

Neural Dynamic Network for Brain Tumor Classification: An Attention-Based Feature Selection Approach

Muchammad Naseer^{*1}, Nova Agustina², Harya Gusdevi³, Niken Riyanti⁴

^{1,2,3,4}Department of Informatics, Faculty of Creative Industries, Universitas Teknologi Bandung, Bandung, Jawa Barat, Indonesia

e-mail: ^{*1}naseer@utb-univ.ac.id, ²nova@utb-univ.ac.id, ³devi@utb-univ.ac.id, ⁴niken@utb-univ.ac.id

Abstract

Magnetic Resonance Imaging (MRI) plays a vital role in the early detection of brain tumors. However, standard Convolutional Neural Network (CNN) models often struggle to extract truly relevant features from complex MRI structures. This limitation creates a gap in achieving robust and clinically interpretable classifications, as feature redundancy and weak attention toward tumor-specific regions may reduce diagnostic reliability. To address this gap, this study introduces a Neural Dynamic Network (NDN) that integrates EfficientNetV2S with a dynamic attention-based mechanism to adaptively highlight informative features while suppressing noise. The proposed model was evaluated using a 5-fold cross-validation scheme and tested on unseen data. Compared with the baseline CNN, the NDN consistently demonstrated higher accuracy, precision, recall, and F1-score across folds and final testing, reflecting improved robustness and balanced sensitivity. NDN yielded significant improvements, with the 5-fold validation averaging an accuracy of 88.44%, a precision of 87.84%, a recall of 87.88%, and an F1-score of 87.82%. Beyond numerical performance, interpretability analysis utilizing Grad-CAM demonstrated that NDN generates more concentrated and clinically consistent heatmaps. In contrast, the baseline CNN produced dispersed activations that exhibited less alignment with tumor regions. Overall, the findings confirm that incorporating a dynamic attention-based mechanism substantially enhances both feature selection and visual interpretability. This makes the NDN architecture more reliable for MRI-based brain tumor classification and highly suitable as a decision-support tool in clinical workflows.

Keywords— Neural Dynamic Network (NDN), Attention-Based Mechanism, Grad-CAM, Brain Tumor Classification

1. INTRODUCTION

Brain tumors represent one of the most critical life-threatening conditions, imposing a substantial global health burden. In 2022, approximately 322,000 new cases were reported worldwide [1]. Nevertheless, the accurate diagnosis of these tumors remains a formidable challenge. Their irregular shapes, heterogeneous structures, and proximity to complex healthy brain tissue complicate detection. Magnetic Resonance Imaging (MRI) stands as one of the primary medical imaging modalities [2] used to identify abnormalities, including brain tumors, by leveraging high-resolution imaging outputs [3], [4], [5]. MRI produces highly detailed images, which are essential for distinguishing healthy regions from abnormal tissues and enabling precise diagnosis and tumor localization [6]. Compared to CT scans, MRI offers several advantages: it is non-invasive, non-ionizing, and provides superior soft tissue contrast [6], [7]. Through multiple sequences, such as structural imaging, contrast-enhanced active tumors, edema, and fluid-adjacent edema, MRI highlights brain abnormalities from various perspectives, facilitating both anatomical and pathological analysis. Nevertheless, scan interpretation remains challenging. Radiologists frequently encounter difficulties in delineating indistinct or infiltrative tumor boundaries due to inter-observer variability and visual overlap between tumors and surrounding healthy tissues [8],

[9]. This complexity has driven the adoption of artificial intelligence-based computational methods, such as radiomics and radiogenomics, to precisely detect tumor margin infiltration and enhance diagnostic transparency [10]. Accurate AI-driven computational approaches are crucial for determining optimal treatment strategies to improve patient survival, making this a vibrant research focus, especially in deep learning [7].

Recent advancements in deep learning have accelerated research in MRI-based tumor classification, with CNNs extensively employed to automate pattern recognition and assist clinical decision-making. Prior works, such as ensemble models using VGG16, DenseNet121, and Inception-ResNet-v2, achieving around 86% accuracy [11], while ResNet-based approaches incorporating feature selection have reached 86.77% accuracy [12], demonstrate promising progress. Despite these promising results, significant challenges persist. CNNs often generate an extensive array of feature maps in each convolutional layer, many of which are redundant, noisy, or irrelevant. These weakly discriminative features reduce model generalization on heterogeneous MRI datasets, particularly when images are acquired from diverse institutions or varying acquisition protocols [13]. Furthermore, CNNs with fixed receptive fields struggle to simultaneously capture global context and fine-grained tumor details, which limits their sensitivity to small or anatomically complex lesions [14].

To address these limitations, this study proposes the Neural Dynamic Network (NDN), an attention-based deep learning architecture that integrates EfficientNetV2S with a Dynamic Attention mechanism for adaptive feature selection designed to adapt its feature selection process to each MRI input. Within the attention-based learning paradigm, the proposed approach differs from conventional static attention modules. Unlike static attention-based modules, where attention weights are fixed after training, Dynamic Attention generates attention weights conditioned on the extracted feature representations of each MRI input. This attention-based yet dynamically adaptive mechanism allows the network to highlight discriminative tumor-related features while suppressing redundant or misleading ones. Through this dynamic attention-based feature selection process, NDN aims to produce more stable classification performance and more concentrated Grad-CAM visualizations that align with clinically relevant tumor regions. The proposed model is evaluated against a baseline CNN using a five-fold cross-validation scheme and independent test data to assess its robustness, interpretability, and applicability for real-world clinical support.

In Indonesia and the broader Southeast Asian region, research on MRI-based tumor classification has predominantly focused on traditional CNN architectures or handcrafted feature extraction, with limited attention given to adaptive attention-based mechanisms. Furthermore, significant variability in MRI acquisition across local hospitals poses substantial challenges to model robustness. By introducing an adaptive attention-based framework, this study contributes to the enhancement of AI-assisted diagnostic tools in regional healthcare settings. The approach aligns with ongoing national initiatives aimed at advancing medical imaging analytics and improving early cancer detection.

Despite the success of CNN-based methods, existing models continue to encounter difficulties in adaptively selecting salient tumor-related features, leading to dispersed activation maps and reduced robustness across imaging variations. These limitations hinder clinical interpretability and practical deployment. This study aims to address these weaknesses by integrating an attention-based mechanism into a CNN architecture, allowing the model to emphasize meaningful features while suppressing irrelevant information. Specifically, this study proposes the Neural Dynamic Network (NDN), a deep learning framework built upon EfficientNetV2S combined with a dynamic attention-based mechanism. The main contributions are: (1) the development of the NDN architecture, which adaptively adjusts attention weights to highlight tumor-specific features and suppress noise; (2) comparison between the baseline CNN and the proposed NDN using 5-fold cross-validation and independent test evaluation; and (3) the integration of Grad-CAM to provide visual interpretability, enabling clinicians to verify whether the model focuses on medically relevant tumor regions.

2. LITERATURE REVIEW

To clarify the research gap and analytically position the contribution of the proposed NDN model, Table 1 presents a comparative summary of prior studies, highlighting their methodological characteristics, inherent limitations, and the manner in which the proposed approach addresses these shortcomings.

