Journal of Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.98549795

SELF_ATTENTION FOR BRAIN TUMOR CLASSIFICATION IN MRI IMAGES

[1]Para Rajesh [2]A. Punitha
[1,2]Department of Computer Science and Engineering
[1,2]Faculty of Engineering and Technology
[1,2]Annamalai University, India
Email: [1]pprr21@gmail.com, [2]charuka17@gmail.com

ABSTRACT

The most prevalent and dangerous illness, brain tumors have a very low life expectancy in their highest degree. Consequently, a crucial step in enhancing patients' quality of life is treatment planning. Broadly speaking, tumors in the brain, lung, liver, breast, prostate, etc. are evaluated using a variety of imaging modalities, including computed tomography (CT), magnetic resonance imaging (MRI), and ultrasound images. In particular, MRI scans are employed in this work to identify brain tumors. Nevertheless, the massive amount of data produced by an MRI scan makes it impossible to manually classify a tumor as opposed to a non-tumor at a certain moment. Nevertheless, it has a restriction in that only a small number of images can get precise quantitative data. This research focuses on developing an automated brain tumor classification method using convolutional neural network in combination with self-attention mechanism. This combination helps in achieving higher accuracy in classifying MRI images as meningioma, pituitary, glioma, and no tumor.

Keywords: *Magnetic Resonance Imaging, brain tumor, classification, convolutional neural network, sel-attention, meningioma, pituitary, glioma.*

1. INTRODUCTION

A brain tumor is a growth of abnormal cells close to the brain. The brain tumors can develop in the tissue region of the brain or near the brain tissue. The pituitary gland, pineal gland, and membranes that surround the surface of the brain are nearby structures that may be affected by the tumor. For the detection of tumors and the modeling of tumor progression, brain MRI images are primarily used. Tumor detection and treatment procedures mostly employ the information available in the MRI images. More details about a specific medical condition are provided by an MRI than by a CT or ultrasound image. The comprehensive information provided by an MRI image on brain anatomy helps the physician in the process of identification of anomalies in brain tissue. But analyzing a large volume of MRI images require more expert radiologist and it is often prone to error. To alleviate such issues, developing an automated and reliable mechanism by following the principles of computer aided diagnosis is essential. In order to classify images into a set of predetermined labels, the process of image classification requires understanding of the contextual information in images. The descriptiveness and discriminative capability of features collected are crucial for achieving strong classification performance in image classification challenges. Intensity histograms, filter-based features, the widely used scale-invariant feature transform (SIFT), and local binary patterns (LBP) are examples of feature extraction methods often employed in medical imaging. Typically, a classification model is trained using the retrieved feature vectors. CNN has shown to be quite

effective in resolving image classification issues. CNN has the ability to learn local and global structures from image [3, 4].
Neural networks' attention mechanisms have a propensity to resemble humans' cognitive attention. The basic goal of this function is to try to de-emphasize the irrelevant information while emphasizing the crucial information. This procedure is essential for avoiding overtaxing a system's memory since working memory is constrained in both human and machines. Attention may be seen as a vector of important weights in deep learning.
Understanding human cognition is essential because attention process is at the crossroads of computer science, biology, psychology, and cognitive science. There is currently no unifying theoretical framework for reference in the area of human visual attention processes, and many cognitive aspects of human vision have to be investigated. The study of computer information processing can only benefit from furthering research on human attention processes, researching and modeling the principles of human vision in information processing. Human pay different attention to various things when low-level and high-level visual characteristics are combined and these aspects also contribute differently. So, it is crucial to pick the right weighting techniques to get it closer to how human eyes naturally observe things. Self-attention might act as a fundamental building component for image classification models, and this study considers pairwise self-attention [6], a generalization of traditional dot-product attention. In comparison to convolutional networks, our paired self-attention networks perform better. Self-attention networks may offer important advantages in terms of resilience and generalization which could be inferred from the experimental results presented in this study by examining the robustness of learned representations [5].

