# CDANET: CHANNEL SPLIT DUAL ATTENTION BASED CNN FOR BRAIN TUMOR CLASSIFICATION IN MR IMAGES

*Tapas Kumar Dutta*[†] *Deepak Ranjan Nayak*[⋆]

[†] Department of IT, Indian Institute of Engineering Science and Technology, Shibpur, India
[⋆] Department of CSE, Malaviya National Institute of Technology, Jaipur, India

## ABSTRACT

Brain tumor is the most common type of cancer that causes a high mortality rate among individuals of all age groups. Hence, accurate diagnosis of brain tumor and its type at early stages is of utmost importance to preclude its severity and ultimately helps patients for timely and better treatment. The current literature has witnessed the usage of convolutional neural networks (CNN) for brain tumor classification in MR images; however, such traditional CNN methods may fail to identify the minute variations of tumor lesions. This paper proposes an attention module called channel split dual attention (CSDA) coupled with a backbone network to handle this issue. The CSDA explores more detailed and discriminative features using channel splitting and two parallel attention blocks: position attention block (PAB) and channel attention block (CAB). The PAB and CAB are introduced to capture feature dependencies in spatial and channel dimensions. Extensive experiments on a publicly available dataset show that our CDANet significantly improves the state-of-the-art CNN results and obtains higher classification accuracy than existing brain tumor detection methods.

***Index Terms***— Brain tumor classification, Channel split dual attention, Position attention, Channel attention, CNN.

## 1. INTRODUCTION

Brain tumors are the most common cause of cancer-related death in individuals of all age groups, including children across the globe [1]. Hence, early detection of brain tumors plays a crucial role in better treatment plans and patient care, thereby enhancing the survival rate. However, the presence of different varieties of brain tumors such as meningioma, glioma, and pituitary, and their similar appearances make the detection task more challenging. A non-invasive and radiation-free imaging technique called MRI has been profoundly used to identify brain tumors since it provides a more detailed picture of brain tissues. But, manual inspection of MR images at a larger scale to detect brain tumors is time-consuming, expert-dependent, and error-prone. Thus, researchers have shown a significant rise in interest in designing accurate and automated computer-aided detection systems for brain tumor classification.

The earlier brain tumor detection systems usually include multiple stages such as preprocessing, feature extraction and selection, dimensionality reduction, and classification. Feature extraction and classification are the two crucial stages in these systems. Usman et al. [2] integrated various features like intensity, local neighborhood, intensity differences, and wavelet texture extracted from brain MR images and used random forest classifier for tumor prediction. Gupta et al. [3] employed a fusion of texture features computed using local binary pattern (LBP), gray level co-occurrence matrices (GLCM), and gray level run length matrices (GLRLM), and an ensemble classifier to achieve higher classification performance. However, choosing an appropriate feature extraction method and a classifier in these systems remains challenging.

On the other hand, deep learning techniques, specifically convolutional neural networks (CNNs), help to automatically learn high-level features and representations from input images without any domain knowledge. Hence, CNNs have been extensively applied in medical image classification tasks and have recently become popular for brain tumor classification [1]. Paul et al. [4] used both fully connected and CNN for brain tumor classification. Afshar et al. [5] investigated the effectiveness of capsule networks (CapsNets) for brain tumor classification. Swati et al. [6] proposed a block-wise fine-tuning (BFT) strategy and used a pretrained VGG19 model to classify brain tumor types. Recently, Ghassemi et al. [7] trained a generative adversarial network (GAN) on a different set of MR images and then used a deep neural network as a discriminator in GAN to distinguish three tumor classes. Although CNN-based brain tumor detection methods have obtained better performance than traditional methods, their practical application is still challenging due to the following reasons. First, the tumor size varies greatly depending on the type of brain tumor, and traditional CNNs may fail to capture the subtle variations of tumor lesions. Second, different tumor types are similar in appearance.

Attention mechanisms have recently become a crucial

This work is partially supported by the Science and Engineering Research Board, Department of Science and Technology, Govt. of India under project No. SRG/2020/001460.

