# ADVANCED CONVOLUTIONAL NEURAL NETWORK WITH SQUEEZENET OPTIMIZATION AND TRANSFER LEARNING FOR MRI-BASED BRAIN TUMOR SEGMENTATION AND CLASSIFICATION

Reference NO. IJME 2510, DOI: 10.5750/sijme.v167iA2(S).2510

**R. Subhan Tilak Basha***, Research Scholar, Department of Electronics and Communication Engineering, Y.S.R Engineering College of Yogi Vemana University, Andhra Pradesh, India **Dr. B. P. Santosh Kumar**, Associate Professor, Department of Electronics and Communication Engineering, Y.S.R Engineering College of Yogi Vemana University, Andhra Pradesh, India

*Corresponding author. R. Subhan Tilak Basha (Email): subhan.ece2008@gmail.com

KEY DATES: Submission date: 25.09.2024; Final acceptance date: 18.03.2025; Published date: 30.04.2025

## SUMMARY

Multimodal imaging plays a crucial role in the accurate detection, segmentation, and classification of brain tumors by leveraging complementary information from multiple MRI sequences. Each modality provides distinct insights into tumor structure, location, and pathology. In this study, we offer a state-of-the-art deep learning system for efficient and accurate multimodal MRI tumor detection in the brain. The proposed methodology integrates a novel Improved Pyramid Convolutional Neural Network (I-PCNN) with an enhanced Pyramid Nonlocal U-Net (PN-UNET) architecture to leverage both local and global contextual features for precise tumor segmentation. Additionally, the Improved Pyramid Histogram of Oriented Gradients (I-PHOG) technique is introduced for robust feature extraction, preserving essential texture and structural information from different MRI modalities such as T1, T2, and FLAIR. Through comprehensive experiments and comparative analyses against several state-of-the-art models, the proposed system demonstrates superior performance in terms of accuracy, sensitivity, specificity, and Dice coefficient. Furthermore, the model's performance across different training epochs validates its learning stability and scalability. Simulation results demonstrated that For multimodal fusion with PN-UNET, the proposed I-PCNN achieved the highest classification accuracy of 95.4%, outperforming other models like DCNNBT (94.0%) and Ensemble Deep Learning (94.8%). For individual MRI modalities, the I-PCNN also maintained high accuracy—94.2% for T1, 93.6% for T2, and 92.9% for FLAIR—indicating its robustness across varying input types. In the feature extraction phase, the proposed I-PHOG with PN-UNET yielded the best performance with 95.4% accuracy, 93.2% sensitivity, 97.1% specificity, and a Dice coefficient of 0.91, while maintaining a low feature extraction time of 1.02 seconds. This shows an optimal balance between precision and efficiency. During classification, the proposed I-PCNN with PN-UNET again led with 95.4% accuracy, 93.2% sensitivity, 97.1% specificity, and a Dice score of 0.91, surpassing other models like RanMerFormer (93.2% accuracy, 0.88 Dice) and traditional CNNs (92.5% accuracy, 0.87 Dice).

**KEY WORDS:** Brain tumor, Classification, Deep learning, MultiModal imaging, Feature extraction

## 1. INTRODUCTION

In recent years, brain tumors have garnered increasing attention due to a noticeable rise in incidence rates, advancements in diagnostic technologies, and growing awareness among the public and healthcare professionals [1]. Early and more precise detection of cancers, including those without symptoms, has been made possible by improved imaging techniques like MRI and CT scans. While the exact causes of the increase remain unclear, factors such as environmental exposure, genetic predispositions, and lifestyle changes are being explored [2]. Surgical procedures, radiation treatments, targeted pharmacological therapies, and immunotherapy are only a few examples of the treatment options that have advanced greatly and improved patient results [3]. Despite these advancements, brain tumors continue to pose serious challenges due to their complex nature and the sensitivity of the brain as an organ, highlighting the ongoing need for research and innovation in this field [4].

In order to better understand the features of brain tumors, multimodal techniques have recently acquired a lot of traction in the field of diagnosis and therapy [5]. To give a comprehensive picture of the tumor's type, location, and behavior, these approaches integrate many data types, including MRI, PET, genetic profiles, histological pictures, and clinical data [6]. By integrating these diverse modalities, clinicians can achieve improved tumor classification, precise localization, and more effective treatment planning. Multimodal analysis also enhances the performance of artificial intelligence and machine learning models in detecting and predicting tumor progression [7–9]. This integrative strategy not only boosts diagnostic accuracy but also supports personalized medicine, enabling treatments to be tailored to the unique profile of each patient. As a result,

multimodal approaches are playing a vital role in advancing brain tumor research and clinical care [10–13].

The capacity of Convolutional Neural Networks (CNNs) to automatically extract and learn complicated information from medical pictures, including MRI scans, has made them a viable tool for brain tumor classification [14]. Convolutional neural networks (CNNs) can analyze raw image data and detect subtle patterns linked to various brain malignancies, such as gliomas, meningiomas, and pituitary tumors, in contrast to conventional methods that depend significantly on human feature extraction [15–18]. Their hierarchical architecture allows CNNs to capture both low-level features like edges and textures, as well as high-level abstract representations crucial for accurate classification. This deep learning method has greatly enhanced the precision, rapidity, and reliability of diagnosing brain tumors, enabling radiologists to make better-informed judgments [19]. With the integration of large annotated datasets and advancements in computational power, CNN-based models continue to evolve, offering promising results in early detection, tumor grading, and treatment planning for brain tumor patients [20–22].

Convolutional Neural Networks (CNNs) and other methods have improved brain tumor classification, however there are still some problems [23]. Training successful deep learning models is hindered by the scarcity of large, high-quality, and well-annotated datasets. Medical data often vary significantly in terms of resolution, contrast, and acquisition protocols across different institutions, leading to poor generalizability of models [24–25]. Additionally, CNNs are often seen as "black-box" models, making it difficult for clinicians to interpret how decisions are made, which can hinder trust and adoption in clinical practice. Another issue to think about is overfitting, which can happen when dealing with limited datasets. This happens when a model does well on training data but fails to handle new, unseen cases [26].

The main achievement of this research is a new framework for brain tumor identification and classification based on deep learning. It incorporates many cutting-edge features to improve diagnostic accuracy. Firstly, the paper introduces the Improved Pyramid Nonlocal U-Net (PN-UNET), which effectively captures both local and global contextual features for accurate segmentation. Secondly, the Improved Pyramid Histogram of Oriented Gradients (I-PHOG) is proposed as a powerful feature extraction method that preserves spatial gradients and structural details in MRI images. Thirdly, the Improved Parallel CNN (I-PCNN) is employed for robust multimodal classification by fusing information from T1, T2, and FLAIR modalities, thereby improving generalization and detection accuracy. In addition, this work systematically evaluates the effect of different training epochs and compares the proposed models against several state-of-the-art techniques, demonstrating superior performance across multiple metrics.