2. LITERATURE SURVEY

Radiologists can now diagnose human disease in various body regions with the aid of computer-aided detection and diagnosis, which is software that combines artificial intelligence and computer vision to analyze radiological and pathology images. Examples of such applications include colorectal cancer detection and segmentation [10, 11] and lung cancer categorization [12–14]. The extraction and selection of features phases have been incorporated into the training process of deep learning [15], which was inspired by knowledge of the neural networks found in the human brain [16]. Models based on deep learning are typically represented as a hierarchy of layers, where each layer is made up of the weighted sum of the components of the layer before it. The data is represented by the top layer, and the output, or solution, is represented by the bottom layer. Complex mapping functions can be replicated using multiple layers, enabling deep learning algorithms to tackle complex problems with typically less human involvement than conventional machine learning techniques. Deep learning has been successfully used to many problems in the analysis of medical images [18] and now outperforms machine learning methods [17].
Compared to traditional machine learning techniques, CNNs have improved accuracy [19]. CNNs have received a lot of attention in deep learning for their ability to automatically identify deep features while responding to minute changes in the images [20]. The structure of a conventional deep CNN-based classification of brain tumors is shown in Fig. 1. A basic rule of thumb is to have at least roughly 10 times as many samples as parameters in the neural network for the problem's successful generalization when training a CNN model with millions

of parameters [21]. If the training sample is too small, overfitting may happen during the training phase [22]. As a result, several research [23–24] use two-dimensional brain image slices generated from three-dimensional brain MRI volumes to address this issue, which boosts the initial dataset's sample size and reduces the class imbalance issue. Additionally, it has the benefit of lowering the starting data dimension and simplifying training computations.

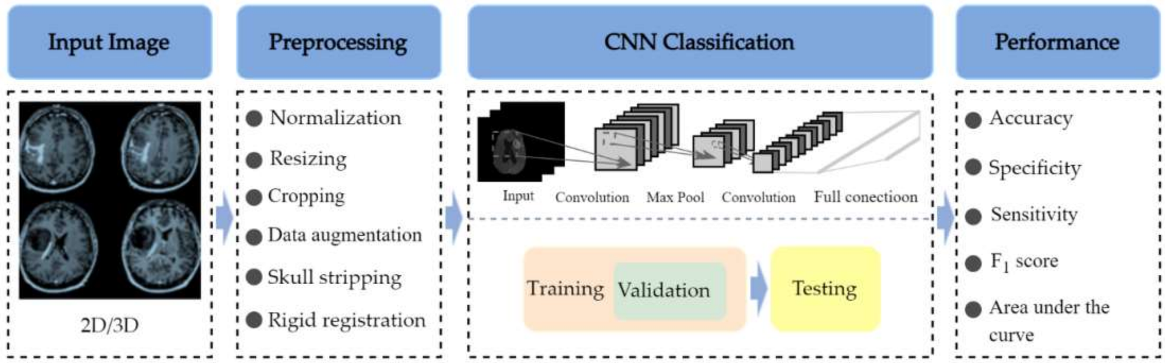

Fig. 1 The schematic view of flow of process in CNN based classification [3]

Self-attention was employed in computer vision applications up until very recently as a complement to convolutions, primarily in the form of layers that were added to, used to modify the outcome of, or otherwise combined with convolutions. Attention weights in channelwise attention models [30, 29, 28] reweight activations in various channels. Other methods [25, 32, 27] combined channel and spatial attention. There are several techniques available to maintain the fundamental idea behind convolutional feature creation, such as reweighting convolutional activations or offsetting the inputs of convolutional kernels [26, 29, 30, 32, 34]. Others [31, 33] used self-attention in particular modules attached to convolutional structures.

Convolutional and self-attention processing streams were integrated [35], but they discovered that the global self-attention they employed was insufficient to completely replace convolutions. In [36] investigation of adaptive filter networks, which generalized convolutions, was expensive in terms of memory and processing power and could not handle images with high resolutions and larger datasets. The most recent findings presented in [37] and [38] are most closely related to the proposed work. In contrast to past constructs that applied self-attention globally throughout the entire feature map [31, 35], one of their important contributions is limiting the scope of self-attention to a local patch. The successful implementation of self-attention across the network, including the initial high-resolution layers, depends on such local attention, which is essential for minimizing the amount of RAM and CPU used by the model. Our research expands on these findings by investigating a wider range of self-attention formulations. In particular, rather than using a common scalar weight, our basic self-attention mechanisms calculate a vector attention that adjusts to different channels. Additionally, we investigate a group of patchwise attention operations that strictly generalize convolution and differ fundamentally from the designs used in [37, 38]. We demonstrate the scalability, favorable parameter, and FLOP budgets of all the self-attentional models provided.