Table 1. Comparison of Related Works and the Proposed NDN Model

Author	Method	Performance	Limitation / Research Gap	Comparison to This Study
Musthafa et al. (2024)	ResNet50	Accuracy 98.52%, Precision 98–99%, Recall 97–99%, F1-score 98%	Focuses on binary classification only (tumor vs. non-tumor), not multi-class tumor recognition.	NDN performs multi-class brain tumor classification, not merely binary tumor/no-tumor classification.
Pacal et al. (2024)	Enhanced EfficientNetV2-Small dengan Global Attention Mechanism (GAM) dan Efficient Channel Attention (ECA).	99.76%, Precision 99.76%, Recall 99.75%, F1-score 99.75%	Dataset not validated with cross-validation, only train–val–test, leading to potential dataset-specific overfitting.	NDN is evaluated using 5-fold cross-validation, preventing overfitting and demonstrating a generalization gain (+1.72%).
Chaoyang et al. (2024)	FDR-TransUNet	Lung segmentation (COVID-19 Radiography Database).	FDR-TransUNet lacks interpretability; no mechanism, such as Grad-CAM, is provided.	NDN provides Grad-CAM overlays that highlight regions actually used by the model during prediction.
Iftikhar et al. (2025)	CNN (Conv–Pool–BN–Dense)	99.21% accuracy, F1-score 99.20%.	Generalization drop: –5.0% (from 99.21% to 94.72%).	NDN does not suffer from generalization drop; instead, it achieves a generalization gain of +1.72% (88.44% to 90.16%).
Jebin et. al. (2025)	U-Net	96.69% (SARTAJ), 97.31% (Br35H), 98.18% (Figshare)	Output consists of segmentation overlays (tumor contours) that only show anatomical boundaries, not feature-level interpretability.	Our study applies overlays on Grad-CAM results, not merely segmentation, allowing visualization of the actual image regions used by the model for classification.
Muksimova et al. (2025)	Dense CNN	Accuracy 98.40%, Sensitivity 98.10%, Specificity 99.21%	Interpretability exists, but not via Grad-CAM; the method does not reveal reasoning behind classification decisions.	Attention-based in NDN adapts to each input rather than relying on static, connection-wise dynamic attention.
Sánchez-Moreno et al. (2025)	Majority Voting Ensemble CNNs (VGG16, DenseNet121, Inception-ResNet-v2)	Accuracy: 86.17%, Precision: 86.29%, Recall: 85.71%, F1-score: 85.63%.	Grad-CAM++ does not influence feature selection; it only visualizes attention after the model has made a prediction.	Attention-based filters feature from the beginning of the pipeline, resulting in Grad-CAM visualizations that are more representative and clinically meaningful.

A synthesis of the studies summarized in Table 1 reveals several consistent limitations across existing deep learning approaches for brain tumor analysis. For instance, Musthafa et al. [15] demonstrate high accuracy using ResNet50; their model is restricted to binary classification, thereby limiting its applicability to multi-class tumor scenarios. While Pacal et al. achieve exceptional performance using EfficientNetV2 enhanced with GAM and ECA, their evaluation relied solely on a single train–val–test split, rendering the model vulnerable to dataset-specific overfitting [16]. Similarly, segmentation-focused architectures such as FDR-TransUNet [17] and U-Net-based

pipelines [7] often lack feature-level interpretability, providing only anatomical contours rather than insights into the underlying decision-making process. Furthermore, ensemble CNN models [11] often depend on post-hoc XAI modules such as Grad-CAM++, which explain predictions without enhancing feature extraction, consequently producing heatmaps sensitive to noise. This instability is also evident in the work of Iftikhar et al. [18], where a notable generalization drop was observed. Finally, although architectures like the Dense CNN with connection-wise attention proposed by Muksimova et al. [19] incorporate an attention mechanism, they rely on static attention, which fails to adapt to heterogeneous tumor appearances and MRI variations.

Collectively, these studies highlight clear research gaps: including limited multi-class capability, reliance on static (non-adaptive) feature selection, and dependence on post-hoc interpretability. Furthermore, the absence of intrinsic attention mechanisms often results in unstable generalization across datasets. To address these limitations, the proposed Neural Dynamic Network (NDN) introduces an attention-based mechanism that adaptively filters salient tumor features during the training process. Unlike previous approaches, the NDN produces intrinsically focused representations and more consistent Grad-CAM overlays. Experimental results demonstrate a positive generalization gain across five-fold cross-validation and test sets, effectively overcoming the shortcoming identified in prior literature.

3. RESEARCH METHODS

This study utilizes an exploratory approach by comparing two neural network architectures: a baseline CNN and the proposed Neural Dynamic Network (NDN), which incorporates EfficientNetV2S as its backbone. The main research phases comprise MRI preprocessing, feature extraction using EfficientNetV2S and Global Average Pooling (GAP), the implementation of an attention-based mechanism within the NDN, and multi-class brain tumor classification. To ensure robust results, we implemented a 5-fold cross-validation on the training dataset and evaluated performance using accuracy, precision, recall, F1-score, and confidence intervals during both validation and final testing on unseen data. The comprehensive research workflow for implementing the baseline CNN and the proposed NDN in MRI-based brain tumor classification is presented in Figure 1.

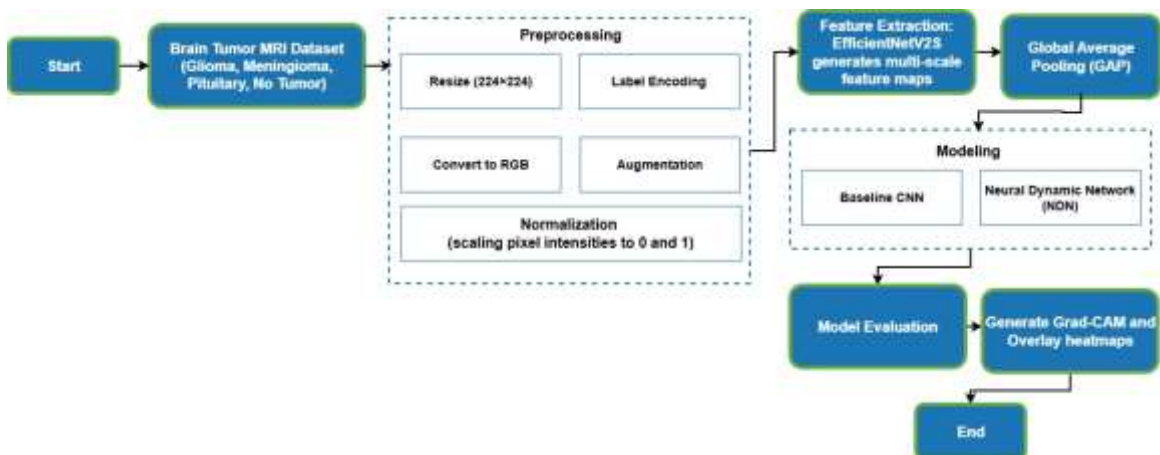


Figure 1. Research Workflow Highlighting Attention-based in the NDN Architecture.

