Proceeding of the IEEE
International Conference on Robotics and Biomimetics
Dali, China, December 2019

Dilated Capsule Network for Brain Tumor Type Classification Via MRI Segmented Tumor Region

Kwabena Adu, Yongbin Yu, Jingye Cai
School of Information and Software Engineering
University of Electronic Science and Technology of China,
Chengdu, China, 610054
adukwabena4u@gmail.com
{*ybyu, jycai*}*@uestc.edu.cn*

Nyima Tashi
School of Information Science and Technology
Tibet University, Lhasa,
China, 850000

***Abstract*—Brain tumor recently is considered among the deadliest cancers according to research statistics and have several categories, based on the different characteristics of the tumor. Early detection of the tumor types help to devise treatment plans and achieve high survival rate. Human inspection is noted to be cost effective, error prone and time-consuming, which have led the interest in Convolutional Neural Networks (CNNs) to automatize the problem. However, CNNs fail to consider the precise location of the features as beneficial, which is harmful, because tumor location and its relationship with the surrounding tissue provide high influence on the brain tumor type. In addition, the CNNs require large amount of dataset for accurate training and prediction. CNNs increasingly reduce image resolution, which result to decrease in classification accuracy. In this work, we incorporated recently developed Capsule Networks (CapsNets) which overcome these drawbacks. The focused contribution is to enhance CapsNets with dilation to maintain the image resolution and improve classification accuracy. We proposed a less trainable CapsNet architecture for brain tumor classification, which takes the segmented tumor regions as inputs within the structure and has the capability of ensuring an increase focus of the CapsNets. While the baseline CapsNets consist of single convolutional layer, our proposed model introduced multiple convolutional layers which achieved an improved performance of 95.54% compared to the related works. Our results indicate that the proposed approach can improve brain tumor classification problem.**

***Index Terms*—brain tumor classification, capsule networks, convolutional neural network, dilated convolution**

I. Introduction

According to American Brain Tumor Association (ABTA), Brain and Central Nervous System (CNS) tumors are the third common cancer among adolescents and young adults, and cause of mortality rate in human across the world [1][2]. Siegel et al., 2016 [3], showed that brain tumor is the leading cause of high cancer mobility rate around the world. The tumor severely influence the life of patients. Early detection and diagnosis of tumor type is a step to increase survival rate and treating the cancer. Brain tumor have different categories such as Meningioma, Glioma and Pituitary [2] depending on factors such as the location, texture and shape of the tumor. To correctly detect the brain tumor type is of crucial importance as it can help physicians to devise treatment plans and better predict patients' survival.

Medical image processing has been well considered to be the common, accurate and precise techniques for use on cancer type classification [3]. Among the available screening technologies, the Magnetic Resonance Imaging (MRI) is used as the appropriate method for brain tumor classification. However, cancer recognition using MRI image with human inspection is cost-effective, error-prone, and time-consuming procedure, as it solely relies on the experience of radiologist. Based on these issues, it necessary to develop an innovative and proper brain tumor type classification approach, which this paper seeks to adopt as its focus.

Related work: In this section of the related work, our focus is based on the various trends used in brain tumor classification and shotcomings that have resulted in the proposing of different approaches. We show the various techniques as follows:

Convolutional Neural Network: The drawbacks of the conventional study of machine learning (ML) problems have paved way towards the adoption of deep learning, which in particular Convolutional Neural Networks (CNNs) [4][5], to perform cancer detection, diagnosis, segmentation and classification. The interest to use CNNs for classifying brain tumor has recently increased [6], where CNNs and neural networks (NN) are combined with different pre-processing methods such as data augmentation. Reference [7] developed six layer CNN, which extracts feature from a given brain MRI images for the purpose of performing brain tumor classification. Although, CNNs have overcome many image processing approaches, however, the network does not necessary need prior knowledge on existing feature type. The CNNs have achieved an improved learning capacity but associated with some key shortcomings [8][9]. For the CNNs to consider the spatial relation between objects in an input image becomes a challenge, which result to lack of rotation and affine transformation. In addition, CNNs require huge amount of training samples to improve the robustness of the networks, which in the case of brain tumor classification problem is not always available.