3. METHODOLOGY

In order to categorize medical images into a set of predetermined labels, the process of image classification requires the spatial and contextual information in images. Neural networks' attention mechanisms frequently resemble humans' cognitive attention. The purpose of this attention mechanism is to de-emphasize the irrelevant information while emphasizing the crucial information required for detection and classification of tumors. An image model must be trained to be able to extract features from the key elements of the image. Training attention processes is the most effective approach as in this investigation tumor images are being analyzed. Identifying which aspect of the image is most likely to cause cancer is crucial. During the pre-processing step the images are scaled, resized, and augmented. The process of expanding the amount of data utilized to train a model is known as data augmentation. The deep learning models frequently need a large amount of training samples, which is not always available, in order to make accurate predictions. As a result, new information is added to the old data to improve the generalized model.

3.1 Dataset

The experimental analysis is implemented using a dataset available in the figshare dataset which is publicly accessible [26] and is frequently used to assess classification and retrieval techniques. It consists of 3064 MRI images of brain captured from 233 patients who have been identified as having one of the following types of tumor namely meningioma, glioma and pituitary. The coronal, sagittal, and axial views are all available in the T1-CE MRI modality. From the Br35H dataset the images with no tumors have been taken. Table 1 presents the distribution of It includes 1426 brain MRI images with glioma of 89 brain tumor affected people, 708 meningioma images of 82 brain tumor patients, and the remaining 930 pituitary tumor images belonging to 62 patients. The pixel values of the MRI images are scaled between 0 and 1 using a min-max normalization approach. The image was resized to 224x224. The three channels were then produced by replicating the gray scale values of the images.

Table 1. Distribution of image classes

SNo	Class	Number of Samples
1	No tumor	1274
2	Meningioma	1042
3	Glioma	1090
4	Pituitary	1164

3.2 Overview of Self-Attention Network

The foundation of CNN is a layer of convolutional neurons. Multiple 2D matrices are produced as the with reference to one or more 2D matrices (or channels) as the convolutional layer's input. There may be a difference in the number of input and output matrices. The process required to compute a single output matrix is as follows:

$$A_j = f(\sum_{i=1}^{N} I_i * K_{i,j} + B_j) \qquad \text{Eq.1}$$

Initially, a kernel matrix $K_{i,j}$ corresponding to each input matrix I_i is convolved. Next, each element of the resultant matrix is added to a bias value B_j, and the total of all the convoluted matrices is calculated. In order to create one output matrix A_j, a non-linear activation function *f* is lastly applied to each component of the matrix A_j. Each set of kernel matrices is a local feature extractor that takes the input matrices and extracts local features. Finding sets of kernel matrices *K* that extract useful discriminative features for use in image classification is the goal of the learning process. The kernel matrices and biases may be trained as shared neuron connection weights in this setup by using the back propagation technique, which optimizes neural network connection weights [7][35].

For CNN to reduce feature dimension, the pooling layer is crucial. Pooling methods should be used to merge the neighboring elements in the convolution output matrices in order to decrease the quantity of output neurons in the convolutional layer. The max-pooling and average-pooling operations are two popular pooling techniques. The highest value from the four adjacent input matrix elements is chosen in max-pooling layer using a 2X2 kernel size to create one member in the output matrix. Just the neurons that contribute to the pooling output must receive the gradient signal during the error back propagation process. Methods like batch learning, momentum, and weight decay are used to speed up and stabilize neural network training. To increase learning accuracy and speed, batch learning is used. In the experiments and results 128 input samples in a batch were processed before changing the connection weights once for the whole batch, as opposed to updating them after each backpropagation. For to improve the convergence process momentum [8] and weight decay [9] mechanisms were introduced; the weight update can be define as follows;

$$\Delta\omega_i(t+1) = \omega_i(t) - \eta\frac{\partial E}{\partial w_i} + \alpha\Delta\omega_i(t) - \lambda\eta\omega_i \qquad \text{Eq. 2}$$