Figure 1 illustrates the end-to-end workflow of the proposed MRI-based brain tumor classification framework, from data preprocessing to evaluation and interpretability. The pipeline consists of standardized preprocessing, feature extraction using EfficientNetV2S and Global Average Pooling, and a comparative modeling stage between a baseline CNN and the proposed Neural Dynamic Network with an attention-based mechanism. Model performance is assessed using 5-fold cross-validation and multiple classification metrics, while Grad-CAM is utilized to provide visual explanations of model decisions, thereby highlighting the contribution of salient

MRI regions to tumor classification process.

a. Data Collection

The Brain Tumor MRI dataset utilized in this study was sourced from Kaggle [20] and contains a total of 7,023 human brain MRI images. These images are categorized into four classes: glioma, meningioma, pituitary, and no tumor. The dataset is split into a training set of 5,712 images (1,321 glioma, 1,339 meningioma, 1,457 pituitary, and 1,595 no tumor) and a testing set of 1,311 images (300 glioma, 306 meningioma, 300 pituitary, and 405 no tumor). Although the original images exhibit varying resolutions, the dataset was selected for its substantial volume, balanced class labels, high-quality annotations, and public availability, all of which facilitate the replication of this study.

b. Preprocessing

Preprocessing was applied to all MRI images in this study to ensure a uniform format for model training. The preprocessing pipeline consisted of the following stages:

i. MRI Image Resizing

All MRI images were resized to 224×224 pixels. This balanced spatial detail with computational efficiency.

ii. RGB Conversion

Following resizing, the images were converted to the RGB color space. This step ensured compatibility with the input requirements for pre-trained ImageNet weights, which necessitate three color channels.

iii. Normalization

In this step, pixel values were rescaled from the $[0, 255]$ range to the $[0, 1]$ range by dividing each pixel by 255. This process stabilized the optimization during training and accelerated model convergence.

iv. Label Encoding

After normalization, label encoding was applied. Categorical labels were transformed into numerical representations using one-hot encoding across four dimensions. This ensures compatibility with the categorical cross-entropy loss during training.

v. Data Augmentation

To enhance model generalization, data augmentation was performed using geometric transformations, including horizontal flipping and 90-degree rotation. These augmentations introduced variability, rendering the model more robust to orientation changes and anatomical symmetry, without altering the underlying class labels.

c. Feature Extraction

This study utilizes EfficientNetV2S as the architectural backbone for feature extraction, selected for its proven efficacy in capturing complex patterns within medical imaging tasks. The resulting feature maps are subsequently passed to both the baseline CNN and the Neural Dynamic Network (NDN) framework. In the baseline CNN (Convolutional Neural Network), the feature maps are directly connected to the fully connected classification layer for final category prediction. For example, the feature extraction results from an MRI image at the pixel level are illustrated in Figure 2.

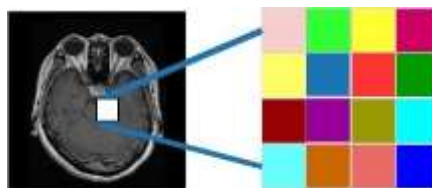


Figure 2. The original brain MRI image and extracted feature maps using EfficientNetV2S

In contrast, within the NDN, the feature maps are first processed through an attention-based mechanism (that emphasizes significant regions of the feature maps) using a dynamic attention mechanism before being passed into the fully connected layer for final classification. Each input image (X) with dimensions $X \in \mathbb{R}^{244 \times 244 \times 3}$ is mapped into a set of feature maps (F) through the EfficientNetV2S. In the convolution process, several components are involved. The indices i and j represent the spatial positions of pixels in the feature map, while u and v denote the positions of the convolution kernel. The component c refers to the input channels, such as the RGB channels, whereas k indicates the filter index in the convolutional layer. The kernel weights, denoted as $W_{u,v,c}^{(k)}$ serve as trainable parameters to generate the feature map in the- k . The activation function applied is ReLU, defined as $\text{ReLU}(x) = \max(0, x)$, where x is the convolution output obtained from the multiplication of kernel weights with the input patch plus a bias term (b). In summary, the convolution process in EfficientNetV2S for generating feature maps can be formulated as shown in Equation (1).

$$F'_{i,j,k} = \text{ReLU}(\sum_u \sum_v \sum_c W_{u,v,c}^{(k)} X_i + u, j + v, c + b_k) \quad (1)$$

The multiplication of kernel weights (W) captures local visual patterns, such as edges, tumor textures, intensity variations, and morphologies, while the summation across channels (c) aggregates information from all input channels, enabling the model to learn multi-channel feature interactions. The kernel's movement over spatial positions (i, j) maps how tumor-related patterns appear in different regions of the image. The ReLU activation suppresses negative values and retains only signals that are considered relevant for tumor representation. As the network goes deeper, the learned features become increasingly abstract, from simple edges to edema textures, to high-level tumor shape patterns. This process forms a hierarchical transformation that converts raw MRI images into meaningful semantic representations. The extracted feature maps are smaller in spatial dimensions compared to the original image but possess greater depth (i.e., a larger number of channels). This process illustrates that the deeper the network, the more complex the semantic representations obtained, which are subsequently utilized in the pooling stage.

d. Global Average Pooling (GAP)

The next stage following feature extraction is pooling, which aims to reduce the spatial dimensions of the feature maps while preserving information. In this study, pooling is implemented using Global Average Pooling (GAP). Unlike conventional pooling methods, such as max pooling, which select the maximum value from a patch, GAP computes the average value of all pixels within each channel, reducing each channel to a single value for a more compact feature vector. Let F' denote the feature maps from the convolutional process, with dimensions H (height feature maps) $\times W$ (width feature maps) $\times C$ (channels), where c is the channel index. The GAP operation to obtain scalar (z_c) for each channel is defined in Equation (2).

$$z_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W F'_{i,j,c} \quad (2)$$

Through GAP, global information from each channel is summarized, thereby reducing computational complexity before reaching the fully connected layer. This process also helps mitigate the risk of overfitting, as it prevents the retention of excessive spatial information that may not be relevant for classification. The feature vector produced by GAP serves as the input to the fully connected layers in both the baseline CNN and the attention-based model using a dynamic mechanism within the NDN.

e. Baseline CNN

The Baseline Convolutional Neural Network (CNN) in this study was designed as a

comparative model against the proposed Neural Dynamic Network (NDN) architecture. This model utilizes EfficientNetV2S as the backbone for feature extraction, after which the extracted features are processed through Global Average Pooling (GAP) and then through fully connected layers. An overview of the baseline CNN architecture is presented in Figure 3.

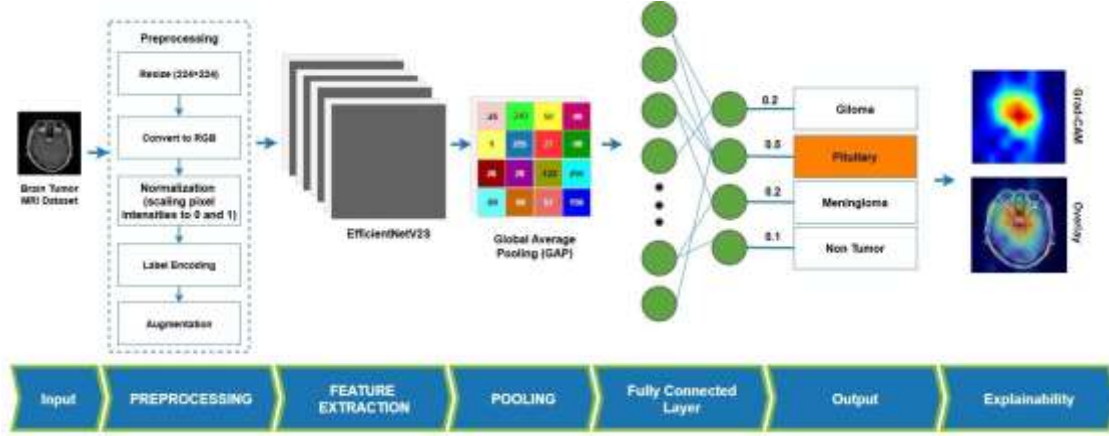


Figure 3. Baseline CNN architecture

As illustrated in Figure 3, the fully connected layers are implemented using several sequential Dense layers with the ReLU (f) activation function. Additionally, Dropout layers are inserted between the Dense layers to prevent overfitting by randomly deactivating a portion of neurons during training. Mathematically, the computation in the fully connected layer of the baseline CNN, denoted as ($h^{(1)}$), is formulated as shown in Equation (3).