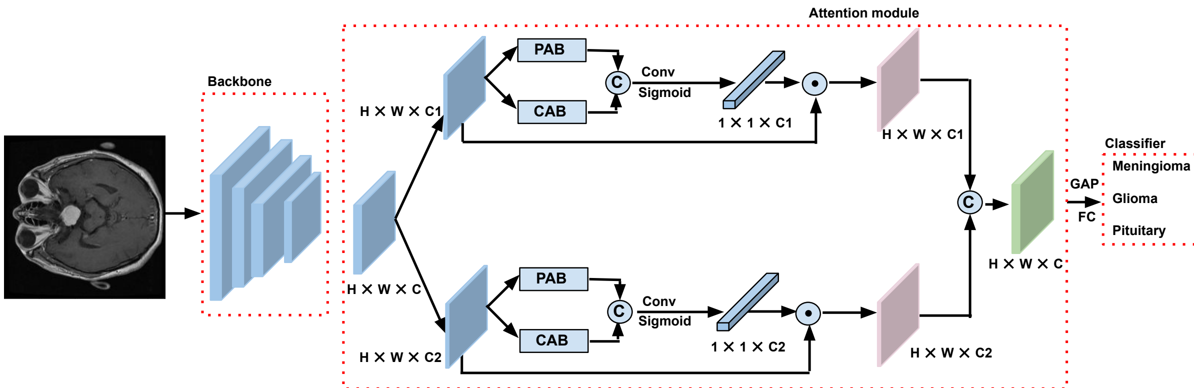


**Fig. 1**. Overall architecture of the proposed CDANet

component to enhance the performance of CNN, which enables to focus on essential features of the input while suppressing the irrelevant information [8, 9, 10]. Hence, in this paper, we introduce an attention module called channel split dual attention (CSDA) that captures more detailed features for effective brain tumor classification. The proposed attention module overcomes the above issues by learning fine-grained features and preserving the brain tumor lesion information. The key contributions of this paper are outlined as follows: (1) We introduce an attention module called CSDA on the top of a backbone network using channel splitting and two parallel attention blocks, (2) We split the feature map into two groups along the channel dimension and for each group we propose to use position attention block (PAB) and channel attention block (CAB) simultaneously to learn more detailed small brain tumor lesion features. Then, we aggregate the features from each group to obtain a refined feature map, and (3) We compare the effectiveness of the proposed attention module with popular attention mechanisms (e.g., CBAM [10] and SE [11]) and state-of-the-art brain tumor detection approaches.

## 2. PROPOSED METHODOLOGY

As shown in Fig. 1, the overall architecture of the proposed CDANet contains three major components: a backbone network, attention module, and classifier. The attention module called CSDA block (CSDAB) includes PAB and CAB to learn dependencies of features in both spatial and channel dimensions. The classification stage contains a global average pooling (GAP) layer followed by a fully connected (FC) layer at the end. This section discusses the proposed network and its components in detail.

### 2.1. Backbone Network

We take an MRI image as input to the proposed CDANet. Then, we use an ImageNet pre-trained CNN as a backbone network which facilitates obtaining the global feature maps $F \in \mathbb{R}^{H \times W \times C}$, where C, W, and H indicate the number of channels, width, and height of the feature maps, respectively. It is worth mentioning that the feature maps $F \in \mathbb{R}^{H \times W \times C}$ have been extracted from the final convolutional layer of the pre-trained CNN model since they incorporate high-level feature representations and are given as input to the attention block. We demonstrate the effectiveness of CDANet using various pre-trained CNNs in Section 3.3.

### 2.2. Channel Split Dual Attention Block (CSDAB)

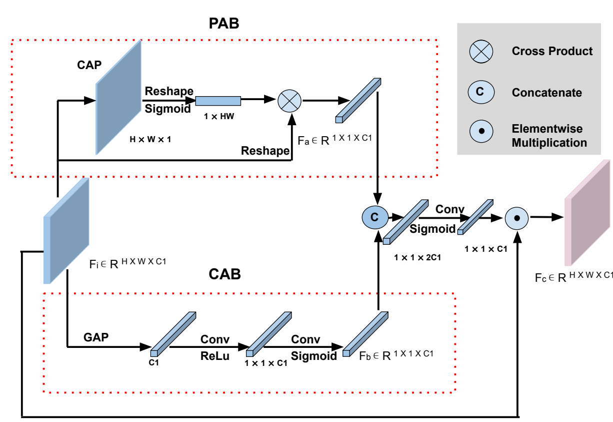


**Fig. 2**. An overview of the proposed CSDA block.

The channel split dual attention block (CSDAB) is designed to learn more detailed and discriminative features from the brain tumor lesions. The CSDAB first splits the feature maps into two groups and then uses two parallel attentions (position and channel attention) on each group of feature maps to obtain more focussed feature representations. Finally, the refined feature maps are concatenated. Fig. 2 depicts the overview of the proposed CSDA block.