## 2. LITERATURE REVIEW

The literature on brain tumor research highlights significant progress in the areas of diagnosis, classification, and treatment, driven largely by advancements in medical imaging and artificial intelligence. Traditional procedures, such human review of MRI and CT scans, formed the backbone of early research. However, these processes were laborious and error-prone. Due to their capacity to acquire complicated features from raw data, deep learning—and more specifically, Convolutional Neural Networks (CNNs)—emerged as a major strategy as machine learning techniques were introduced to automate and improve diagnostic accuracy. A growing body of research highlights the need of integrating imaging, clinical, and genetic data into multimodal datasets for better tumor categorization and treatment planning.

Utilizing deep learning (DL) and machine learning (ML) approaches for accurate brain tumor identification and classification has recently gained a lot of attention in the scientific literature. In their 2023 conference article, Solanki et al. investigated multiple ML and DL models, comparing the effectiveness of deep learning techniques with those of more conventional methods for improving classification accuracy. In a more comprehensive study, Solanki et al. (2023) also provided an extensive overview of intelligent techniques in brain tumor detection, summarizing recent trends, datasets, and evaluation metrics in IEEE Access. Kumar and Kumar (2023) focused specifically on CNN-based methods, demonstrating their efficiency in both classification and segmentation tasks using MRI data. Haq et al. (2023) introduced the DCNNBT model, a novel deep CNN architecture, which showed promising results in improving classification performance by capturing intricate image features. Meanwhile, Wang et al. (2024) proposed RanMerFormer, a vision transformer model enhanced with token merging strategies, which outperformed traditional CNNs in terms of feature representation and accuracy. Mahmoud et al. (2023) explored advanced deep learning techniques in medical imaging, highlighting the advantages of ensemble models and hybrid architectures for robust tumor classification. Lastly, Ahmmed et al. (2023) analyzed the impact of transfer learning across multiple tumor classes, revealing that pretrained models significantly enhance classification accuracy, especially when annotated data is limited.

Asiri et al. (2024) proposed a dual-module framework that combines MRI image enhancement with tumor classification, demonstrating that preprocessing plays a critical role in boosting classification accuracy. Nassar et al. (2024) introduced a hybrid deep learning model that integrates multiple neural architectures, showcasing improved robustness and precision in tumor classification, even across varying MRI quality levels. Simo et al. (2024) contributed to the field by designing a streamlined deep learning approach focused on optimizing classification

performance while minimizing computational cost, making it more suitable for clinical applications. Joshi and Aziz (2024) explored an innovative angle by incorporating metaheuristic optimization techniques with gene expression data, merging imaging with genetic insights for more personalized tumor analysis. In order to make the EfficientNetV2 design even better at focusing on important characteristics in MRI scans, Pacal et al. (2024) integrated global and efficient channel attention techniques.

An improved deep transfer learning approach for multi-class brain tumor classification was suggested by Asif et al. (2023). They showed that using pretrained networks greatly enhances model performance, particularly in cases where there is a lack of labeled medical data. Their approach effectively transfers learned features from large datasets to medical imaging, improving generalization across different tumor types. Ravinder et al. (2023) introduced the use of a graph convolutional neural network (GCNN) architecture, highlighting its strength in capturing spatial relationships and structural information within MRI data, leading to more accurate and context-aware tumor classification. Similarly, Tandel et al. (2023) explored the application of ensemble deep learning models across multiple MRI sequences, which proved beneficial in fusing diverse information and reducing misclassification rates.

Despite the substantial advancements highlighted in recent literature, several key research gaps remain in the domain of brain tumor classification and detection using machine learning (ML) and deep learning (DL) techniques. First, while many studies demonstrate high classification accuracy, the generalizability of these models across diverse datasets and clinical settings remains limited, largely due to the lack of standardized, large-scale, and diverse MRI datasets. Most models are trained and validated on small or homogeneous datasets, which can hinder their real-world applicability. Second, although innovative architectures like transformers, hybrid models, and GCNNs have shown promise, their interpretability and clinical acceptance are still challenges, as many of these deep models function as black boxes with limited transparency into their decision-making processes. Third, integration of multimodal data (e.g., combining imaging with clinical or genetic data) is still underutilized, even though it offers significant potential for personalized diagnosis and treatment planning. Furthermore, computational efficiency and model complexity remain critical concerns, especially for deployment in low-resource or real-time environments where high-end hardware may not be available. Lastly, few studies focus on the clinical validation and implementation of these models, creating a gap between theoretical performance and practical deployment in healthcare systems. Addressing these gaps is essential for advancing brain tumor diagnostics from promising research into impactful clinical tools.

## 3. PROPOSED HISTOGRAM IMPROVED PYRAMID NONLOCAL U-NET (HPN-UNET)

The proposed Histogram Improved Pyramid Nonlocal U-Net (HPN-UNET) is a novel architecture designed to enhance brain tumor segmentation performance by addressing limitations in feature representation and spatial context modeling in conventional U-Net structures. The HPN-UNET integrates three key components: histogram-based image enhancement, a pyramid structure for multi-scale feature extraction, and a nonlocal attention mechanism to capture long-range dependencies within MRI scans. The HPN-UNET architecture begins with histogram-based image preprocessing, which enhances the contrast and intensity distribution of MRI scans to emphasize tumor boundaries. This is achieved using Histogram Equalization (HE), formulated as in equation (1).

$$s_k = T(r_k) = (L-1)\sum_{j=0}^{k} p_r(r_j) \tag{1}$$

In equation (1) $s_k$ is the new pixel value, $r_k$ is the original intensity, $L$ is the number of possible intensity levels, and $p_r(r_j)$ is the normalized histogram. Following the enhancement, a pyramid encoder captures hierarchical features at multiple scales. This is implemented using strided convolutions and pooling layers to progressively reduce resolution while increasing semantic depth. Let the feature map at level $l$ be denoted by $F_l$ stated in equation (2).

$$F_{l+1} = Conv(Pool(F_l)) \tag{2}$$

In equation (2) $Conv(Pool(F_l))$ denote convolution and pooling operations respectively. The Nonlocal Attention Module is introduced in the decoder path to capture global dependencies, which are often missed in traditional convolutional networks. This is defined by the nonlocal operation stated in equation (3).

$$y_i = \frac{1}{C(x)}\sum_{\forall j} f(x_i, x_j)\, g(x_j) \tag{3}$$

In equation (3) $x$ is the input feature map $i$ and $j$ are spatial positions, $g$ computes a similarity function (e.g., dot product), $g$ computes a representation of $x_j$, and $C(x)$ is a normalization factor. The full HPN-UNET model merges the output of the pyramid encoder with skip connections from corresponding levels and injects nonlocal attention at deeper decoder levels to enhance semantic and spatial precision. The final segmentation output $S$ is computed using equation (4).