$$h^{(1)} = f(W^{(1)}z + b^{(1)}) \quad (3)$$

Subsequently, the final layer is a Dense layer with a Softmax activation function (\hat{y}_k), which generates probability distributions for the four brain tumor classes: glioma, meningioma, pituitary, and non-tumor. The Softmax function operates by normalizing the exponential values of the output logits (ok). The output logit score ok is obtained from the result of Equation (4).

$$ok = W_k^{out}h^{(1)} + b_k^{out} \quad (4)$$

Thus, each output value lies within the range $[0,1]$ and the sum of all predicted probabilities equals 1. This output layer represents the likelihood that a given MRI image belongs to each class- k out of the total C classes. In summary, this computation is expressed in Equation 5.

$$\hat{y}_k = \frac{e^{ok}}{\sum_{m=1}^C e^{ok_m}} \quad (5)$$

f. Neural Dynamic Network (NDN)

The Neural Dynamic Network (NDN) extends the baseline CNN by integrating a dynamic attention-based mechanism following the feature extraction stage. Similar to the baseline model, EfficientNetV2S serves as the architectural backbone to extract feature maps, which are then aggregated using Global Average Pooling (GAP). However, unlike the baseline approach, the NDN further processes the feature maps through dynamic attention-based processing before sending them to the fully connected layer. This critical step enables the model to suppress irrelevant feature noise while simultaneously amplifying salient diagnostic patterns. Specifically, the dynamic attention-based mechanism computes an attention weight vector using two sequential non-linear operations. In the first phase, the output from the convolution layer passes through a

ReLU activation function in (FC_1) , which removes negative values. Figure 4 provides an overview of the NDN architecture.

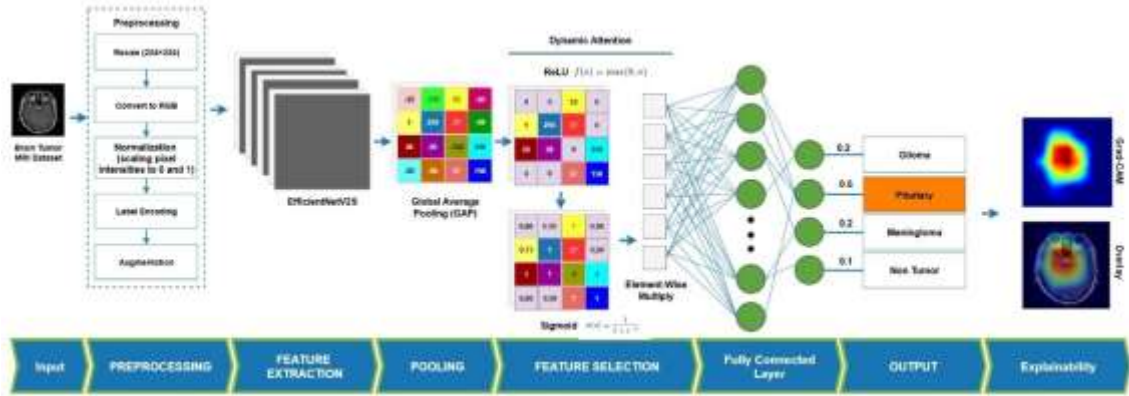


Figure 4. Neural Dynamic Network architecture using dynamic attention-based

As illustrated in Figure 4, the dynamic attention-based mechanism computes an attention weight vector using two sequential non-linear operations. First, the output from the convolution layer passes through a ReLU activation function in (FC_1) , which removes negative values. Next, the result is fed into a Sigmoid function in (FC_2) , mapping the values into the range $[0,1]$ so they can serve as attention weights. The operations performed in FC_1 and FC_2 correspond to Equations (6) and (7), respectively.

$$FC_1 = \text{ReLU}(W_{1z} + b_1) \quad (6)$$

$$FC_2 = \text{Sigmoid}(W_{2FC_1} + b_2) \quad (7)$$

After obtaining FC_1 and FC_2 , these functions are combined with the feature maps through element-wise multiplication, denoted by the symbol \odot . This process, referred to as element-wise multiplication (z'), integrates the attention weights with the feature maps. The computation of z' is formulated in Equation (8).

$$z' = z \odot FC_2 \quad (8)$$

Subsequently, the output layer of the NDN, which has passed through element-wise multiplication, is forwarded to the fully connected layer. This layer consists of several sequential Dense layers with the ReLU (f). activation function. Similar to the baseline CNN, Dropout layers are inserted between the Dense layers. The computation of the fully connected layer ($h^{(2)}$) in the NDN is formulated in Equation (9). The calculation of the output logit scores (ok) is expressed in Equation (10).

$$h^{(1)} = f(W^{(1)}z' + b^{(1)}) \quad (9)$$

$$ok = W_k^{out}h^{(1)} + b_k^{out} \quad (10)$$

Finally, in the output layer, each output value in the NDN is transformed to fall within the range $[0,1]$, ensuring that the total probability sums to 1. The final stage of the NDN process is represented in Equation (11).

$$\hat{y}_k = \frac{e^{ok}}{\sum_{m=1}^C e^{om}} \quad (11)$$

FC1 (ReLU) serves as an information compression step that reduces dimensionality and

captures the core relationships across channels. FC2 (Sigmoid) transforms the activations into attention weights ranging from 0 to 1, functioning as an importance estimation mechanism. The element-wise multiplication (\odot) performs feature refinement by amplifying tumor-related features and suppressing irrelevant ones. Dynamic attention-based recalculates the channel weights (FC2) for each MRI image individually, meaning that every image produces a new weighting pattern based on tumor location, edema intensity, and mass size. In this stage, the model learns to adaptively adjust feature emphasis according to the characteristics of the input. Unlike SE/ECA, which compute static weights that do not change across samples, dynamic attention-based updates its weights according to the MRI characteristics, for instance, assigning higher weights to texture-sensitive channels when the tumor is small, and reducing edge-sensitive channels when extensive edema is present.

g. Explainability with Grad-CAM

In general, Grad-CAM operates by leveraging the gradients of the feature maps generated by the network. In the baseline CNN, gradients are computed with respect to the feature maps from the final convolutional layer. This occurs before Global Average Pooling (GAP). In contrast, in the NDN, gradients are computed with respect to the feature maps obtained after feature selection in the dynamic attention-based layer. These gradients represent the contribution of each channel to the target class. They are then averaged to obtain the importance weight of each channel. These weights are subsequently used to recombine the channels in the feature maps, producing the class activation map, also known as a heatmap. The heatmap is normalized to the range $[0,1]$, resized to match the original image resolution (224×224), and then overlaid on the MRI image to highlight the model's focus areas. To generate Grad-CAM, the first step is to compute the class score gradient (α_k^c) with respect to the feature maps. This can be formulated using Equation (12) for the baseline CNN and Equation (13) for the NDN.

$$\alpha_k^c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W \frac{\partial o^c}{\partial z_c} \quad (12)$$

$$\alpha_k^c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W \frac{\partial o^c}{\partial z'} \quad (13)$$