Capsule Networks: To overcome the above-mentioned drawbacks of the CNNs, CapsNets [10][11] have been pro-

This work was supported by National Natural Science Foundation of China (NSFC Grant No. 61550110248), Sichuan Science and Technology Program (Grant No. 2019YFG0190) and Research on Sino-Tibetan multi-source information acquisition, fusion, data mining and its application (Grant No. H04W170186)

978-1-7281-6321-5/19//$31.00 ©2019 IEEE

posed and have the focused to enhance the robustness to transformations. The term capsule can be define as a group of neurons, which can substitute different instantiation parameters which is associated with different objects, and the probability of their existence. The unique property of this network is known as "Routing by Agreement", which generally means capsule in lower level predicts output of capsules in the high levels. The high-level capsule gets activated when these predictions agree. P. Afshar et al., 2018 [12] investigated and used capsule networks for brain tumor type classification. The work explored some potential structures of CapsNets to accurately identify the architecture that best maximize prediction accuracy for the brain tumor problem. In other work, P. Afshar et al., 2019 [13] modified CapsNet structure to access the surrounding tissue of the tumor without distraction from the actual target for the brain tumor classification, which accepts the boundaries of tumor coarse as an additional input in the pipeline to enhance the focus of the CapsNets. The recent work by P. Afshar et al., 2019 [14] explored the interpretability of Capsule networks via Radiomic analaysis which achieved significant results in brain tumor classification.

The down-sampled techniques by the CNNs increasingly decrease the resolution of feature maps, which result to lower classification accuracy. This paper based on these drawbacks, enhance the CapsNet with dilation factor to keep the resolution of feature maps high for improved performance and high resolution during reconstruction. The introduction of dilation in the network reduce parameter to be learned and decrease training time. We used the segmented tumor regions of MRI images as input to the CapsNet. Our results indicate that the proposed method outperforms other related works.

The rest of this paper is organized as follows: Section 2 describes problem formulation and background for CNNs and CapsNets. In section 3, we present our proposed method followed by an experimental setup in section 4. Finally, Section 5 concludes the paper.

II. PROBLEM FORMULATION

The goal of this work is to develop a deep learning (DL) architecture to classify brain tumor into different types using MRI segmented tumor images. Although, there are many types of brain tumor, this work focused on 3 types of the tumors (i.e. Meningioma, Glioma and Pituitary). In this section, we briefly reviewed available techniques for brain tumor classification problem such as CNNs, which is the building block for deep neural networks (DNNs). We further highlight the potential drawbacks of CNNs that have resulted to CapsNets and the design of other architectures as follows:

A. Convolutional Neural Networks

CNNs consist of stack convolutional layers, pooling layers, activation layers and fully connected layers (FC) [15]. The method of weight sharing over the entire input is an important property of the CNNs. This significantly decrease the computational cost and enhance the network with the capability of extracting elementary and higher other local features. The neurons in higher layer receive inputs from the lower layers which are localized in a closely neighborhood. In this way, layers such as first layer extracts edges, textures and corners while the extracted features are integrated in the upper layers to extract high dimensional complex features. CNNs have achieved popularity in the medical image processing problems due to the fact that the networks does not require prior information on the types of features to be extracted [16].

Finally, CNNs usually implement sub-sampling techniques in the layers to reduce the size of input. Though, CNNs are accepted to provide useful results in many research fields due to the beneficial feature and characteristic it carries. In this work, we noted that the convolution layers do not consider the precise location of the features as beneficial. This makes the approach harmful, because there are variations of information for different instances.

B. Capsule Networks

The concept of capsule is defined as the collective neurons indicating activity vectors of the neurons which represent the existing pose parameter. The length of the various vectors show the probability existence of a specific entity. The pooling layer in CNNs mostly pose a challenge to the structure by reducing resolution. This shortcoming of the CNNs have resulted in the proposing of CapsNets, which has successfully substituted the pooling layer with routing by agreement method. The "Routing by Agreement" by CapsNet has shown appropriate replacement of the pooling layer. Based on this criteria, outputs from the lower layer are fed to available parent capsules in the high layers; however, the coupling coefficients among them tend to be different from each other. The capsules found in the lower layer predict the output of the parent capsules. The coupling coefficient in between these capsules increase if the prediction matches with the output of the parent capsule. Initialize the term u_i to be the output of capsule *i*. The predicted output of u_i for the parent capsule *j* can be computed as