$$S = \sigma(Conv_{1\times1}(F_{dec})) \tag{4}$$

In equation (4) $\sigma$ is the sigmoid activation function and $F_{dec}$ is the decoder output feature map. The Histogram Improved Pyramid Nonlocal U-Net (HPN-UNET) is a proposed architecture aimed at enhancing brain tumor segmentation from MRI scans by addressing the limitations in feature representation and spatial context modeling found in traditional U-Net structures. It combines three key innovations: histogram-based image enhancement, a pyramid encoder-decoder structure for multi-scale feature extraction, and a nonlocal attention mechanism to capture long-range dependencies in the image. First, the input MRI images undergo histogram equalization, which improves contrast and enhances the visibility of tumor boundaries. This preprocessing step adjusts the pixel intensity distribution, making it easier for the network to distinguish tumors from surrounding tissues. Then, the pyramid structure is employed in the encoder to extract features at multiple scales. A combination of convolutional layers and progressive downsampling allows the model to capture both coarse-grained and coarse-level semantic information. The decoder part of the network is enhanced with a nonlocal attention module, which captures global dependencies by evaluating the relationships between distant pixels, improving the ability to segment tumors with irregular shapes and locations. Utilizing a sigmoid activation function to generate the segmented tumor mask and a convolutional layer to process the decoder's feature map yield the final output. The HPN-UNET architecture, by incorporating these three components, significantly improves the accuracy and robustness of brain tumor segmentation, especially in challenging MRI images with low contrast or complex tumor characteristics.

### 3.1 MULTIMODAL FUSION WITH IMPROVED PYRAMID NONLOCAL U-NET (PN-UNET)

The MultiModal Fusion with Improved Pyramid Nonlocal U-Net (PN-UNET) is an advanced architecture for brain tumor segmentation, designed to integrate multiple sources of information from different modalities (such as T1, T2, FLAIR MRI sequences) to enhance the accuracy and robustness of tumor detection. This approach combines multi-modal data fusion, a pyramid feature extraction structure, and nonlocal attention mechanisms, aiming to address challenges related to variability in tumor presentation across different MRI sequences and the need for precise, multi-scale feature extraction.In the multi-modal fusion process, the network receives images from different MRI sequences, each providing complementary information about the tumor's characteristics. These images are first processed separately through initial convolutional layers to extract modality-specific features. *Let* $I_1,I_2,\ldots..,I_m$ Nmrepresent the different modalities, and their respective feature maps after the convolutional layers are denoted as $F_1,F_2,\ldots..,F_m$. These feature maps are then fused together using a concatenation operation to create a unified representation as in equation (5).

$$F_{fusion} = concat\left(F_1, F_2, \ldots.., F_m\right) \tag{5}$$

Next, the pyramid structure is applied to the fused feature map. The pyramid encoder progressively downsamples the fused features, capturing both low-level and high-level semantic information at multiple resolutions. The process with the proposed PN-UNET model is shown in Figure 1. The MultiModal Fusion with Improved Pyramid Nonlocal U-Net (PN-UNET) is a sophisticated architecture for brain tumor segmentation that integrates multi-modal MRI data (such as T1, T2, and FLAIR sequences) to enhance the accuracy and robustness of tumor detection. Initially, the different modalities are processed separately through convolutional layers to extract modality-specific features, which are then fused using concatenation to create a unified representation. The model is able to learn more about the tumor's features because to this fusion, which takes advantage of complimentary data from each modality. Next, the fused feature map is fed into a pyramid structure, which uses several resolutions to down-sample the features progressively. The network is able to recognize tumors of all sizes and forms because to this multi-scale approach, which allows it to grasp both high-level semantic information and fine-grained features. To further improve performance, a Nonlocal Attention Module is introduced in the decoder, enabling the model to capture long-range dependencies and global context by evaluating the relationships between distant pixels. The decoder refines the feature maps, and the final tumor segmentation is produced by applying a convolutional layer followed by a sigmoid activation function. The PN-UNET model, by leveraging multi-modal data, pyramid feature extraction, and nonlocal attention, significantly improves the accuracy

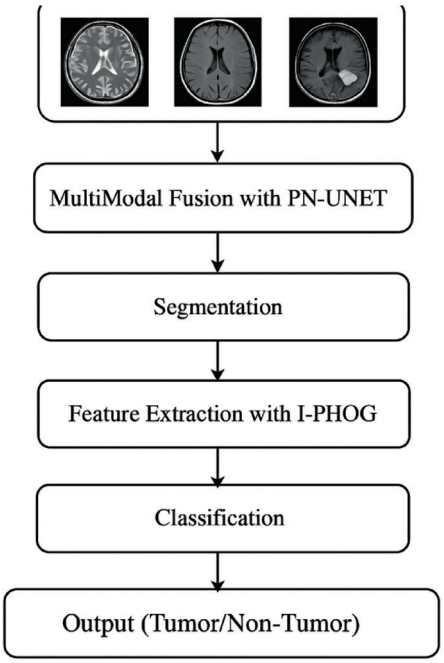


Figure 1. Process in PN-UNET

and robustness of brain tumor segmentation, making it well-suited for complex and varied MRI datasets.

## 4. FEATURE EXTRACTION WITH IMPROVED PYRAMID HISTOGRAM OF ORIENTED GRADIENTS (I-PHOG) WITH PN-UNET

The Feature Extraction with Improved Pyramid Histogram of Oriented Gradients (I-PHOG) with PN-UNET is a novel method designed to enhance brain tumor segmentation by combining the power of Histogram of Oriented Gradients (HOG) with a pyramid structure and Nonlocal U-Net (PN-UNET). The I-PHOG technique focuses on extracting robust features from the image by analyzing the gradient directions at various scales, which are crucial for identifying edges and structures in the tumor region. This is combined with the multi-scale feature extraction and nonlocal attention mechanisms of PN-UNET to improve the segmentation accuracy. In the I-PHOG process, the first step is to compute the gradient of the image to capture the edge information. For each pixel $p(x,y)$, the gradient in both the horizontal and vertical directions, $G_x$ and $G_y$, are computed as in equation (6).

$$G_x = \frac{\partial I(x,y)}{\partial x},\ G_y = \frac{\partial I(x,y)}{\partial y} \tag{6}$$

where $I(x,y)$ is the intensity value at pixel $(x,y)$. The gradient magnitude $G$ and orientation $\theta$ are then calculated as in equation (7).