In Equations (12) and (13), α_k^c is obtained by dividing the sum of gradients by $H \times W$ of the feature map. This value is then multiplied by the total number of spatial units in each channel. The process is similar to average pooling, but it pools gradients instead of activation values. In the baseline CNN, the gradient $\frac{\partial o^c}{\partial z_c}$ represents the sensitivity of the class logit score to changes in the activation value at position (i,j) in the- k channel after feature selection. The next step is to construct the Class Activation Map (CAM). This is achieved by multiplying each channel of the feature maps by its corresponding weight α_k^c . Then, sum the results across all channels. This operation produces an activation map that highlights the MRI regions most important to class prediction (c). To keep only positive contributions, apply the ReLU activation function to the sum. Mathematically, this is shown in Equation (14) for the baseline CNN, and Equation (15) for the NDN.

$$L_{Grad-CAM}^c = ReLU(\sum_k \alpha_k^c z_c) \quad (14)$$

$$L_{Grad-CAM}^c = ReLU(\sum_k \alpha_k^c z') \quad (15)$$

h. Architecture and Training Setup

In this study, three fully connected layers were used sequentially. The first is a Dense layer with 512 units and ReLU activation, accompanied by a Dropout rate of 0.3 to prevent overfitting. Next is a Dense layer with 128 units and ReLU activation. The final output layer is a Dense layer with 4 units and a Softmax activation function, corresponding to the number of target classes. The training process employed a 5-fold stratified K-fold cross-validation strategy. This

helped ensure more stable performance evaluation and minimized bias in data partitioning. Training also included an Early Stopping mechanism (patience 10, monitor val_loss) to prevent overfitting. The AdamW optimizer was used with an initial learning rate of 1×10^{-4} along with categorical cross-entropy as the loss function. Model evaluation was conducted using accuracy, precision, recall, and F1-score metrics. This provided a more comprehensive assessment of performance across each validation fold. The training scheme included a batch size of 32 and a maximum of 100 epochs. To further mitigate overfitting, early stopping was applied (patience = 10). Adaptive learning rate adjustment was performed with ReduceLROnPlateau (reduction factor = 0.2, minimum learning rate 1×10^{-6}). Model checkpointing was used to store the best weights based on validation accuracy.

i. Computational Environment Specifications

All experiments in this study were conducted in a computing environment equipped with an NVIDIA A100 Tensor Core GPU, using Python 3 as the programming language. The system was configured with 83.5 GB of RAM, 40.0 GB of GPU memory, and a storage capacity of 235.7 GB.

4. RESULT AND DISCUSSION

After preprocessing, feature extraction, model training, and applying the dynamic attention-based mechanism, the next step was to evaluate model performance. This section presents the experimental results from the baseline CNN and the Neural Dynamic Network (NDN). Results are presented under the 5-fold cross-validation scheme and from final testing on held-out test data. Quantitative evaluation used accuracy, precision, recall, and F1-score metrics. Model interpretability was also analyzed with Grad-CAM to provide deeper insights into regions of focus in MRI-based brain tumor classification. To reduce repetition, the overall trends are summarized visually in Figures 5–10 and supported with statistical testing and analytical interpretation to address reviewer feedback.

a. Model Performance

In the baseline model, the Early Stopping mechanism was triggered in Folds 1 (epoch 54), 2 (epoch 51), 4 (epoch 72), and 5 (epoch 80). This means that after these points, the validation loss no longer showed significant improvement for 10 consecutive epochs. As a result, the training process stopped earlier. In contrast, in Fold 3, Early Stopping was not triggered, so training continued for the full 100 epochs. The activation of Early Stopping in the baseline model suggests the validation loss had plateaued. Meanwhile, training performance continued to improve without corresponding validation gains, indicating the onset of overfitting.

In the NDN model, Early Stopping was not activated. The validation loss consistently improved until it approached the maximum number of epochs. This demonstrates that the model equipped with the dynamic attention-based mechanism maintained a stable downward trend in loss. As a result, it did not meet the criteria for early termination. This also indicates the NDN is more resistant to overfitting symptoms and has better generalization capability compared to the baseline CNN. For clarity, Figure 5 presents the average validation performance per fold for the baseline CNN. Figure 6 illustrates the results obtained from the proposed Neural Dynamic Network (NDN).

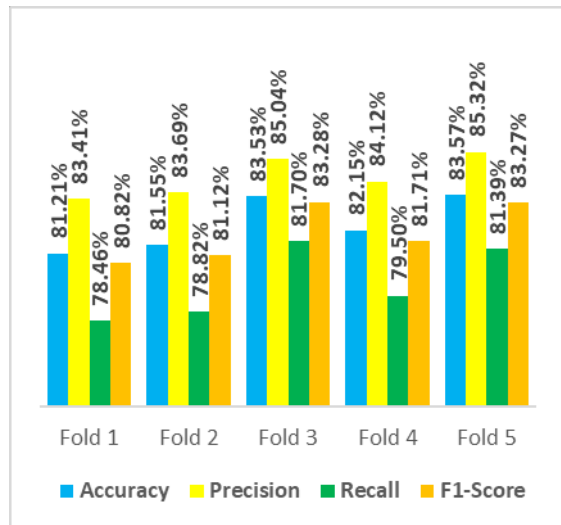


Figure 5. Average Validation Performance per Fold

(Baseline CNN)

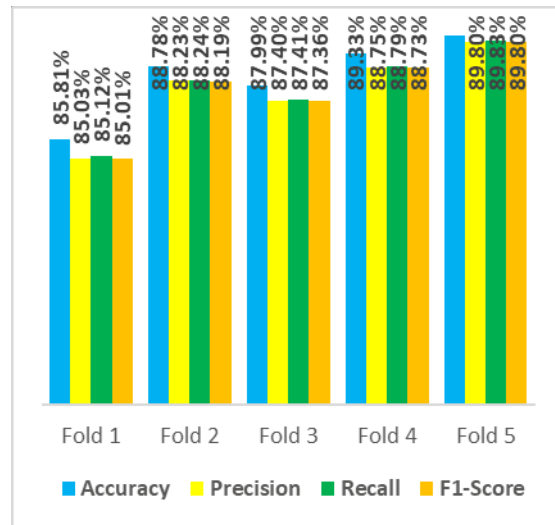


Figure 6. Average Validation Performance per Fold

(NDN)

In the baseline CNN (Figure 5), the validation accuracy ranged from 81.12% to 83.56%, with an average of approximately 82%. Precision ranged from 80.82% to 85.32%. Recall varied from 78.46% to 81.71%. The F1-score stayed stable, ranging from 80.28% to 83.41%. These results indicate the baseline CNN could extract features, but still struggled to maintain consistent recall. This suggests the model's sensitivity to true positives was relatively low, limiting its ability to detect all tumor classes evenly.

Conversely, the proposed NDN model (Figure 6) achieved higher performance in all folds. Validation accuracy ranged from 85.81% to 88.90%. Precision and recall consistently exceeded 85%. The F1-score remained above 87%, reaching 88.90% in the best fold. This improvement can be attributed to the dynamic attention-based mechanism, which emphasizes tumor-relevant features and suppresses noisy ones. These findings confirm NDN maintains a better balance between precision and recall, yielding a significant improvement in F1-score compared to the baseline CNN. For completeness, Table 1 summarizes the average results, while Figures 5–6 provide fold-wise visual summaries to streamline result presentation and avoid repetitive textual descriptions.