$$\hat{\boldsymbol{u}}_{\boldsymbol{j|i}} = \boldsymbol{W_{ij}} \boldsymbol{u_i}, \tag{1}$$

where $u_{j|i}$ in (1) indicates the prediction vector of capsule j^{th} output in the higher layer computed by capsule i in the lower layer. The weighting matrix defined as W_{ij} learns in backwards pass. The result is based on the conformation existing between lower capsules and parent capsule. In addition, coupling coefficient is defined as c_{ij}. Where more weights are assigned to the predictions that have higher agreement with the actual output. The c_{ij} is computed as

$$\boldsymbol{c_{ij}} = \frac{\boldsymbol{exp(b_{ij})}}{\sum_{\boldsymbol{k}} \boldsymbol{exp(b_{ik})}}, \tag{2}$$

where the log probability is defined as b_{ij}, and indicates whether capsule i couples with capsule j. At the initial stage of the routing by agreement process, b_{ij} is initialized to 0. Based on this, input vector to the parent capsule j is calculated as

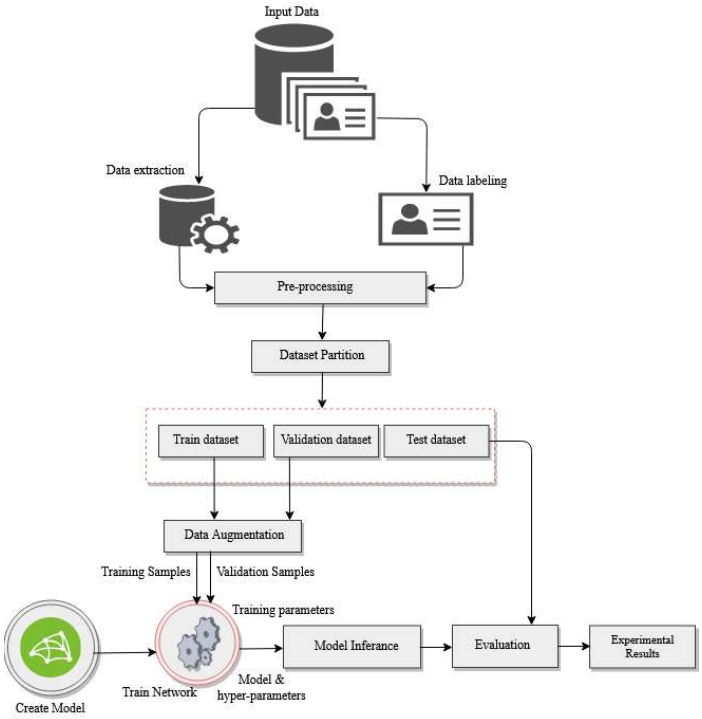


Fig. 1: Block data flow implementation of the proposed model.

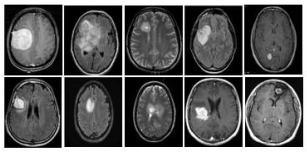


Fig. 2: Sample brain tumor MRI images. The images show whole MR images captured by radiologist from patients.

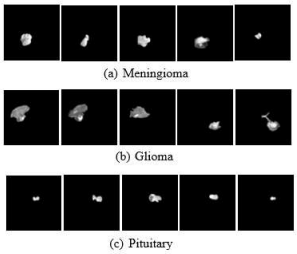


Fig. 3: Sample brain tumor MRI segmented images. In (a) is Meningioma, which is seen to grow slowly. With this tumor, 90% are considered to be benign (i.e. non-cancerous) and mostly occurs in the brain. In (b) is another type of brain tumor called Gliomas, which comprises about 30% of all the brain tumors and central nervous system tumors. The tumor forms 80% percent of all malignant (i.e. cancerous) brain tumors. The Final type of tumor in (c) is Pituitary. Represent nearly 16.2% of all primary brain tumors and rarely become malignant.

$$\boldsymbol{s_j} = \sum_{\boldsymbol{i}} \boldsymbol{c_{ij}} \hat{\boldsymbol{u}}_{\boldsymbol{j|i}} \tag{3}$$

The CapsNet applies nonlinear squash function to ensure the output vectors of capsules do not exceed 1 and forms the final output of each capsule based on the initial vector value defined in (3).