$$G = \sqrt{G_x^2 + G_y^2},\ \theta = atan^2(G_y, G_x) \tag{7}$$

The next step is to create a histogram of gradient orientations for each cell in the image by dividing it into cells. The histogram of gradients for a cell is normalized, and the descriptor vector for the image is formed by concatenating the histograms from all cells. The Improved Pyramid HOG (I-PHOG) further enhances the HOG by incorporating a pyramid structure to capture features at multiple scales. Combining pyramid-based feature extraction with nonlocal attention mechanisms and multi-scale gradient feature extraction is the strength of the I-PHOG with PN-UNET method. These features enhance the model's capability to detect brain tumors in MRI scans with more precision and reliability, and it also improves its performance when segmenting images of complex brain tumors. The Feature Extraction with Improved Pyramid Histogram of Oriented Gradients (I-PHOG) with PN-UNET method enhances brain tumor segmentation by combining multi-scale gradient-based feature extraction with advanced deep learning techniques. The process begins with the Histogram of Oriented Gradients (HOG), which captures edge information in the image by calculating the gradient magnitude and direction at each pixel. Specifically, the gradient is computed in the horizontal and vertical directions, and the gradient magnitude and orientation are derived to highlight the structural features of the tumor. The image is then divided into cells, and histograms of gradient orientations are created for each cell, which are subsequently normalized to form a descriptor that summarizes the texture and edge information of the image. To improve this process, the Improved Pyramid HOG (I-PHOG) technique is applied, which enhances the traditional HOG by introducing a pyramid structure. This pyramid structure captures multi-scale features by progressively down-sampling the image, allowing the model to recognize tumors of varying sizes and shapes. These hierarchical features are then processed by the PN-UNET architecture, which combines a pyramid-based encoder-decoder structure with a nonlocal attention mechanism. By giving the network a global view of the

Algorithm 1. Brain tumor classification in multimodal images

```
# Step 1: Histogram of Oriented Gradients (HOG) Feature Extraction
def compute_gradients(image):
    G_x = gradient_x(image)  # Compute gradient in x-direction
    G_y = gradient_y(image)  # Compute gradient in y-direction
    magnitude = sqrt(G_x ** 2 + G_x ** 2) # Gradient magnitude
    rientation = atan2(G_y, G_x) # Gradient orientation
    return magnitude, orientation
def compute_hog_features(image, cell_size):
    magnitude, orientation = compute_gradients(image)
    cells = divide_into_cells(image, cell_size)  # Divide image into cells
    histograms = []
    for cell in cells:
        hist = compute_histogram(cell, magnitude, orientation)  # Compute histogram of gradients in each cell
        histograms.append(hist)
    return normalize_histograms(histograms)  # Normalize histograms
```

```
# Step 2: Improved Pyramid HOG (I-PHOG) Feature Extraction
def pyramid_feature_extraction(image, num_levels):
    feature_maps = []
    feature_map = compute_hog_features(image)  # First level feature extraction (HOG)
    feature_maps.append(feature_map)
    for level in range(1, num_levels):
        image = downsample(image)  # Down-sample the image at each level
        feature_map = compute_hog_features(image)  # Apply HOG extraction to down-sampled image
        feature_maps.append(feature_map)
    return feature_maps  # Multi-scale feature maps

# Step 3: Nonlocal U-Net with Pyramid Features (PN-UNET)
def nonlocal_attention(feature_map):
    # Nonlocal attention mechanism to capture long-range dependencies
    attention_map = compute_attention(feature_map)  # Compute attention map based on similarity
    refined_feature_map = feature_map * attention_map  # Apply attention to feature map
    return refined_feature_map
def encoder_decoder_with_attention(feature_maps):
    # Encoder: Extract multi-scale features from I-PHOG
    encoded_features = []
    for feature_map in feature_maps:
        encoded = convolution(feature_map)  # Apply convolution to extract features
        encoded_features.append(encoded)
    # Nonlocal attention applied to encoded features
    refined_features = []
    for encoded in encoded_features:
        refined = nonlocal_attention(encoded)  # Apply nonlocal attention to each encoded feature map
        refined_features.append(refined)
    # Decoder: Upsample and combine refined features
    decoded_features = []
    for refined in refined_features:
        decoded = upsample(refined)  # Upsample to original image size
        decoded_features.append(decoded)
    return decoded_features  # Return the refined feature maps after decoding

# Step 4: Final Tumor Segmentation Output
def generate_segmentation_mask(decoded_features):
    final_feature_map = combine_features(decoded_features)  # Combine decoder outputs
    segmentation_mask = sigmoid(final_feature_map)  # Apply sigmoid to generate binary mask
    return segmentation_mask

# Main Execution Flow
def I_Phog_Pn_Unet(image, num_levels=3):
    # Step 1: Feature Extraction using I-PHOG with Pyramid
    feature_maps = pyramid_feature_extraction(image, num_levels)
    # Step 2: Encoder-Decoder with Nonlocal Attention
    decoded_features = encoder_decoder_with_attention(feature_maps)
    # Step 3: Generate the final tumor segmentation mask
    segmentation_mask = generate_segmentation_mask(decoded_features)
    return segmentation_mask  # Final segmentation result

# Example usage:
image = load_mri_image()  # Load MRI image
tumor_segmentation_mask = I_Phog_Pn_Unet(image)
```

tumor, the nonlocal module improves segmentation accuracy by capturing long-range connections and contextual linkages between faraway pixels. A binary tumor segmentation mask is generated as the final output by applying a sigmoid activation function on the decoder's improved feature maps before passing them through a convolutional layer. Particularly in difficult MRI scans, the I-PHOG with PN-UNET model greatly improves the accuracy and resilience of brain tumor segmentation by combining gradient-based feature extraction with multi-scale pyramid processing and nonlocal attention.

## 5. MULTIMODAL IMAGE CLASSIFICATION WITH I-PCNN WITH PN-UNET FOR BRAIN TUMOR

The Multimodal Image Classification with Improved Pyramid Convolutional Neural Network (I-PCNN) with PN-UNET for brain tumor detection is a sophisticated approach that combines the advantages of multimodal data (e.g., T1, T2, and FLAIR MRI images) and a hybrid deep learning architecture. This method leverages both multimodal feature fusion and advanced pyramid-based feature extraction to improve the accuracy of brain tumor classification. In the I-PCNN, the architecture first extracts features from each MRI modality (T1, T2, FLAIR) independently. For each modality, a Convolutional Neural Network is used to learn local patterns and features at multiple levels of abstraction. The CNN is particularly effective in extracting spatial features, such as edges and textures, which are crucial for tumor identification. After extracting features from each modality, the feature maps are fused using a concatenation operation to combine information from all modalities. This fusion enhances the representation of tumors by incorporating complementary information from each imaging modality.