#### 2.2.1. *Channel Splitting*

Let $F \in \mathbb{R}^{H \times W \times C}$ denote the feature map obtained from the last convolutional layer of a pre-trained CNN. Then, we first split $F$ into two groups along the channel dimension, that is, $F = [F_1, F_2]$, $F_i = \mathbb{R}^{H \times W \times C/2}$, where each group helps in learning more specific tumor lesion features. Each $F_i$ is then fed to two attention blocks, PAB and CAB, simultaneously to capture pixel-wise and channel-wise dependencies, which are concatenated to enrich the feature representations further.

#### 2.2.2. *Channel Attention Block (CAB)*

All channels of the high-level feature maps may not preserve important class-specific information, which can be seen experimentally in Section 3.3. The channel attention exploits the interdependencies between the channels and demonstrates the significance of each channel while suppressing less informative channels. Hence, we design a channel attention block for learning channel-wise feature interdependencies.

For an input feature map $F_1 \in \mathbb{R}^{H \times W \times C_1}$, where $i = 1$ and $C_1 = C/2$, we first apply global average pooling (GAP) to obtain channel-wise information as $F' \in \mathbb{R}^{1 \times 1 \times C}$. Then, we apply two $1 \times 1$ convolution layers with ReLU and sigmoid activation, respectively to obtain channel-wise attention weights $F_b \in \mathbb{R}^{1 \times 1 \times C_1}$ which can be represented as follows

$$F_b = \sigma(Conv(\alpha(Conv(GAP(F_1))))) \qquad (1)$$

where $\alpha$ and $\sigma$ represent the ReLU and sigmoid functions, respectively.

#### 2.2.3. *Position Attention Block (PAB)*

Since the size of tumor lesions varies significantly, it is indispensable for a model to focus both on small and large regions. Also, the location of the tumor provides valuable information for classification. To focus on these informative regions, position attention block has been introduced which explores the interdependencies of features in the spatial dimension. In general, it demonstrates the significance of each spatial position of the feature map. The PAB is mathematically represented as follows.

As shown in Fig. 2, for a given input feature map $F_1 \in \mathbb{R}^{H \times W \times C_1}$, we apply cross channel average pooling (CAP) to obtain a feature map $F' \in \mathbb{R}^{H \times W \times 1}$. Then, we reshape it to $\mathbb{R}^{1 \times M}$, where $M = H \times W$ is the total number of positions and apply a sigmoid function. Further, we reshape $F_1$ to $\mathbb{R}^{M \times C}$ and perform cross product to obtain the position attention weights $F_a \in \mathbb{R}^{1 \times 1 \times C_1}$ which can be defined as

$$F_a = \sigma(Reshape(CAP(F_1))) \otimes Reshape(F_1) \qquad (2)$$

where, $\otimes$ denotes the cross product.

#### 2.2.4. *Aggregation Strategy*

After generating the position and channel attention weights, we aggregate them using concatenation $\mathbb{R}^{1 \times 1 \times 2C_1}$ which makes the model to deal with size irregularities of tumor lesions and similar appearances of different classes. Then, we use a $1 \times 1$convolution layer followed by a sigmoid activation to make the channel number same as $F_1$ i.e., $\mathbb{R}^{1 \times 1 \times C_1}$. At last, we perform element-wise multiplication between the final attention weights and the input feature map ($F_1$) to obtain a feature map $F_c 1 \mathbb{R}^{H \times W \times C_1}$ which can be stated as

$$F_c = \sigma(Conv(Concat(F_a, F_b))) \odot F_1 \qquad (3)$$

where $F_c$ is the output of our attention module over a single feature group and $\odot$ represents element-wise multiplication. Similarly, for $F_2$, we compute the output as $F_c \in \mathbb{R}^{H \times W \times C_2}$. The outputs of $F_1$ and $F_2$ are finally concatenated to obtain a refined feature map ($\mathbb{R}^{H \times W \times C}$).

### 2.3. Classifier

For classification of MR images, we employ a GAP layer followed by FC layer with three neurons to decide the class label as meningioma, glioma or pituitary.

## 3. EXPERIMENTS AND RESULTS

### 3.1. Dataset and Experimental Setup

To evaluate the proposed approach, a publicly available brain tumor dataset [12, 13] is used, which includes 3064 T1-weighted MRI images of three tumor types: glioma (1426 images), meningioma (708 images), and pituitary (930 images). The images are obtained from 233 patients in coronal, sagittal, and axial views and the images have a resolution of $512 \times 512$ pixels.