Table 1. Comparative Results of Baseline-Based Model (CNN) and NDN

Performances	Based-Model (CNN)		NDN	
	Training	Testing	Training	Testing
Accuracy	82.66%	84.89%	88.44%	90.16%
Precision	84.78%	85.85%	87.84%	90.39%
Recall	79.90%	84.21%	87.88%	89.70%
F1-Score	82.18%	77.48%	87.82%	89.41%

Table 1 shows that the baseline CNN achieved an average accuracy of 82.66% across all folds during training, with a precision of 84.78%, a recall of 79.90%, and an F1-score of 82.18%. In contrast, the NDN model demonstrated consistent improvements across all average metrics during training, with an accuracy of 88.44%, precision of 87.84%, recall of 87.88%, and F1-score of 87.82%. These results confirm that the dynamic attention-based mechanism in NDN enhanced the quality of feature representation during training, making the model not only more accurate but also more balanced in its ability to identify both positive and negative cases. In testing, NDN consistently outperformed the baseline CNN, improving test accuracy from 84.89% (CNN) before, to 90.16%, precision from 85.85% to 90.39%, recall from 84.21% to 89.70%, and F1-score from 77.48% to 89.41%. These improvements demonstrate that integrating the dynamic attention-based mechanism into NDN enhances feature selection, resulting in a more robust model

with superior generalization for MRI-based brain tumor classification. The next evaluation involves measuring model loss for CNN and NDN in the best-performing fold (Fold 5), illustrated in Figures 7 and 8.

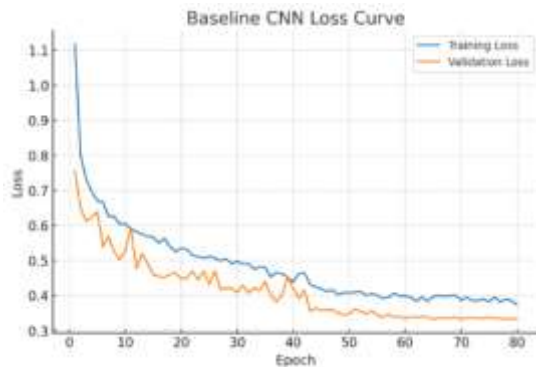


Figure 5. Loss Curve of the Baseline CNN

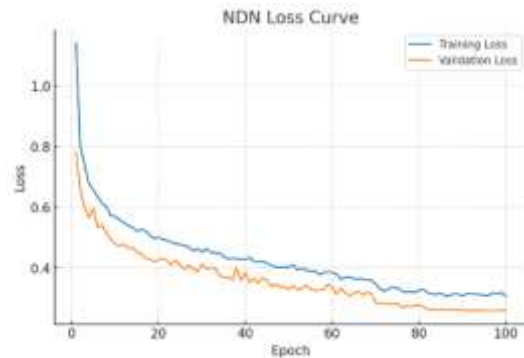


Figure 6. Loss Curve of the NDN

In Figure 7, the baseline CNN exhibits a validation loss curve that appears fairly fluctuating during the initial epochs before gradually converging. Although the training loss decreases progressively, a noticeable gap remains between the training and validation curves, indicating that the model tends to be less consistent in generalizing. Conversely, as shown in Figure 8, the NDN demonstrates a smoother and more stable convergence pattern. Both training loss and validation loss consistently decrease throughout the epochs, with a smaller gap between them compared to the baseline CNN. This pattern highlights the effectiveness of the dynamic attention-based mechanism in enhancing feature selection and suppressing irrelevant information, thereby enabling the model to generalize better and deliver more reliable performance on the test data. Next, to evaluate the dynamic aspects of this study, i.e., the extent to which the application of GAP, Fully Connected Layers, and dynamic attention-based influences model performance, the results are summarized in Table 2.

Table 2. Ablation Study Results of CNN Variants and Full NDN

Model Variant	Accuracy	Precision	Recall	F1-Score
Baseline CNN with GAP and Fully Connected Layer	82.66%	84.78%	79.90%	82.18%
Baseline CNN without GAP and Fully Connected Layer	23.39%	5.850%	25.00%	9.480%
Full NDN with dynamic attention-based, CNN, GAP, and FC Layers	88.44%	87.84%	87.88%	87.82%

In Table 2, the ablation study shows that the Global Average Pooling (GAP) and Fully Connected Layers are essential components in the CNN architecture. When both components are removed, the model's performance drops drastically, with accuracy falling to only 23.39% and the F1-score to 9.48%, indicating that the model is almost unable to perform classification. When the GAP and FC layers are reinstated, performance improves, achieving an accuracy of 82.66%, demonstrating that a standard CNN can still recognize basic patterns in MRI images. However, the greatest improvement is obtained when a dynamic attention-based model is integrated into the architecture (NDN), increasing accuracy to 88.44% along with improvements in precision, recall, and F1-score. This indicates that the feature selection mechanism in NDN is able to highlight relevant features and suppress noise, resulting in more stable representations and greater sensitivity to tumor presence. Overall, the ablation study shows that each architectural component, especially the dynamic attention-based, contributes to improving model performance. To further evaluate the model's generalization capability, a cross-dataset experiment was conducted using two different binary MRI datasets, i.e., the Brain Tumor MRI Dataset (MRI Dataset 4 Class) [20] and the Mendeley MRI Dataset (MRI Dataset 2 Class) [21], as shown in Table 3.

Table 3. Cross-Dataset Classification Performance Between the Mendeley MRI Dataset (Binary) and the Kaggle Brain Tumor MRI Dataset (Binary)

Model	Train Dataset	Test Dataset	Accuracy	Precision	Recall	F1-Score
CNN	MRI Dataset 4 Class	MRI Dataset 2 Class	76.98%	69.58%	95.70%	80.58%
CNN	MRI Dataset 2 Class	MRI Dataset 4 Class	94.32%	95.67%	97.30%	96.48%
NDN	MRI Dataset 4 Class	MRI Dataset 2 Class	79.52%	71.92%	96.69%	82.49%
NDN	MRI Dataset 2 Class	MRI Dataset 4 Class	93.91%	96.95%	94.47%	95.69%

The test results presented in Table 3 show that when the model was trained on the dataset containing 4 classes and tested on the dataset containing 2 classes, both CNN and NDN experienced a decrease in accuracy (76.98% and 79.52%). Nevertheless, both models still produced very high recall values (95.70% for CNN and 96.69% for NDN). This condition indicates that the model becomes highly sensitive to the presence of tumors but lacks precision in distinguishing positive and negative cases. The cause is the characteristics of the MRI Dataset 4 Class, which has higher variations in intensity and noise, causing the model to generalize by predicting “tumor” more frequently, resulting in lower precision and reduced accuracy. Conversely, when the model was trained using the dataset with 2 classes and tested on the dataset with 4 classes, performance improved significantly. CNN achieved an accuracy of 94.32% and NDN 93.91%, with precision, recall, and F1-score all above 94%. The dataset with 2 classes contains cleaner and more uniform image quality, allowing the model to learn tumor patterns more stably. These learned patterns can then be transferred effectively when the model is exposed to the more varied Dataset 4. In this scenario, the performance difference between CNN and NDN is not very large, because the high quality of the training data already sufficiently supports CNN’s learning process. However, NDN still maintains an advantage in prediction stability, as seen from the highest precision value of 96.95% when tested on the dataset with 4 classes.

However, the MRI Dataset 2 Class contains only two classes, i.e., tumor and non-tumor, so models trained on this dataset tend to learn simpler and less diverse patterns. The binary class setting makes the learning process more stable and less noisy, but also limits the model’s understanding of more complex tumor morphology variations. As a result, when tested on MRI Dataset 4, which has more diverse image characteristics (such as intensity variations, tumor shape variations, and anatomical background differences), the model still performs well because the main tumor pattern has been learned, but it does not experience a significant performance boost from the use of dynamic attention-based. This condition explains why the performance gap between CNN and NDN becomes relatively small in the training scenario using MRI Dataset 2 Class.