$$\boldsymbol{v_j} = \frac{||\boldsymbol{s_j}||^{\boldsymbol{2}}}{\boldsymbol{1} + ||\boldsymbol{s_j}||^{\boldsymbol{2}}} \frac{||\boldsymbol{s_j}}{||\boldsymbol{s_j}||}, \tag{4}$$

where s_j is the input vector to the j^{th} capsule and v_j is the output. A. Shahroudnejad et al., 2018 [17] showed that, CapsNets used nonlinear squashing function on output vectors v_j in each iteration. It actually bounds probability of these vectors between 0 and 1, which means that it suppresses small vectors and preserves long vectors in the unit length

$$\begin{cases} v_j \approx ||s_j|| s_j = 0, & \text{if } s_j \text{ is small} \\ v_j \approx \frac{||s_j||}{||s_j||}, & \text{otherwise} \end{cases} \tag{5}$$

The log probabilities are updated in the routing process based on the agreement between v_j and using the fact that if the two vectors agree, they will have a large inner product. Therefore, agreement a_{ij} for updating log probabilities and coupling coefficients can be calculated as follows

$$\boldsymbol{a_{ij}} = \boldsymbol{v_j} \hat{\boldsymbol{u}}_{\boldsymbol{j|i}} \tag{6}$$

Capsule k in the last layer is associated with a loss function l_k, which puts high loss value on capsules with long output

instantiation parameters when the entity does not actually exist. The loss function l_k is calculated as

$$l_k = \mathrm{T}_k \max(0, m^+ - ||v_k||)^2 + \lambda(1-\mathrm{T}_k)\max(0, ||v_k||, -m^-)^2, \quad (7)$$

Let T_k be 1 whenever the class k is present, and is 0 otherwise. The hyper-parameters to be initialized for training process are the terms m^+, m^-, and λ.

The baseline CapsNets architecture consist of 1-layer convolutional network and 2 capsule layers. Finally, it has 3 fully connected networks which perform reconstruction on inputs by using the instantiation parameters from capsules associated with the true labels.

III. Proposed Pipeline

As mentioned above, the focus of this study is to perform classification of brain tumor into three (3) categories namely meningioma, glioma and pituitary using MRI segmented tumor images. The MRI images are widely considered as best modality and method for detecting brain diseases.

CNNs have major shortcomings which is limited when applied to real-world problems. The CNNs fail to consider the precise location of the features as beneficial, which is harmful, because there are variations of information for different instances. The lost of spatial information in CNNs are crucial in the issue of brain tumor classification, because tumor location and the relationship with surrounding tissue provide high influence on the brain tumor type. In addition, the CNNs require large amount of dataset for accurate training and prediction. Based on this, the network performs poorly when trained with small dataset, which in the case of most medical images are not available such as brain MRIs

The recently proposed network called CapsNets reviewed in section II above have successfully replaced the pooling layers of CNNs with appropriate method called "Routing by Agreement", which have the potential to maintain the spatial relation and make it more suitable model for brain tumor classification. CapsNets have breakthrough the problems of CNNs.

A. Dilated Networks

Image classification using CNNs increasingly reduce image resolution. This happens till the image is fully represented by small feature map where the spatial structure of input image gain loss of recognizability. The loss of spatial vision can result to decrease in image classification performance. Furthermore, the loss can introduce complication on the transfer of model to applications that requires precise and detailed understanding of image. In the above observation, we employed the approach of dilation to avoid the need of up-sampling and maintain high resolution on the convolutional layer. This was done for improvement of accuracy on classification of image on the network. Dilated convolution can be written as

$$(\boldsymbol{F} *_l \boldsymbol{k})(\boldsymbol{p}) = \sum_{\boldsymbol{s}+l\boldsymbol{t}=\boldsymbol{p}} \boldsymbol{F}(\boldsymbol{s})\boldsymbol{k}(\boldsymbol{t}) \quad (8)$$

where (8) shows the dilated convolution respectively. Let $*l$ is referred to as a dilated convolution. The known discrete convolution * simply describes the 1-dilated convolution. Dilated convolution is given by l>1 while l = 1 is seen as the standard convolution.