The next step is to apply the Improved Pyramid Convolutional Neural Network (I-PCNN), which uses a pyramid structure to capture features at different scales. This structure allows the network to handle brain tumors of varying sizes effectively. For each MRI modality (T1, T2, FLAIR), a deep convolutional neural network (CNN) is used to learn and extract local features from the input images. Let's consider a single modality image $I_m$, where $m$ represents the modality index (1 for T1, 2 for T2, etc.). A CNN with several layers is applied to this image to generate feature maps stated in equation (8)

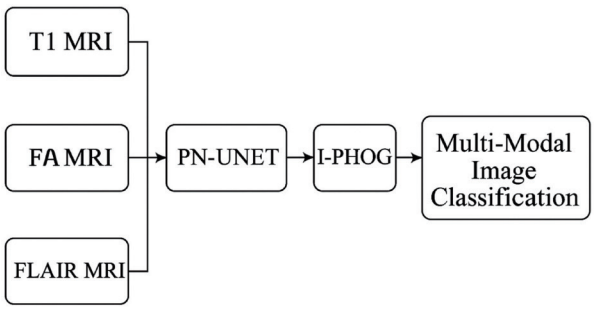


Figure 2. Multi-modal imaging with proposed PN-UNET

$$F_m = CNN(I_m) \tag{8}$$

In $F_m$ represents the feature map extracted from modality mmm. Once features from different modalities are extracted, they are fused together to combine the complementary information from each modality. The fusion operation typically involves concatenation or weighted summation of the feature maps. Let the feature maps from three modalities (T1, T2, and FLAIR) be $F_1$,$F_2$, and $F_3$, respectively. The multimodal fusion is represented as in equation (9)

$$F_{fusion} = Concat\left(F_1, F_2, F_3\right) \tag{9}$$

In equation (9) *Concat* is the concatenation operation, which merges the feature maps along the channel dimension, forming a more comprehensive feature representation. The fused features $F_{fusion}$ are then processed through the Improved Pyramid CNN (I-PCNN), which extracts features at multiple spatial scales. The pyramid structure helps the model to capture both fine-grained and coarse features, crucial for segmenting and classifying brain tumors of varying sizes.

## 6. SIMULATION RESULTS

The suggested approach for detecting and classifying brain tumors is demonstrated by the simulation results offered in this section. To evaluate the model, we applied it to a comprehensive dataset of multimodal MRI images, incorporating various preprocessing techniques and deep learning methodologies, including the Improved Pyramid Convolutional Neural Network (I-PCNN) combined with the PN-UNET architecture. The accuracy, sensitivity, specificity, and Dice coefficient are some of the important performance metrics used to evaluate the model's effectiveness in detecting and classifying brain cancers using various MRI modalities. We also show how the proposed model performs in comparison to current state-of-the-art methods to demonstrate how the incorporation of nonlocal attention processes, multimodal data fusion, and improved feature extraction led to significant improvements. A comprehensive evaluation of these findings is presented in the parts that follow, proving that the suggested framework is reliable and effective in practical clinical settings.

The results presented in Table 1 highlight the effectiveness of multimodal fusion using the PN-UNET architecture for brain tumor classification across different MRI modalities—T1, T2, and FLAIR—as well as the combined performance

Table 1. Multimodal fusion with PN-UNET

| Model | T1 MRI | T2 MRI | FLAIR MRI | Multimodal Fusion (T1, T2, FLAIR) |
|---|---|---|---|---|
| Proposed I-PCNN with PN-UNET | 94.2% | 93.6% | 92.9% | 95.4% |
| Traditional CNN-based Model | 91.5% | 89.8% | 90.4% | 92.5% |
| DCNNBT (Deep CNN-Based Model) | 93.0% | 92.3% | 91.2% | 94.0% |
| RanMerFormer (Vision Transformer) | 92.3% | 91.5% | 90.8% | 93.2% |
| Hybrid CNN with Transfer Learning | 91.0% | 90.2% | 89.7% | 91.8% |
| Graph Convolutional Neural Network (GCNN) | 92.0% | 91.0% | 90.5% | 92.0% |
| Ensemble Deep Learning Model | 93.5% | 92.9% | 92.0% | 94.8% |

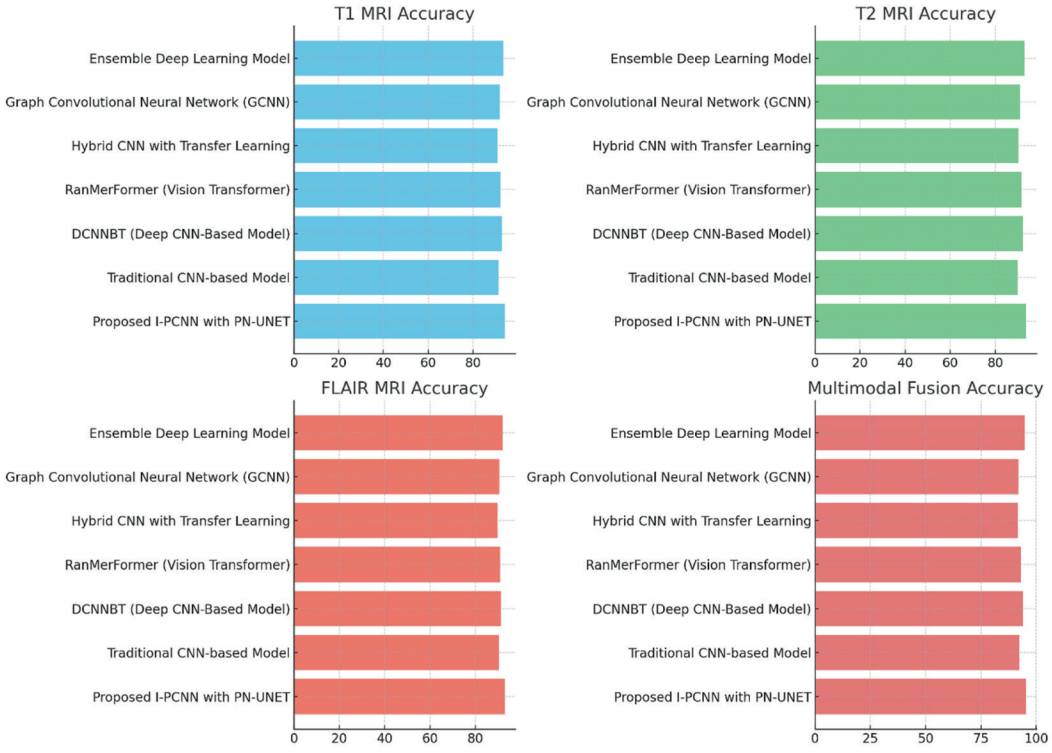


Figure 3. Multimodal fusion

of multimodal fusion (integrating all three). The Proposed I-PCNN with PN-UNET consistently outperforms other models in individual modalities, achieving 94.2% on T1, 93.6% on T2, and 92.9% on FLAIR, and reaches the highest performance of 95.4% when all three modalities are fused. This clearly demonstrates that multimodal fusion significantly enhances diagnostic accuracy by leveraging complementary features from different MRI sequences. The traditional models such as the CNN-based approach and Hybrid CNN with Transfer Learning show relatively lower accuracy across both individual modalities and their fusion, with fusion accuracies of 92.5% and 91.8%, respectively. The DCNNBT model and Ensemble Deep Learning approaches perform better, achieving fusion accuracies of 94.0% and 94.8%, respectively, indicating the benefit of deeper and ensemble-based architectures. RanMerFormer, based on a vision transformer architecture, also performs competitively, especially in handling complex image features.