We adopt a similar experimental setting as reported in [12, 4, 6, 5] to test the performance of our method, i.e., we use five-fold cross-validation while ensuring that images from one patient do not present in training and testing set simultaneously. For data augmentation, we use vertical flips, horizontal flips, and rotation (+$15^0$,-$15^0$). The input resolution of the proposed CDANet is $224 \times 244$. We train all the models with a batch size of 36 and base learning rate 0.0002 for 50 epochs. The networks are optimized using Adam optimizer with a cross-entropy loss function. We implement our method based on Keras with Tensorflow backend.

### 3.2. Results

To exhibit the generality of the CDANet, we adopt various ImageNet pre-trained CNN models such as DenseNet-121 [14], MobileNet [15], ResNet-50 [16], SqueezeNet [17], and EfficientNet B0 [18] as backbone networks. Further, to verify the effectiveness of the proposed attention block with the above CNN architectures, we compare it with other attention mechanisms such as CBAM [10] and SE [11], and the results are shown in Table 1. It can been seen that the proposed method with CSDA block leads to an improved performance compared to only baseline models and baseline with other attention mechanisms. Further, DenseNet-121 achieved the higher classification accuracy of 96.60%. It is worth mentioning here that all models are evaluated under similar experimental setup.

We perform the ablation studies of our model to investigate each component of the proposed CSDA block. As shown in Table 2, the performance of the model decreases significantly without channel splitting and with only PAB and CAB. However, all these components together result in higher classification accuracy. For this experiment, we consider only the best performing baseline model i.e., DenseNet-121.

**Table 1**. Classification results of CDANet with different backbone networks and attention mechanisms

| Backbone | Attention Type | Average Accuracy (%) |
|---|---|---|
| DenseNet-121 | Ours | **96.60** |
| | CBAM | 96.16 |
| | SE | 96.22 |
| | None | 95.88 |
| MobileNet | Ours | **94.88** |
| | CBAM | 94.68 |
| | SE | 94.12 |
| | None | 93.92 |
| ResNet-50 | Ours | **95.77** |
| | CBAM | 95.11 |
| | SE | 95.13 |
| | None | 94.92 |
| SqueezeNet | Ours | **93.52** |
| | CBAM | 93.13 |
| | SE | 93.38 |
| | None | 92.92 |
| EfficientnetB0 | Ours | **92.83** |
| | CBAM | 92.28 |
| | SE | 92.48 |
| | None | 92.09 |

**Table 2**. Comparison of different parts of CSDA block

| Model | Attention Type | Average Accuracy (%) |
|---|---|---|
| DenseNet-121 | Ours | **96.60** |
| | PAB | 95.95 |
| | CAB | 96.33 |
| | No channel split | 96.21 |
| | None | 95.88 |

To observe the effectiveness of the CSDA block intuitively, we generate heatmaps using Grad-CAM [19] for DenseNet-121 model with and without (w/o) attention, as depicted in Fig. 3. We can clearly see that the model with CSDA only focuses on more relevant regions of tumor lesions. Also, we compare the proposed approach with various existing brain tumor detection schemes on the same dataset as shown in Table 3. It can be observed that the proposed CDANet achieves promising performance compared to others which can be majorly attributed to its capability of learning more detailed and small lesion features using CSDA. It can be noted that the results of the existing methods are obtained from their original papers and all methods have been evaluated using the same train-test split.

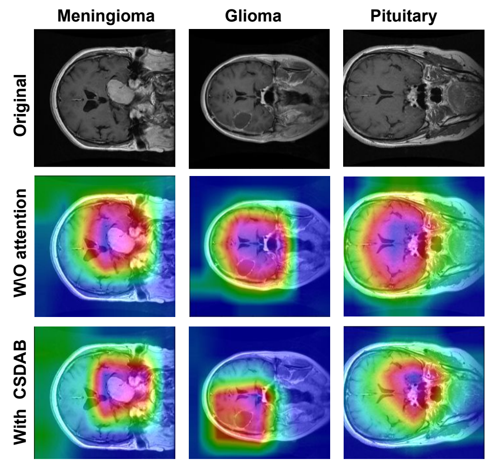


**Fig. 3**. Visualization results of the proposed attention block

**Table 3**. Comparison with existing CNN based classification methods on the same dataset

| Author | Method | Accuracy (%) |
|---|---|---|
| Paul et al. [4] | Custom CNN | 90.26 |
| Afshar et al. [5] | CapsNet | 86.56 |
| Swati et al. [6] | VGG19 with BFT | 94.82 |
| Ghassemi et al. [7] | GAN with custom CNN | 93.01 |
| Ours | CDANet | **96.60** |