Let F_0, $F_1, ..., F_{n-1} : \mathbb{Z}^2 \to \mathbb{R}$ be discrete functions and let k_0, $k_1, ..., k_{n-2} : \Omega 1 \to \mathbb{R}$ be discrete 3×3 kernels. We considered and applied the kernels in an exponentially increasing dilation:

$$F_{i+1} = F_{i *2^i} k_i \text{ for } i = 0, 1, ..., n-2. \quad (9)$$

Receptive field of an element p is defined in F_{i+1} representing the set of elements in F_0 which modifies the values of $F_{i+1(p)}$. The size of receptive field represents the number of these element. At this point, the size of the receptive can be seen in each element in F_{i+1} is $(2^{i+2}$ - $1) \times (2^{i+2}$ - $1)$. The receptive field is viewed as a square of exponentially increasing size.

This work was motivated by the above-mentioned issues, where in this work, we designed a CapsNet pipeline that allows brain images as input. The network also used dilation (Fig.5) to enhance and keep the resolution of the input images high for better and accurate classification. We show that dilation on the architecture reduced the number of parameters to be learned and also decreased training time. In this work, we note that addition of more convolutional layers can improve accuracy. Hence, we explored multiple convolutions and identified the best one which maximize accuracy.

The proposed structure shown Fig. 2 is described in detail as follows:

- Input: Down-sampled 64×64 from 512×512 patch captured from MRI image. The down-sampled patches reduce the parameters to be learned and decrease the training time of the structure.
- Convolutional layer with 64 filters, 9×9 kernel, stride of 1, dilation=1 and ReLU as nonlinear activation function. This leads to 56×56×64 feature maps.
- Convolutional layer with 64 filters, 3×3 kernel, stride of 1, dilation=2 and ReLU as nonlinear activation function. This leads to 52×52×64 feature maps.
- Convolutional layer with 64 filters, 3×3 kernel, stride of 1, dilation=3 and ReLU as nonlinear activation function. This leads to 46×46×32 feature maps.
- The Primary capsule layer resulted from 256×9×9 and stride of 2 convolutions. The component capsules of the layer is 32 and consists of 8 dimension each, which result to feature map of size 19×19. That is each component available in the primary capsule contains 19×19 individual localized capsules.
- The final part of the encoder referred to as class capsule includes 3 capsules representing each type of brain tumor (i.e. meningioma, glioma and pituitary) to be classified. The dimension of this encoder is 16.
- The decoder part of our structure consists of 3 fully connected layers resulting to 512, 1024 and 4096 re-

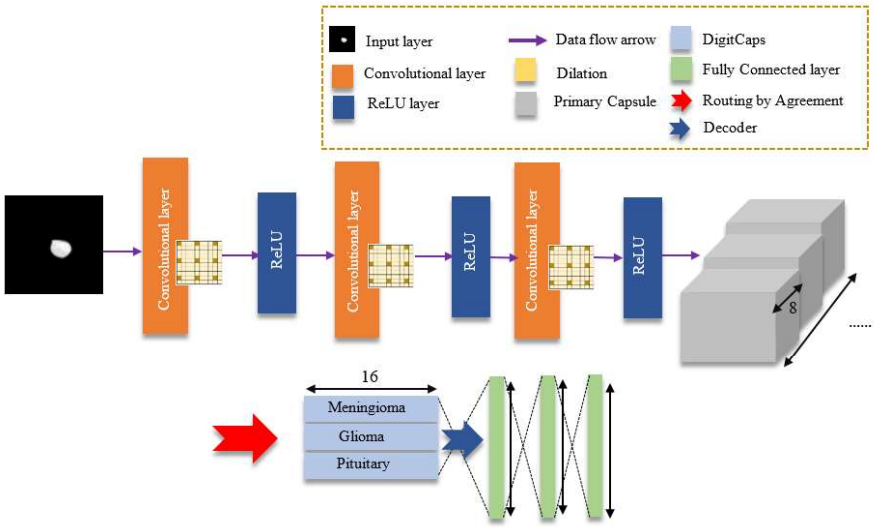


Fig. 4: Proposed pipeline for brain tumor classification. The architecture is composed of input, 3 convolutional layers with dilation, primarycaps layer, digitcaps and 3 fully connected layers. Best view in color.