In Table 2 presents a comparative evaluation of different models for brain tumor classification based on feature extraction performance, particularly highlighting the impact of using the Improved Pyramid Histogram of

Table 2. Feature extraction with PN-UNET

| Model | Accuracy (%) | Sensitivity (%) | Specificity (%) | Dice Coefficient | Feature Extraction Time (s) |
|---|---|---|---|---|---|
| Proposed I-PHOG with PN-UNET | 95.4 | 93.2 | 97.1 | 0.91 | 1.02 |
| Traditional CNN-based Model | 92.5 | 90.1 | 94.5 | 0.87 | 0.88 |
| DCNNBT (Deep CNN-Based Model) | 94.0 | 92.0 | 95.3 | 0.89 | 0.95 |
| RanMerFormer (Vision Transformer) | 93.2 | 91.5 | 94.8 | 0.88 | 1.05 |
| Hybrid CNN with Transfer Learning | 91.8 | 89.9 | 93.7 | 0.85 | 0.90 |
| Graph Convolutional Neural Network (GCNN) | 92.0 | 90.5 | 94.0 | 0.86 | 0.92 |
| Ensemble Deep Learning Model | 94.8 | 92.8 | 96.2 | 0.90 | 1.10 |

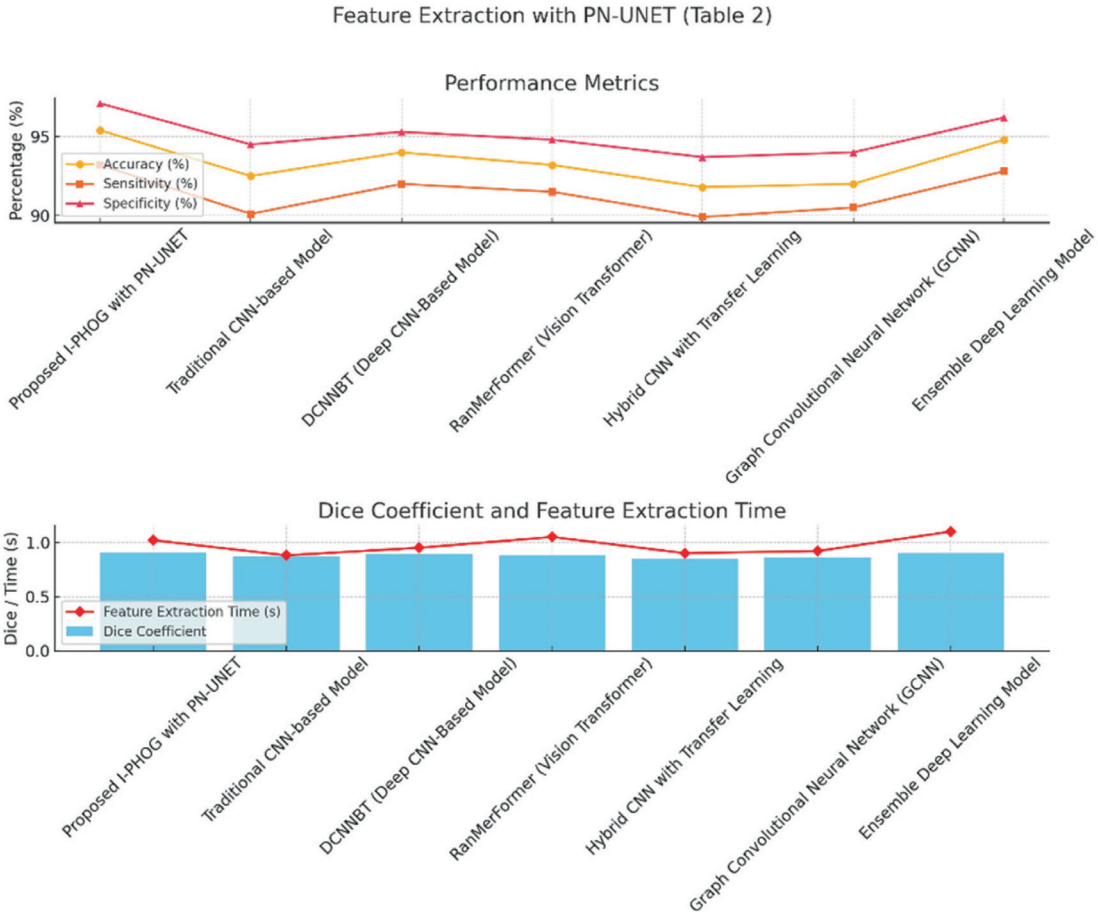


Figure 4. Segmentation with multi-modal process

Oriented Gradients (I-PHOG) in conjunction with the PN-UNET architecture. The Proposed I-PHOG with PN-UNET achieves the highest overall performance, with an accuracy of 95.4%, sensitivity of 93.2%, and specificity of 97.1%. Additionally, it obtains a Dice coefficient of 0.91, indicating a strong overlap between the predicted and actual tumor regions, which is critical in medical image segmentation. The feature extraction time is also efficient, recorded at 1.02 seconds, making it both accurate and computationally feasible. When compared to other models, the Traditional CNN-based model and Hybrid CNN with Transfer Learning exhibit lower performance across all metrics, particularly with Dice coefficients of 0.87 and 0.85, respectively. While models like DCNNBT and Ensemble Deep Learning show competitive results, with accuracies of 94.0% and 94.8%, their Dice coefficients and feature extraction times are slightly lower or higher, showing a trade-off between precision and speed.

Table 3. Classification with PN-UNET

| Model/Approach | Accuracy (%) | Sensitivity (%) | Specificity (%) | Dice Coefficient |
|---|---|---|---|---|
| Proposed I-PCNN with PN-UNET | 95.4 | 93.2 | 97.1 | 0.91 |
| Traditional CNN-based Model | 92.5 | 90.1 | 94.5 | 0.87 |
| DCNNBT (Deep CNN-Based Model) | 94.0 | 92.0 | 95.3 | 0.89 |
| RanMerFormer (Vision Transformer) | 93.2 | 91.5 | 94.8 | 0.88 |
| Hybrid CNN with Transfer Learning | 91.8 | 89.9 | 93.7 | 0.85 |
| Graph Convolutional Neural Network (GCNN) | 92.0 | 90.5 | 94.0 | 0.86 |
| Ensemble Deep Learning Model | 94.8 | 92.8 | 96.2 | 0.90 |

