizer, R. E. Johnston, J. P. Ericksen, B. C. Yankaskas, and K. E. Müller, "Contrast-limited adaptive histogram equalization: Speed and effectiveness," in *Proc. 1st Conf. Vis. Biomed. Comput.*, 1990, pp. 337–345, doi: 10.1109/VBC.1990.109340.
[59] J. He, M. Chen, and Z. Li, "Edge-aware local Laplacian filters for medical X-ray image enhancement," in *Proc. 5th Int. Conf. Health Inf. Sci.*, X. Yin, J. Geller, Y. Li, R. Zhou, H. Wang, and Y. Zhang, Eds., Cham, Switzerland: Springer, 2016, pp. 102–108.
[60] M. Tan and Q. V. Le, "EfficientNetV2: Smaller models and faster training," 2021, *arXiv:2104.00298*.
[61] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A ConvNet for the 2020s," 2022, *arXiv:2201.03545*.
[62] L. Tan, H. Wu, J. Xia, Y. Liang, and J. Zhu, "Skin lesion recognition via global-local attention and dual-branch input network," *Eng. Appl. Artif. Intell.*, vol. 127, Jan. 2024, Art. no. 107385, doi: 10.1016/j.engappai.2023.107385.
[63] Z. Yang, W. Xia, Z. Lu, Y. Chen, X. Li, and Y. Zhang, "Hypernetwork-based physics-driven personalized federated learning for CT imaging," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Dec. 15, 2024, doi: 10.1109/TNNLS.2023.3338867.
[64] Z. Yang, Y. Chen, H. Huangfu, M. Ran, H. Wang, X. Li, and Y. Zhang, "Dynamic corrected split federated learning with homomorphic encryption for U-shaped medical image networks," *IEEE J. Biomed. Health Informat.*, vol. 27, no. 12, pp. 5946–5957, Dec. 2023, doi: 10.1109/JBHI.2023.3317632.

ADNAN SAEED received the M.Sc. degree from the Virtual University of Pakistan, Lahore, Pakistan, in 2019, and the M.Phil. degree from Lahore Leads University, Pakistan, in 2022. He is currently an Accomplished Researcher of computer science, specializing in image processing, computer vision, and pattern recognition.

KHURRAM SHEHZAD received the B.S. degree from the COMSATS Institute of Information and Technology, Islamabad, Pakistan, and the M.S. degree in medical imaging from Northeast University, Shenyang, China. He is currently a Secondary School Teacher (SST) of computer science with the FG Public School (Boys), Multan Cantonment. He has published several research articles in renowned international journals. His research interests include medical imaging, artificial intelligence, motion detection, and generative AI.

SHAHZAD SARWAR BHATTI received the B.S. degree in electrical engineering from the University of Engineering and Technology, Lahore, Pakistan, the M.S. degree in telecommunication engineering from the University of Engineering and Technology, Peshawar, Pakistan, and the Ph.D. degree in computer science and technology from Shanghai Jiao Tong University, China. He is currently an Assistant Professor with the Faculty of Computing and Emerging Technologies, Emerson University, Multan, Pakistan. He has published several papers in well-known international journals and conference proceedings, including IEEE TRANSACTIONS ON MOBILE COMPUTING, IEEE TRANSACTIONS ON SERVICES COMPUTING, and IEEE International Conference on Web Services (ICWS). His research interests include artificial intelligence, machine learning, and deep learning.

SAIM AHMED received the B.S. degree in electronics from the Sir Syed University of Science and Technology, Pakistan, in 2009, the M.E. degree in industrial control and automation from Hamdard University, Pakistan, in 2013, and the Ph.D. degree in control science and engineering from Nanjing University of Science and Technology, China, in 2019. He is currently a Postdoctoral Researcher with Prince Sultan University, Saudi Arabia. His research interests include the theory and applications of adaptive control, sliding mode control, time delay control, robotic manipulators, nonlinearities, and their compensation.

AHMAD TAHER AZAR is currently a Full Professor with Prince Sultan University, Riyadh, Saudi Arabia. He is also a Full Professor with the Faculty of Computers and Artificial Intelligence, Benha University, Egypt. He is the Leader of the Automated Systems and Soft Computing Laboratory (ASSCL), Prince Sultan University. He has expertise in artificial intelligence, control theory and applications, robotics, machine learning, computational intelligence, and dynamic system modeling. He has authored/co-authored over 450 research papers in prestigious peer-reviewed journals, book chapters, and conference proceedings. He is also an Editor of IEEE SYSTEMS JOURNAL, IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, *Human-Centric Computing and Information Sciences* (Springer), and *Engineering Applications of Artificial Intelligence* (Elsevier).