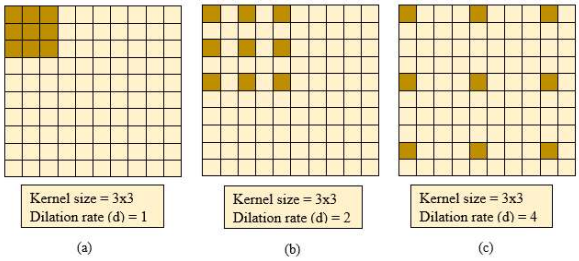


Fig. 5: Increase receptive field with a dilation rate. We can see that each receptive field is composed of kernel size 3×3 and associated with different dilation rate of 1, 2 and 4 respectively. In (a), the dilation rate of 1 means skip 0 pixel. This initialization gives the general standard convolutional kernel where the 3×3 kernels are kept without any skip of pixels. In (b), the given dilation rate of 2 means skip 1 pixel from a kernel. Finally in (c), it follows the same pattern as in (a) and (b) by skipping 3 pixels given dilation rate of 4. Generally, this is viewed as generating zeros in between the kernels. Best view in color.

spectively. In this structure, we note that the number of existing neurons in the last FC layer is equally the same compared to the given input image, as we seek to minimize the sum of squared difference between the input image and the reconstructed ones.

- Output: softmax classifier, 3 classes. This corresponds to the total tumor types to be classified.

We note that, the proposed model achieved the least total trainable parameters of 6.7M compared to the previous related works that used CapsNet such as P. Afshar et al. 2018 with the total trainable parameters of 7.9M.

IV. Experimental Setup

To evaluate the proposed structure shown in Fig. 4, we used MRI images dataset by J. Cheng et al., 2016, J. Cheng et al., 2015 [8] [18]. The tumor dataset contains 3,064 MRI images of 233 patients consisting three kinds of brain tumor namely (i) meningioma (708 slices), (ii) glioma (1426 slices), and (iii) pituitary tumor (930 slices). The dataset is composed of both brain images and the segmented tumors, which in this work, we used the segmented tumors for the experiments. Fig. 2 and 3 show the brain images and the segmented tumors of the dataset. Table 1 shows the detail values of the hyper-parameters adopted to train the proposed structure. Our framework was implemented using Pytorch deep learning libraries. On a server with NVIDIA GeForce GTX 1060 6GB GPU, and 64bit windows10 as OS. In Table 2, we compare the proposed structure with existing related works which trained their model using the segmented brain tumor regions. In Fig. 6 (a), we show the training and test accuracy curve while (b) shows the training and test loss curve.

TABLE I: Hyper-parameters used for training brain tumor classification.

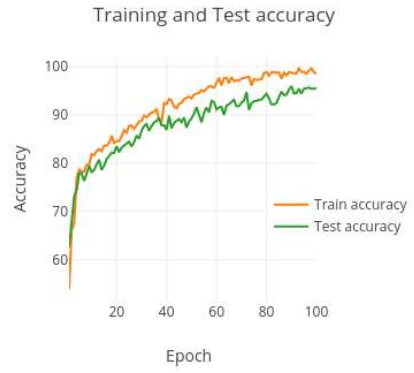
Hyper-parameter	Optimized Value
Batch size	32
Number of Epochs	100
Learning rate	0.001
Dilation rate	1, 2, 3
Routing iteration	3
Activation function	ReLU
Optimizer	Adam [20]

V. Conclusion

In this work, we highlighted a Capsule Networks structure for brain tumor classification problem. We used dilation in the

TABLE II: Results comparison between related works and proposed approach.

	Method	Classification Type	Accuracy
1	CNN [6]	Multi	61.97%
2	CNN [6]	Multi	72.13%
3	CapsNet [12]	Multi	78%
4	CapsNet [12]	Multi	86.56%
5	P. Afshar [13]	Multi	90.89%
6	**Proposed Approach**	Multi	**95.54**%



(a)

(b)

Fig. 6: Training and test curve for the proposed model. The (a) in this figure shows training and test accuracy while (b) is the training and test loss curve. Best view in color.

architecture to enhance the image resolution and to aggregate the contextual information in the brain tumor. The dilation also facilitates the architecture to expand the receptive fields without losing resolution and reduced learnable parameters, and training time. We evaluated CapsNet with multiple convolutional layers, the structure consists of 3 convolutional layers that feed input to the primary capsule. Our architecture was evaluated on segmented MRI images and achieved the state-of-the-art performance. We showed that, the architecture can improve brain tumor classification problem.