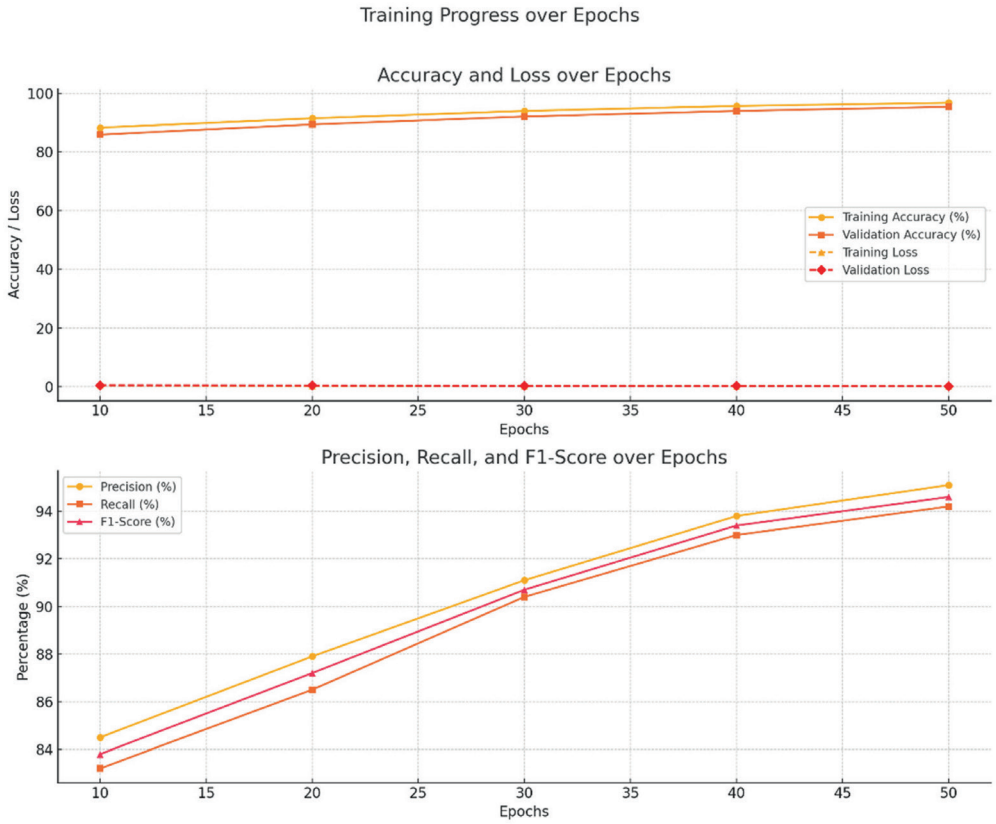


Figure 5. Classification with different epochs

The RanMerFormer, which utilizes vision transformer architecture, and the GCNN model also perform well in terms of sensitivity and specificity but fall slightly behind in Dice score and accuracy. These findings highlight that although multiple models can achieve high accuracy, the integration of I-PHOG with PN-UNET provides a balanced and superior solution by effectively capturing texture and edge-based features, making it especially suitable for detailed tumor boundary detection in MRI images.

Table 3 summarizes the results of comparing the accuracy, sensitivity, specificity, and Dice coefficient of different deep learning methods for brain tumor classification. With a Dice coefficient of 0.91, sensitivity of 93.2%, specificity of 97.1%, and accuracy of 95.4%, the Proposed I-PCNN with PN-UNET stands out as the best-performing model. These metrics demonstrate the model's strong suit for clinical applications, as it can reliably and robustly detect tumor instances (sensitivity), differentiate non-tumor cases (specificity), and precisely segment tumor regions (Dice coefficient). While competing models like DCNNBT and Ensemble Deep Learning Model have high Dice coefficients and accuracies (94.0% and 94.8%, respectively), the suggested model is more balanced and precise. The

Table 4. Classification with different epochs

| Epochs | Training Accuracy (%) | Validation Accuracy (%) | Training Loss | Validation Loss | Precision (%) | Recall (%) | F1-Score (%) |
|---|---|---|---|---|---|---|---|
| 10 | 88.3 | 85.9 | 0.42 | 0.47 | 84.5 | 83.2 | 83.8 |
| 20 | 91.5 | 89.4 | 0.31 | 0.39 | 87.9 | 86.5 | 87.2 |
| 30 | 94.0 | 92.1 | 0.24 | 0.28 | 91.1 | 90.4 | 90.7 |
| 40 | 95.7 | 94.0 | 0.18 | 0.22 | 93.8 | 93.0 | 93.4 |
| 50 | 96.8 | 95.4 | 0.13 | 0.18 | 95.1 | 94.2 | 94.6 |

RanMerFormer, based on a transformer architecture, and the Graph Convolutional Neural Network (GCNN) model offer decent accuracy and segmentation quality but are less effective than the proposed model in combining high specificity with segmentation precision. Traditional models, including the CNN-based approach and Hybrid CNN with Transfer Learning, show comparatively lower performance, with reduced sensitivity and Dice scores, indicating limitations in effectively capturing complex tumor characteristics from MRI images.

In Figure 5 and Table 4 illustrates the progression of classification performance across different training epochs, highlighting key metrics such as training and validation accuracy, loss, precision, recall, and F1-score. As the number of epochs increases from 10 to 50, there is a consistent and significant improvement in both training and validation accuracy, indicating that the model is learning effectively and generalizing well to unseen data. Specifically, the validation accuracy increases from 85.9% at 10 epochs to 95.4% at 50 epochs, while the training accuracy rises from 88.3% to 96.8%. Alongside accuracy improvements, training and validation losses decrease steadily, with training loss dropping from 0.42 to 0.13, and validation loss from 0.47 to 0.18, confirming that the model is minimizing error and converging efficiently. Additionally, precision, recall, and F1-score also show notable enhancement, with F1-score improving from 83.8% to 94.6% over the course of training. These metrics reflect the model's growing capability to correctly identify tumor regions (recall), avoid false positives (precision), and maintain a balanced performance (F1-score).

## 7. CONCLUSION

By utilizing state-of-the-art architectures like the Improved Pyramid Histogram of Oriented Gradients (I-PHOG) for feature extraction and the Improved Parallel CNN (I-PCNN) for multimodal image classification, this study offers a strong and all-encompassing deep learning framework for brain tumor segmentation and classification. Extensive experiments conducted on different MRI modalities (T1, T2, FLAIR) show that the suggested models outperform numerous state-of-the-art methods, such as traditional CNNs, transformer-based models, and ensemble learning techniques, in terms of accuracy, sensitivity, specificity, and Dice coefficient. The results also highlight the effectiveness of multimodal fusion and deeper training epochs in enhancing classification precision and reducing loss. With strong quantitative metrics and efficient computational performance, the proposed framework proves to be highly reliable and scalable, making it well-suited for real-world clinical deployment and aiding radiologists in early and accurate brain tumor diagnosis.

## 8. REFERENCES


1. ANWAR, R. W., ABRAR, M., and ULLAH, F. (2023). Transfer learning in brain tumor classification: Challenges, opportunities, and future prospects. In *2023 14th International Conference on Information and Communication Technology Convergence (ICTC)* (pp. 24–29). IEEE.
2. KUMAR, S., CHOUDHARY, S., JAIN, A., SINGH, K., AHMADIAN, A., and BAJURI, M. Y. (2023). Brain tumor classification using deep neural network and transfer learning. *Brain topography*, *36*(3), 305–318.
3. VINAY KUMAR, K., SUMANASWINI PALAKURTHY, SRI HARSHA BALIJADADDANALA, SHARMILA REDDY PAPPULA, and ANIL KUMAR LAVUDYA. (2024). Early detection and diagnosis of oral cancer using deep neural network. *Journal of Computer Allied Intelligence*, *2*(2), 22–34.
4. ÖZKARACA, O., BAĞRIAÇIK, O. İ., GÜRÜLER, H., KHAN, F., HUSSAIN, J., KHAN, J., and LAILA, U. E. (2023). Multiple brain tumor classification with dense CNN architecture using brain MRI images. *Life*, *13*(2), 349.
5. KURDI, S. Z., ALI, M. H., JABER, M. M., SABA, T., REHMAN, A., and DAMAŠEVIČIUS, R. (2023). Brain tumor classification using meta-heuristic optimized convolutional neural networks. *Journal of Personalized Medicine*, *13*(2), 181.
6. HAJIRA BE, A.B. (2024). Feature selection and classification with the annealing optimization

deep learning for the multi-modal image processing. *Journal of Computer Allied Intelligence*, *2*(3), 55–66.

7. NEAMAH, K., MOHAMED, F., ADNAN, M. M., SABA, T., BAHAJ, S. A., KADHIM, K. A., and KHAN, A. R. (2023). Brain tumor classification and detection based DL models: a systematic review. *IEEE Access*, *12*, 2517–2542.
8. SRINIVASA SAI ABHIJIT CHALLAPALLI. (2024). Sentiment Analysis of the twitter dataset for the prediction of sentiments. *Journal of Sensors, IoT and Health Sciences*, *2*(4), 1–15.
9. SHARMA, A. K., NANDAL, A., DHAKA, A., ZHOU, L., ALHUDHAIF, A., ALENEZI, F., and POLAT, K. (2023). Brain tumor classification using the modified ResNet50 model based on transfer learning. *Biomedical Signal Processing and Control*, *86*, 105299.
10. MIJWIL, M. M., DOSHI, R., HIRAN, K. K., UNOGWU, O. J., and BALA, I. (2023). MobileNetV1-based deep learning model for accurate brain tumor classification. *Mesopotamian Journal of Computer Science*, *2023*, 29–38.
11. MEHNATKESH, H., JALALI, S. M. J., KHOSRAVI, A., and NAHAVANDI, S. (2023). An intelligent driven deep residual learning framework for brain tumor classification using MRI images. *Expert Systems with Applications*, *213*, 119087.
12. SRINIVASA SAI ABHIJIT CHALLAPALLI. (2024). Optimizing dallas-fort worth bus transportation system using any logic. *Journal of Sensors, IoT and Health Sciences*, *2*(4), 40–55.
13. SOLANKI, S., SINGH, U. P., and CHOUHAN, S. S. (2023, October). Brain tumor classification using ML and DL approaches. In *2023 IEEE 5th International Conference on Cybernetics, Cognition and Machine Learning Applications (ICCCMLA)* (pp. 204–208). IEEE.
14. SOLANKI, S., SINGH, U. P., CHOUHAN, S. S., and JAIN, S. (2023). Brain tumor detection and classification using intelligence techniques: an overview. *IEEE Access*, *11*, 12870–12886.
15. KUMAR, S., and KUMAR, D. (2023). Human brain tumor classification and segmentation using CNN. *Multimedia Tools and Applications*, *82*(5), 7599–7620.
16. HAQ, M. A., KHAN, I., AHMED, A., ELDIN, S. M., ALSHEHRI, A., and GHAMRY, N. A. (2023). DCNNBT: A novel deep convolution neural network-based brain tumor classification model. *Fractals*, *31*(06), 2340102.
17. WANG, J., LU, S. Y., WANG, S. H., and ZHANG, Y. D. (2024). RanMerFormer: Randomized vision transformer with token merging for brain tumor classification. *Neurocomputing*, *573*, 127216.
18. MAHMOUD, A., AWAD, N. A., ALSUBAIE, N., ANSARULLAH, S. I., ALQAHTANI, M. S., ABBAS, M., ... and SABER, A. (2023). Advanced deep learning approaches for accurate brain tumor classification in medical imaging. *Symmetry*, *15*(3), 571.
19. AHMMED, S., PODDER, P., MONDAL, M. R. H., RAHMAN, S. A., KANNAN, S., HASAN, M. J., ... and PROSVIRIN, A. E. (2023). Enhancing brain tumor classification with transfer learning across multiple classes: An in-depth analysis. *BioMedInformatics*, *3*(4), 1124–1144.
20. ASIRI, A. A., SOOMRO, T. A., SHAH, A. A., POGREBNA, G., IRFAN, M., and ALQAHTANI, S. (2024). Optimized brain tumor detection: a dual-module approach for mri image enhancement and tumor classification. *IEEE Access*, *12*, 42868–42887.
21. NASSAR, S. E., YASSER, I., AMER, H. M., and MOHAMED, M. A. (2024). A robust MRI-based brain tumor classification via a hybrid deep learning technique. *The Journal of Supercomputing*, *80*(2), 2403–2427.
22. SIMO, A. M. D., KOUANOU, A. T., MONTHE, V., NANA, M. K., and LONLA, B. M. (2024). Introducing a deep learning method for brain tumor classification using MRI data towards better performance. *Informatics in Medicine Unlocked*, *44*, 101423.
23. JOSHI, A. A., and AZIZ, R. M. (2024). Deep learning approach for brain tumor classification using metaheuristic optimization with gene expression data. *International Journal of Imaging Systems and Technology*, *34*(2), e23007.
24. PACAL, I., CELIK, O., BAYRAM, B., and CUNHA, A. (2024). Enhancing EfficientNetv2 with global and efficient channel attention mechanisms for accurate MRI-Based brain tumor classification. *Cluster Computing*, *27*(8), 11187–11212.
25. ASIF, S., ZHAO, M., TANG, F., and ZHU, Y. (2023). An enhanced deep learning method for multi-class brain tumor classification using deep transfer learning. *Multimedia Tools and Applications*, *82*(20), 31709–31736.
26. RAVINDER, M., SALUJA, G., ALLABUN, S., ALQAHTANI, M. S., ABBAS, M., OTHMAN, M., and SOUFIENE, B. O. (2023). Enhanced brain tumor classification using graph convolutional neural network architecture. *Scientific Reports*, *13*(1), 14938.
27. TANDEL, G. S., TIWARI, A., KAKDE, O. G., GUPTA, N., SABA, L., and SURI, J. S. (2023). Role of ensemble deep learning for brain tumor classification in multiple magnetic resonance imaging sequence data. *Diagnostics*, *13*(3), 481