Overall, this cross-dataset evaluation demonstrates that NDN exhibits more stable performance than CNN, particularly when the training data originate from a noisier dataset. Models trained on high-quality datasets are able to transfer their learned knowledge to other datasets more effectively. These findings reinforce the importance of training data quality in influencing the generalization ability of MRI-based brain tumor classification models.

b. Grad-CAM

This study used Gradient-weighted Class Activation Mapping (Grad-CAM) to enhance model interpretability. Grad-CAM generated heatmaps of the MRI regions contributing most to classification. The activation maps from the CNN were dispersed and less focused on tumor regions. CNN heatmaps showed broad attention, making it difficult to confirm if true pathological areas were fully highlighted. This aligns with the lower recall performance in the baseline CNN (79.90% in training, 84.21% in testing), indicating limited sensitivity in detecting positives. Thus, the CNN could extract discriminative features, but its interpretability remained limited. This may raise concerns in clinical applications. Figure 9 illustrates the baseline CNN results.

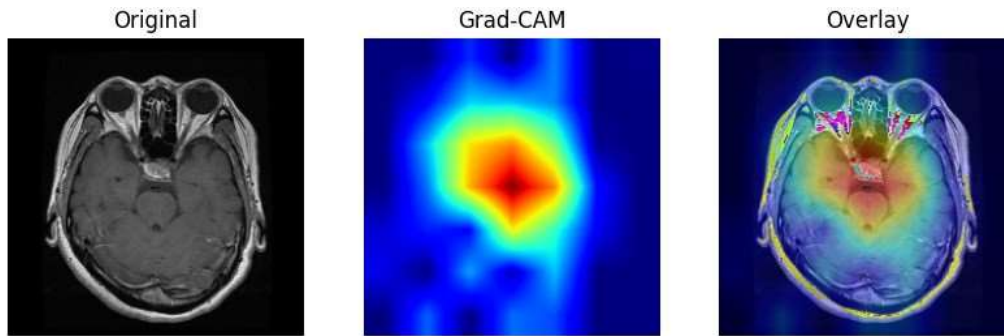


Figure 9. Grad-CAM of the Baseline CNN

In contrast, Figure 10 presents the NDN results, showing a more focused and consistent activation in the tumor regions, especially in the pituitary area, which matches the image label. The generated heatmaps are more directed, with high intensity only in target regions. Other image areas are relatively suppressed. This confirms that the dynamic attention-based mechanism in NDN can adaptively select important features, suppress noise, and guide the model to relevant areas. Higher recall and F1-scores for NDN compared to CNN further support this, indicating a better balance between sensitivity and precision.

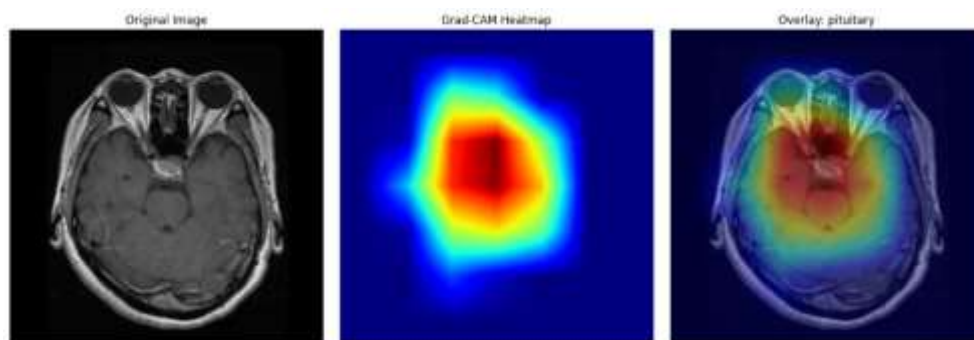


Figure 10. Grad-CAM of the NDN

Overall, this comparison demonstrates that integrating Grad-CAM with NDN not only enhances numerical performance but also provides more focused visual interpretability. Therefore, NDN is more suitable for supporting clinical diagnostic processes, as its predictions can be validated through the visualization of tumor regions in MRI images.

c. Discussion

The experimental results show that the Neural Dynamic Network (NDN) consistently outperforms the baseline CNN across accuracy, precision, recall, and F1-score in both cross-validation and final testing. This study also conducted statistical significance tests to verify whether the performance improvements achieved by NDN were truly meaningful and not due to random variation. Based on the fold-wise results from the 5-fold cross-validation scheme, the paired t-test and Wilcoxon signed-rank test produced p-values of 0.0633 (accuracy), 0.1284 (precision), 0.1418 (recall), and 0.1328 (F1-score) for the paired t-test, and 0.1250, 0.1875, 0.3125, and 0.1875, respectively, for the Wilcoxon test. Since all p-values were greater than 0.05, these performance improvements cannot be considered statistically significant. This outcome reflects the limited sample size and the low variance across folds, two common conditions that reduce the statistical power of significance testing in deep learning experiments. However, NDN still outperformed CNN on all folds, demonstrating its practical advantage despite the lack of statistical significance.

Next, the study evaluated the confidence scores, which represent the average softmax

probability assigned to the correct class. The higher confidence achieved by NDN (93.72% compared to 88.79% for CNN) indicates that NDN produces more stable predictions. From a performance interpretation perspective, the improvement introduced by the dynamic attention-based approach arises from the model's ability to adaptively adjust feature response weights for each MRI input. Rather than treating all extracted features equally, dynamic attention-based strengthening activates related to tumor regions while suppressing irrelevant or noisy signals. This results in clearer class separation, better sensitivity to subtle tumor structures, and more consistent predictions across folds.

This functional benefit is further supported by the interpretability results. The Grad-CAM maps produced by NDN are more focused on the actual tumor regions, indicating that the model is not only more accurate but also more aligned with meaningful spatial patterns. Furthermore, the cross-dataset evaluation provides additional evidence regarding the robustness of NDN. When trained on a noisier and more heterogeneous dataset (MRI Dataset 4 Class), both CNN and NDN showed reduced accuracy but maintained extremely high recall when tested on MRI Dataset 2 Class. This indicates a strong tumor-sensitivity bias, suggesting that models exposed to high-variation data tend to overpredict positive cases. In contrast, when trained on the cleaner MRI Dataset 2 Class and tested on the more complex Dataset 4 Class, both models achieved substantially higher overall performance, with NDN maintaining the highest precision. This demonstrates that NDN benefits more clearly in scenarios involving noisy or heterogeneous training data, where dynamic attention-based effectively suppresses irrelevant activations.

These findings collectively illustrate that dynamic attention-based contributes not only to better numerical performance but also to improved stability across datasets and clearer localization of tumor regions. This has practical implications for clinical decision support, as models that generalize more consistently across datasets are more reliable when deployed across different hospitals or imaging protocols. However, the study still has limitations. Ablation studies have not been performed to isolate the specific contribution of dynamic attention-based relative to other architectural components.