ACKNOWLEDGMENT

This work was supported by National Natural Science Foundation of China (NSFC Grant No. 61550110248), Sichuan Science and Technology Program (Grant No.2019YFG0190) and Research on Sino-Tibetan multi-source information acquisition, fusion, data mining and its application (Grant No. H04W170186). The authors would like to thank the editor and the reviewers for their helpful suggestions and valuable comments.

REFERENCES

[1] Brain Tumor Statistics. *American Brain Tumor Association.* January 2017.

[2] Key Statistics for Brain and Spinal Cord Tumors. *American Cancer Society.* January 4, 2018.

[3] R.L. Siegel, K.D. Miller, A. Jemal, "Cancer Statistics, 2017," *A Cancer Journal for Clinicians,* vol. 67, issue 1, pp. 7-30, 2017.

[4] K. Usman, K. Rajpoot, "Brain Tumor Classification from Multi-Modality MRI using Wavelets and Machine Learning," *Pattern Analysis and Applications,* vol. 20, no. 3, pp. 871-881, Aug. 2017.

[5] A Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, Jan. 2012, pp. 1097–1105.

[6] J. S. Paul, A. J. Plassard, B. A. Landman, and D. Fabbri, "Deep learning for brain tumor classification," *Proc. SPIE, Med. Imag., Biomed. Appl. Mol., Struct., Funct. Imag.,* vol. 10137, Mar. 2017, Art. no. 1013710. doi: 10.1117/12.2254195.

[7] Z. Li, Y. Wang, et al., "Deep Learning Based Radiomics (DLR) and Its Usage in Noninvasive IDH1 Prediction for Low Grade Glioma," *Scientific Reports,* vol. 7, no. 1, 2017.

[8] J. Cheng, W. Huang, S. Cao, R. Yang, W. Yang, Z. Yun, Z. Wang, Q. Feng, "Enhanced Performance of Brain Tumor Classification via Tumor Region Augmentation and Partition," *PloS one.*, 2015.

[9] N. K. E. Abbadi, N. E. Kadhim, "Brain Cancer Classification Based on Features and Artificial Neural Network," *International Journal of Advanced Research in Computer and Communication Engineering,* vol. 8, no. 1, Jan. 2017.

[10] S. Sabour, N. Frosst, G.E. Hinton, "Dynamic Routing Between Capsules," *31st Conference on Neural Information Processing Systems (NIPS 2017),* 2017.

[11] G.E. Hinton, N.Frosst, and S.Sabour, "Matrix Capsules with EM Routing," *Proc. ICLR,* Feb. 2018.

[12] P. Afshar, A. Mohammadi, K. N. Plataniotis, "Brain Tumor Type Classification via Capsule Network," *Proceedings - International Conference on Image Processing, ICIP,* pp. 3129-3133, Sept. 2018.

[13] P. Afshar, A. Mohammadi, K. N. Plataniotis, "Capsule Networks for Brain Tumor Classification Based on MRI Images and Coarse Tumor Boundaries," *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP* pp. 1368-1372, 2019.

[14] P. Afshar, A. Mohammadi, K. N. Plataniotis, "Capsule Networks' Interpretability for Brain Tumor Classification Via Radiomics Analyses," *IEEE International Conference on Image Processing, ICIP*, 2019.

[15] Y. Lecun, L. Bottou , Y. Bengio , P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE, 1998.*

[16] D. Ravi, C. Wong, et al., "Deep Learning for Health Informatics," *IEEE Journal of Biomedical and Health Informatics,* vol. 21, no. 1, pp. 4-21, 2017

[17] S. Atefeh, A. Parnian, P. Konstantinos N., M. Arash "Improved explainability of capsule networks: Relevance path by agreement"*2018 IEEE Global Conference on Signal and Information Processing, GlobalSIP 2018 - Proceedings,* pp. 549-553 2019.

[18] J. Cheng , W. Yang , M. Huang, W. Huang, J. Jiang, Y. Zhou, R. Yang, J. Zhao, Y. Feng, Q. Feng , W. Chen, "Retrieval of Brain Tumors by Adaptive Spatial Pooling and Fisher Vector Representation," *PloS one.,* 2016.

[19] Marcus D. Bloice, Andreas Holzinger, "A Tutorial on Machine Learning and Data Science Tools with Python," *Machine Learning for Health Informatics,* 2016.

[20] Diederik P. Kingma, Jimmy Ba, "Adam: A Method for Stochastic Optimization," *3rd International Conference for Learning Representations, San Diego,* 2015.