## 4. CONCLUSION

In this paper, we have proposed a channel split dual attention-based network called CDANet for effective brain tumor classification which is end-to-end learnable. In particular, we introduce a CSDA module on the top of a backbone network which includes channel splitting, PAB, and CAB, facilitating learning more detailed and discriminative features from tumor lesions. We also compared our approach with state-of-the-art techniques on the same dataset to prove its effectiveness. The ablation experiments demonstrate that our attention module achieves superior performance with different backbone networks than other attention mechanisms, which exhibits the generality of CDANet.

## 5. REFERENCES


[1] Khan Muhammad, Salman Khan, Javier Del Ser, and Victor Hugo C De Albuquerque, "Deep learning for multigrade brain tumor classification in smart healthcare systems: A prospective survey," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 2, pp. 507–522, 2020.

[2] Khalid Usman and Kashif Rajpoot, "Brain tumor classification from multi-modality mri using wavelets and machine learning," *Pattern Analysis and Applications*, vol. 20, no. 3, pp. 871–881, 2017.

[3] Nidhi Gupta, Pushpraj Bhatele, and Pritee Khanna, "Glioma detection on brain MRIs using texture and morphological features with ensemble learning," *Biomedical Signal Processing and Control*, vol. 47, pp. 115–125, 2019.

[4] Justin S Paul, Andrew J Plassard, Bennett A Landman, and Daniel Fabbri, "Deep learning for brain tumor classification," in *Medical Imaging 2017: Biomedical Applications in Molecular, Structural, and Functional Imaging*. SPIE, 2017, vol. 10137, pp. 253–268.

[5] Parnian Afshar, Arash Mohammadi, and Konstantinos N Plataniotis, "Brain tumor type classification via capsule networks," in *25th IEEE international conference on image processing (ICIP)*. IEEE, 2018, pp. 3129–3133.

[6] Zar Nawab Khan Swati, Qinghua Zhao, Muhammad Kabir, Farman Ali, Zakir Ali, Saeed Ahmed, and Jianfeng Lu, "Brain tumor classification for mr images using transfer learning and fine-tuning," *Computerized Medical Imaging and Graphics*, vol. 75, pp. 34–46, 2019.

[7] Navid Ghassemi, Afshin Shoeibi, and Modjtaba Rouhani, "Deep neural network with generative adversarial networks pre-training for brain tumor classification based on mr images," *Biomedical Signal Processing and Control*, vol. 57, pp. 101678, 2020.

[8] Jun Fu, Jing Liu, Haijie Tian, Yong Li, Yongjun Bao, Zhiwei Fang, and Hanqing Lu, "Dual attention network for scene segmentation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 3146–3154.

[9] Along He, Tao Li, Ning Li, Kai Wang, and Huazhu Fu, "Cabnet: category attention block for imbalanced diabetic retinopathy grading," *IEEE Transactions on Medical Imaging*, vol. 40, no. 1, pp. 143–153, 2020.

[10] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 3–19.

[11] Jie Hu, Li Shen, and Gang Sun, "Squeeze-and-excitation networks," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7132–7141.

[12] Jun Cheng, Wei Huang, Shuangliang Cao, Ru Yang, Wei Yang, Zhaoqiang Yun, Zhijian Wang, and Qianjin Feng, "Enhanced performance of brain tumor classification via tumor region augmentation and partition," *PloS one*, vol. 10, no. 10, pp. e0140381, 2015.

[13] Jun Cheng, Wei Yang, Meiyan Huang, Wei Huang, Jun Jiang, Yujia Zhou, Ru Yang, Jie Zhao, Yanqiu Feng, Qianjin Feng, et al., "Retrieval of brain tumors by adaptive spatial pooling and fisher vector representation," *PloS one*, vol. 11, no. 6, pp. e0157112, 2016.

[14] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger, "Densely connected convolutional networks," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4700–4708.

[15] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.

[16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[17] Forrest N Iandola, Song Han, Matthew W Moskewicz, Khalid Ashraf, William J Dally, and Kurt Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and¡ 0.5 mb model size," *arXiv preprint arXiv:1602.07360*, 2016.

[18] Mingxing Tan and Quoc Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International Conference on Machine Learning*. PMLR, 2019, pp. 6105–6114.

[19] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *IEEE international Conference on Computer Vision*, 2017, pp. 618–626.