$\omega_i(t) - \eta\frac{\partial E}{\partial w_i}$ represent the conventional back-propagation term in which $\omega_i(t)$ is the current weight vector and $\frac{\partial E}{\partial w_i}$ denotes the error gradient with respect to the weight vector. The η denotes the learning rate which usually controls the rate of convergence. The $\alpha\Delta\omega_i(t)$ represent the momentum term wherein the α denotes the momentum rate. The λ represents the weight decay rate which helps to reduce the chance of model overfitting. Each learning iteration results in a tiny reduction or decay of the weight vector towards zero, aiding in the stabilization of the learning process. In our experiment, we set the learning rate to be 0.001, the momentum rate to be 0.9, and the decay rate to be 0.01. These learning parameters were chosen using a grid search mechanism.

The convolution layers of the network serve two purposes in convolutional networks for image recognition. Feature extraction is first, the convolution operation performs aggregation by integrating functions from each place that the kernel has accessed. Feature transformation is the second operation, and it is executed via nonlinear mappings and subsequent linear mappings. These subsequent translations and nonlinear scalar functions procedures deconstruct the feature space and result in complicated mappings in pieces. The two functions of feature aggregation and feature transformation may be isolated, which is one observation that forms the basis of our structure. Feature transformation may be carried out by perceptron layers that individually process each feature vector (for each pixel) if we have a method for feature

aggregation. A perceptron layer consists of a pointwise operation that transforms features into a linear mapping and a nonlinear scalar function.
Thus, feature aggregation is the main emphasis in the design of deep learning based image classification process. The convolution operator combines feature values from a collection of neighboring places linearly using a defined kernel that applies pre-trained weights. The kernel weights are fixed and do not change based on the image content. Moreover, the number of parameters increases linearly with the number of aggregated features since each location must be processed with a unique weight vector. By combining feature aggregation through self-attention with feature transformation using element-wise perceptrons, we are able to create high-performing image classification architectures. Self-attention broadens the CNN's receptive field without increasing the computational cost by including very high kernel sizes. The deep learning models that must reflect global dependencies have become completely dependent on attention mechanisms. In specifically, by paying attention to every location in the same sequence, self-attention, also known as intra-attention, calculates the response at a place in a sequence.
The pairwise self-attention can be expressed mathematically as follows;

$$y_i = \sum_{j \in R(i)} \alpha\left(x_i, x_j\right) \odot \beta(x_j) \qquad \text{Eq. 3}$$

where $\odot$ denotes the Hadamard product, i represents the spatial index of the feature vector denoted by x_i, and $R(i)$ denotes the local print of the aggregation which is nothing but the set of all indexes that indicates which feature vectors are aggregated for constructing the output feature vector y_i. The feature vector represented by $\beta(x_j)$ are generated by the function β and these features are aggregated by the adaptive weight vectors denoted as $\alpha(x_i, x_j)$. The computed weights $\alpha\left(x_i, x_j\right)$ are used for combining the transformed features $\beta(x_j)$. The function α can be decomposed as follows;

$$\alpha(x_i, x_j) = \gamma(\delta\left(x_i, x_j\right)) \qquad \text{Eq. 4}$$

A single vector which represents the features x_i and x_j will be given as output by the relation function δ. The γ function later maps this single vector which is combined with the transformed features generated by $\beta(x_j)$.The dimension of the output from the γ function will not be same as the transformed features as the attention weights were shared among a set of channels. An unique feature of the pair-wise attention mechanism described above is that the feature vectors are processed independently and the computed weights does include only information related to location denoted by i and j. To incorporate the spatial information, the feature maps are augmented as follows;
Step 1. Normalize the horizontal and vertical coordinates along the feature map to a range of [-1, 1] in each dimension.
Step 2. Pass the normalized coordinates are sent to a trainable linear layer which maps them to respective range for m=network layers.
Step 3. The linear mapping function output a 2D feature p_i for every location available in the feature map.
Step 4. The relative location information is encoded after estimating the difference $p_i - p_j$.

Step 5. Augment the weight vectors $\alpha(x_i, x_j)$ by concatenating the $[p_i - p_j]$ before γ function. The self-attention mechanism described above was used to construct a residual block which shall implement both feature aggregation and transformation. The input feature vector is sent via two processing streams. The left unit asses the weights of the attention layer through mapping function δ and a mapper denoted as γ. The right unit reduces the dimensionality of the input by applying a linear transformation β. Then the output of the right and left units are aggregated through a Hadamard product. Finally the combined features are normalized and expanded to their original dimension C.

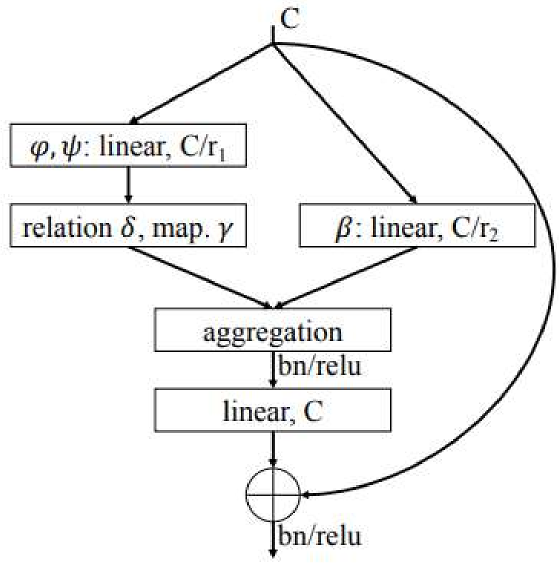

Fig. 2 Schematic view of a self-attention block [2]

3.3 Network Architecture

Residual networks are utilized as baselines in the proposed network architecture and are generally followed [12]. By layering 10 self-attention blocks at various resolutions, a classification model is created. ResNet50 is integrated with this architecture. The suggested architectural designs are wholly focused on self-attention. The five steps that make up the SAN's backbone, each with a distinct level of spatial resolution, result in a 32-resolution reduction factor. There are many self-attention blockages within each level. Transition layers, which decrease spatial resolution and increase channel dimensionality, connect successive stages. A classification layer that consists of global average pooling, a linear layer, and a softmax processes the output of the previous stage. Transition layers increase receptive field while decreasing computing load by diminishing spatial resolution. A batch normalization layer, a ReLU, 2x2 max pooling with stride 2, and a linear mapping that increases channel dimensionality are all included in the transition. The architecture of the self-attention block is presented in Fig. 2.

3.4 Performance Metrics

The performance of developed classifier trained to perform multi classification task was analyzed based on the following performance metrics.

- **Matthew's Correlation Coefficient** (MCC): With reference to the confusion matrix used for evaluating the binary classifiers performance, the MCC is an optimal metric suitable for evaluating the performance of a multi-classification model irrespective of the class imbalance. The coefficient value ranges between 0 and 1, a value closer to one indicates a good classification model and a value closer to zero indicates poor training result.

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad \text{Eq. 5}$$

- **Log Loss Function:** It is an error function that estimated the log loss termed as cross-entropy loss and hence the classifier that yields minimum value of the log function is the optimal one. It estimates the loss value based on the probability scores. Also it can be inferred that it calculates probability membership scores. This takes the uncertainty of the model in to account. The metrics which uses prediction labels for quantifying the model performance certainly hides the uncertainty of the model. It highly penalizes the test instances for which the model predicted class membership has lower scores.

$$\boldsymbol{loss = -\sum_{c=1}^{M} y_{o,c} \log(p_{o,c})} \quad \text{Eq. 6}$$

where M – number of classes present in the dataset

Y – a binary value indicating the correct classification for observation' o'

P – predicted probability score indicating observation o is of type c.

- **Cohen's Kappa Coefficient**: For qualitative (categorical) items, Cohen's kappa coefficient (k) is a statistic that is used to assess inter-rater reliability as well as intra-rater reliability. Since it considers the potential that the agreement could have happened by chance, it is typically believed to be a more reliable measurement than a simple % agreement estimate.

$$k = \frac{p_o - p_e}{1 - p_e} \quad \text{Eq. 7}$$

The relative concordance between actual and expected values, denoted by p_o, was observed. This represents the sum of any confusion matrix's diagonal cells divided by the sum of its non-diagonal cells. p_e is the likelihood that true values and false values will coincide by accident.

4. EXPERIMENTS AND RESULTS

As the input size for the network architecture is fixed, the images were scaled to 224x224. A bi-cubic technique was used to resize the photos. The model underwent 100 training epochs. The training set's image samples were split into two distinct, non-overlapping groups for training and validation. The ideal split of the data samples for training and testing was achieved with the aid of the k-fold cross validation process. In order to provide more training images for the training, flip, rotate, and elastic deformation were applied to the training images. The architecture weights were initialized with random weight values before the training. The initial learning rate was set at 0.0001, and the training period batch size was set at 16. While image augmentation by flipping and rotating keeps the original shape and size of the tumors visible in the cancer images, elastic augmentation shifts each pixel, distorting the image. It's possible that some tumors would no longer be visibly related to others that underwent elastic deformation to become malignant tumors. It was shown that elastic augmentation increased accuracy up to 88.64 percent while flip and rotation picture augmentation increased accuracy up to 91.83 percent.

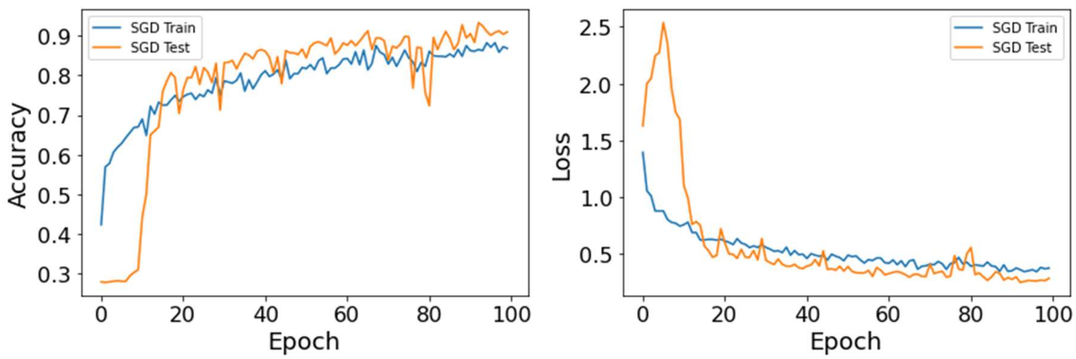

Fig. 3 Analysis of Accuracy and Loss for Attention based classifier with SGD optimization

On images that have been rotated and flipped, the experiment tests trained networks. Images from the figshare dataset are rotated and flipped in this experiment in one of two different directions: clockwise 90°, and inverted flip about the horizontal axis. This is zero-shot testing because no such manipulations were used during training. Table 2 presents the findings. Given that pairwise self-attention is essentially a set operator, it was hypothesised that pairwise self-attention models would be more resistant to this type of manipulations than convolutional networks (or patchwise self-attention). In fact, we find that patchwise or convolutional self-attention networks are more susceptible than pairwise self-attention models, though the domain shift affects all networks.

Table 2. Performance comparison of proposed self-attention based model with Conv.ResNet model

Method	NoRotation	Clockwise90	Elastic Augmentation
ResNet50	86.40	85.23	83.9
SAN with 10 blocks	91.54	90.45	88.46

A higher MCC value for the attention model classification is possible because the prediction performed well in all four categories of the confusion matrix (true positives, false negatives, true negatives, and false positives), proportionally to the size of the positive elements and the size of the negative elements in the dataset. The Matthews correlation coefficient (MCC) is a more reliable statistical rate. High value of kappa coefficient exhibits that the proposed model shall perform better even if the training samples are highly imbalanced.

Table 3. Summary of Performance Metrics (SGD with LR = 0.0001)

Matthew's Correlation Coefficient	Cohen's Kappa coefficient	Log Loss
0.9406	0.9385	**0.1632**

5. CONCLUSION

The efficacy of image classification model that is wholly based on self-attention has been investigated in this work. In this case, a set operation known as pairwise self-attention is used, which is fundamentally distinct from convolution. In order to effectively change parameters across both spatial dimensions and channels, vector attention is used. Numerous important results from our experiments are obtained. The performance of networks built solely on paired self-attention surpasses the efficiency of convolutional baselines. This shows that convolutional networks are not necessary for the efficacy of deep learning in computer vision; there is another way to achieve comparable or higher discriminatory power, with different and potentially advantageous structural properties like permutation- and cardinality-invariance.

Reference

1. Abiwinanda, Nyoman, et al. "Brain tumor classification using convolutional neural network." *World Congress on Medical Physics and Biomedical Engineering 2018: June 3-8, 2018, Prague, Czech Republic (Vol. 1)*. Springer Singapore, 2019.
2. Zhao, Hengshuang, Jiaya Jia, and Vladlen Koltun. "Exploring self-attention for image recognition." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2020.
3. Xie, Yuting, et al. "Convolutional neural network techniques for brain tumor classification (from 2015 to 2022): Review, challenges, and future perspectives." *Diagnostics* 12.8 (2022): 1850.
4. Ari, Ali, and Davut Hanbay. "Deep learning based brain tumor classification and detection system." *Turkish Journal of Electrical Engineering and Computer Sciences* 26.5 (2018): 2275-2286.
5. Li, Zhenliang, et al. "Deep multi-instance learning with induced self-attention for medical image classification." *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2020.
6. Zhao, Hengshuang, Jiaya Jia, and Vladlen Koltun. "Exploring self-attention for image recognition." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2020.
7. Takahashi, Ryo, Takashi Matsubara, and Kuniaki Uehara. "A novel weight-shared multi-stage CNN for scale robustness." *IEEE Transactions on Circuits and Systems for Video Technology* 29.4 (2018): 1090-1101.

8. Kan, Tao, et al. "Convolutional neural networks based on fractional-order momentum for parameter training." *Neurocomputing* 449 (2021): 85-99.
9. Hernández-García, Alex, and Peter König. "Do deep nets really need weight decay and dropout?." *arXiv preprint arXiv:1802.07042* (2018).
10. Figueiredo, P.; Figueiredo, I.; Pinto, L.; Kumar, S.; Tsai, Y.; Mamonov, A. Polyp detection with computer-aided diagnosis in white light colonoscopy: Comparison of three different methods. Endosc. Int. Open 2019, 7, E209–E215.Diagnostics 2022, 12, 1850 42 of 46
11. Yeung, M.; Sala, E.; Schönlieb, C.B.; Rundo, L. Focus U-Net: A novel dual attention-gated CNN for polyp segmentation during colonoscopy. Comput. Biol. Med. 2021, 137, 104815.
12. Gong, J.; Liu, J.Y.; Sun, X.W.; Zheng, B.; Nie, S.D. Computer-aided diagnosis of lung cancer: The effect of training data sets on classification accuracy of lung nodules. Phys. Med. Biol. 2018, 63, 035036.
13. Nishio, M.; Sugiyama, O.; Yakami, M.; Ueno, S.; Kubo, T.; Kuroda, T.; Togashi, K. Computer-aided diagnosis of lung nodule classification between benign nodule, primary lung cancer, and metastatic lung cancer at different image size using deep convolutional neural network with transfer learning. PLoS ONE 2018, 13, e0200721.
14. Tian, Q.; Wu, Y.; Ren, X.; Razmjooy, N. A new optimized sequential method for lung tumor diagnosis based on deep learning and converged search and rescue algorithm. Biomed. Signal Process. Control 2021, 68, 102761.
15. Schmidhuber, J. Deep learning in neural networks: An overview. Neural Netw. 2015, 61, 85–117.
16. Hu, A.; Razmjooy, N. Brain tumor diagnosis based on metaheuristics and deep learning. Int. J. Imaging Syst. Technol. 2021, 31, 657–669.
17. Tandel, G.S.; Balestrieri, A.; Jujaray, T.; Khanna, N.N.; Saba, L.; Suri, J.S. Multiclass magnetic resonance imaging brain tumor classification using artificial intelligence paradigm. Comput. Biol. Med. 2020, 122, 103804.
18. Shen, D.; Wu, G.; Suk, H.I. Deep learning in medical image analysis. Annu. Rev. Biomed. Eng. 2017, 19, 221–248.
19. Tandel, G.S.; Balestrieri, A.; Jujaray, T.; Khanna, N.N.; Saba, L.; Suri, J.S. Multiclass magnetic resonance imaging brain tumor classification using artificial intelligence paradigm. Comput. Biol. Med. 2020, 122, 103804.
20. Yasaka, K.; Akai, H.; Kunimatsu, A.; Kiryu, S.; Abe, O. Deep learning with convolutional neural network in radiology. Jpn. J. Radiol. 2018, 36, 257–272.
21. Miotto, R.; Wang, F.; Wang, S.; Jiang, X.; Dudley, J.T. Deep learning for healthcare: Review, opportunities and challenges. Brief. Bioinform. 2018, 19, 1236–1246.
22. Deepak, S.; Ameer, P.M. Brain tumor classification using deep CNN features via transfer learning. Comput. Biol. Med. 2019, 111, 103345.
23. Ge, C.; Gu, I.Y.H.; Jakola, A.S.; Yang, J. Deep semi-supervised learning for brain tumor classification. BMC Med. Imaging 2020, 20, 87.
24. B. E. Manjunath Swamy. "Personalized Ranking Mechanism Using Yandex Dataset on Machine Learning Approaches." Proceedings of the International Conference on

Cognitive and Intelligent Computing: ICCIC 2021, Volume 1. Singapore: Springer Nature Singapore, 2022.

25. Burada, Sreedhar,"Computer-Aided Diagnosis Mechanism for Melanoma Skin Cancer Detection Using Radial Basis Function Network." Proceedings of the International Conference on Cognitive and Intelligent Computing: ICCIC 2021, Volume 1. Singapore: Springer Nature Singapore, 2022.
26. Kumar, M. S, et al. "Deep Convolution Neural Network Based solution for Detecting Plant Diseases." Journal of Pharmaceutical Negative Results (2022): 464-471.
27. Sreedhar, B., BE, M.S. and Kumar, M.S., 2020, October. A comparative study of melanoma skin cancer detection in traditional and current image processing techniques. In 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC) (pp. 654-658). IEEE.
28. P. Sai Kiran, and M. S Kumar. "Resource aware virtual machine placement in IaaS cloud using bio-inspired firefly algorithm." Journal of Green Engineering 10 (2020): 9315-9327.
29. Balaji, K., P. Sai Kiran, and M. Sunil Kumar. "Power aware virtual machine placement in IaaS cloud using discrete firefly algorithm." Applied Nanoscience (2022): 1-9.
30. Ananthanatarajan, V., Kumar, M. S., & Tamizhazhagan, V. (2020). Forecasting of wind power using lstm recurrent neural network. Journal of Green Engineering, 10.Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In CVPR, 2018.
31. Fei Wang, Mengqing Jiang, Chen Qian, Shuo Yang, Cheng Li, Honggang Zhang, Xiaogang Wang, and Xiaoou Tang. Residual attention network for image classification. In CVPR, 2017.
32. Xiaolong Wang, Ross Girshick, Abhinav Gupta, and Kaiming He. Non-local neural networks. In CVPR, 2018.
33. Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. CBAM: Convolutional block attention module. In ECCV, 2018.
34. Hengshuang Zhao, Yi Zhang, Shu Liu, Jianping Shi, Chen Change Loy, Dahua Lin, and Jiaya Jia. PSANet: Point-wise spatial attention network for scene parsing. In ECCV, 2018.
35. Xizhou Zhu, Han Hu, Stephen Lin, and Jifeng Dai. Deformable ConvNets v2: More deformable, better results. In CVPR, 2019.