5. CONCLUSION

This study proposes the Neural Dynamic Network (NDN), a modification of EfficientNetV2S with the addition of a dynamic attention-based mechanism to improve MRI-based brain tumor classification performance. The model demonstrates consistent performance improvements compared to the baseline CNN, both in accuracy, sensitivity (recall), and prediction stability. The main contribution of this study is showing that dynamically weighting features for each input, rather than applying static weights across the entire dataset, can reduce irrelevant activations, strengthen tumor representation, and stabilize the learning process. This mechanism provides a new direction in attention design for CNNs by emphasizing contextual weight adjustment without increasing architectural complexity. Although statistical significance tests (paired t-test & Wilcoxon) did not yield p-values < 0.05 due to the small number of folds and low fold-to-fold variance, the performance improvements remain consistent across all folds. In addition, NDN provides a higher prediction confidence level (93.72% compared to 88.79% for CNN) and more focused Grad-CAM attention regions aligned with actual tumor locations. These results enhance the model's reliability, an important aspect in clinical application. Furthermore, the cross-dataset experiment further supports this robustness, showing that NDN maintains more stable precision than CNN when trained on noisier data and tested on an external dataset.

This study has several limitations. First, the ablation studies were not conducted to isolate the specific contribution of dynamic attention-based, and the Grad-CAM evaluation remains qualitative without quantitative metrics. Future work should include component-level ablation, multi-institutional validation, and quantitative interpretability measures such as overlap ratio (IoU) and pointing-game accuracy to empirically verify the model's attention quality. Evaluating

multimodal MRI inputs and integrating clinical metadata also represent promising directions to further improve robustness and clinical applicability.

6. ACKNOWLEDGMENTS

The research was funded under the Main Contract dated 28 May 2025 (Contract No. 125/C3/DT.05.00/PL/2025) and supported through the Derivative Contracts dated 04 June 2025 (Contract No. 8078/LL4/PG/2025) and 05 June 2025 (Contract No. 049/WR-RICP/UTB/VI/2025).

REFERENCES

- [1] Filho M Adalberto et al., “Cancers of the brain and central nervous system: global patterns and trends in incidence.,” *J Neurooncol*, vol. 172, pp. 567–578, May 2025, doi: 10.1007/s11060-025-04944-y.
- [2] M. S. Hasibuan, “Integrating Convolutional Neural Networks into Mobile Health: A Study on Lung Disease Detection,” *Journal of Applied Data Sciences*, vol. 6, no. 3, pp. 1495–1503, Sep. 2025, doi: 10.47738/jads.v6i3.660.
- [3] A. ul Haq, J. P. Li, S. Khan, M. A. Alshara, R. M. Alotaibi, and C. Mawuli, “DACBT: deep learning approach for classification of brain tumors using MRI data in IoT healthcare environment,” *Sci Rep*, vol. 12, no. 1, p. 15331, Sep. 2022, doi: 10.1038/s41598-022-19465-1.
- [4] A. Alshuhail et al., “Refining neural network algorithms for accurate brain tumor classification in MRI imagery,” *BMC Med Imaging*, vol. 24, no. 1, p. 118, May 2024, doi: 10.1186/s12880-024-01285-6.
- [5] M. M. Zahoor et al., “Brain Tumor MRI Classification Using a Novel Deep Residual and Regional CNN,” *Biomedicines*, vol. 12, no. 7, p. 1395, Jun. 2024, doi: 10.3390/biomedicines12071395.
- [6] R. Khan et al., “High-precision brain tumor classification from MRI images using an advanced hybrid deep learning method with minimal radiation exposure,” *J Radiat Res Appl Sci*, vol. 18, no. 4, p. 101858, Dec. 2025, doi: 10.1016/j.jrras.2025.101858.
- [7] B. M. Jebin, S. I. Shyla, K. N. Sujantha Bel, and C. J. Jeba Sheela, “Brain tumor detection and classification using u-net and CNN with brain texture pattern analysis,” *Biomed Signal Process Control*, vol. 110, p. 108156, Dec. 2025, doi: 10.1016/j.bspc.2025.108156.
- [8] A. J. Aiya et al., “Optimized deep learning for brain tumor detection: a hybrid approach with attention mechanisms and clinical explainability,” *Sci Rep*, vol. 15, no. 1, p. 31386, Aug. 2025, doi: 10.1038/s41598-025-04591-3.
- [9] H. Fan et al., “Artificial intelligence-based MRI radiomics and radiogenomics in glioma,” *Cancer Imaging*, vol. 24, no. 1, p. 36, Mar. 2024, doi: 10.1186/s40644-024-00682-y.
- [10] Y. Wang, Z. Hu, and H. Wang, “The clinical implications and interpretability of computational medical imaging (radiomics) in brain tumors,” *Insights Imaging*, vol. 16, no. 1, p. 77, Mar. 2025, doi: 10.1186/s13244-025-01950-6.

- [11] L. Sánchez-Moreno, A. Perez-Peña, L. Duran-Lopez, and J. P. Dominguez-Morales, “Ensemble-based Convolutional Neural Networks for brain tumor classification in MRI: Enhancing accuracy and interpretability using explainable AI,” *Comput Biol Med*, vol. 195, p. 110555, Sep. 2025, doi: 10.1016/j.compbimed.2025.110555.
- [12] A. Agrawal and J. Chaki, “CerebralNet meets Explainable AI: Brain tumor detection and classification with probabilistic augmentation and a deep learning approach,” *Biomed Signal Process Control*, vol. 110, p. 108210, Dec. 2025, doi: 10.1016/j.bspc.2025.108210.
- [13] F. Yuan, Z. Tang, C. Wang, Q. Huang, and J. Shi, “A multiple gated boosting network for multi-organ medical image segmentation,” *IET Image Process*, vol. 17, no. 10, pp. 3028–3039, Aug. 2023, doi: 10.1049/ipr2.12852.
- [14] R. R. Selvaraju, A. Das, R. Vedantam, M. Cogswell, D. Parikh, and D. Batra, “Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization,” *Int J Comput Vis*, vol. 128, pp. 336–359, 2016, [Online]. Available: <https://api.semanticscholar.org/CorpusID:15019293>
- [15] M. M. M, M. T. R, V. K. V, and S. Guluwadi, “Enhancing brain tumor detection in MRI images through explainable AI using Grad-CAM with Resnet 50,” *BMC Med Imaging*, vol. 24, no. 1, p. 107, May 2024, doi: 10.1186/s12880-024-01292-7.
- [16] 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.
- [17] Z. Chaoyang, S. Shibao, H. Wenmao, and Z. Pengcheng, “FDR-TransUNet: A novel encoder-decoder architecture with vision transformer for improved medical image segmentation,” *Comput Biol Med*, vol. 169, p. 107858, Feb. 2024, doi: 10.1016/j.compbimed.2023.107858.
- [18] S. Iftikhar, N. Anjum, A. B. Siddiqui, M. Ur Rehman, and N. Ramzan, “Explainable CNN for brain tumor detection and classification through XAI based key features identification,” *Brain Inform*, vol. 12, no. 1, p. 10, Dec. 2025, doi: 10.1186/s40708-025-00257-y.
- [19] S. Muksimova, S. Umirzakova, N. Iskhakova, A. Khaitov, and Y. I. Cho, “Advanced convolutional neural network with attention mechanism for Alzheimer’s disease classification using MRI,” *Comput Biol Med*, vol. 190, p. 110095, May 2025, doi: 10.1016/j.compbimed.2025.110095.
- [20] Msoud Nickparvar, “Brain Tumor MRI Dataset [Data set],” 2021, Kaggle: 1. doi: doi.org/10.34740/kaggle/dsv/2645886.
- [21] A. S. H. Emran, H. Akter, and A. Al Shiam, “Brain Tumor MRI Dataset,” 2025, Mendeley Data: