A novel CNN for Low-Grade Gliomas classification on MRI

Jun Liu[1,2], Geng Yuan[2], Wenbin Zhang[3],Xue Lin[2], XiaoLin Xu[2], Yanzhi Wang[2]

[1]Robotics Institute, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA
[2] Department of Electrical & Computer Engineering, College of Engineering, Northeastern University, Boston, MA, USA
[3] University of Maryland, Baltimore County, MD USA

Abstract. Codeletion of chromosomal arms 1p/19q has been connected to a good response to treatment in low-grade gliomas (LGG) in several studies. The ability to predict 1p19q status is critical for treatment planning and patient follow-up. The goal of this research is to develop a non-invasive approach based on MR images using our novel convolutional neural networks. Although public networks such as VggNet, GoogleNet and other well-known public networks can effectively diagnose brain cancers using transfer learning, but the model contains a huge number of components that are unrelated to medical image. As a result, the diagnostic results are unreliable by transfer learning model. In order to address the issue of trustworthiness, we build the model from the bottom up, rather than relying on transfer learning. Our network structure flexibly uses a deep convolution stack mixed with a dropout and dense operation, which reduced overfitting and enhanced performance. We also augmented the given dataset. The Gaussian noise is introduced during the model training. We use three--fold cross-validation to train the best selection model. Our proposed model is compared to MobileNetV2, InceptionResNetV2 and VGG16 that have been fine-tuned through transfer learning. Our model achieves better results than these models. When classifying images between 1p/19q codeletion and not codeleted, the proposed architecture achieved an F1-score of 96.37%, Precision 97.46%, Recall 96.34% in the test set.

Keywords. low-grade gliomas; 1p/19q; CNN; reliability transfer learning; radiogenomics.

1. Introduction

Low-grade gliomas(LLG) [1] are brain tumors that arise from astrocytes and oligodendrocytes, two separate types of brain cells(1). Low-grade gliomas can cause a variety of symptoms depending on where they are in the brain. The tumor in the area of the brain that governs language may prevent the patient from speaking or understanding. A brain tumor diagnosis can be devastating for patients. The majority of tumors are discovered as a result of a symptom that prompts doctors to perform a brain MRI or CT scan.
Magnetic Resonance Imaging (MRI) is the most effective method for detecting brain malignancies. The scans generate a tremendous amount of picture data. The radiologist examines these images.

Tumors of the brain are difficult to diagnose and treat. The sizes and locations of brain tumors vary dramatically. As a result, fully comprehending the nature of the tumor is quite challenging. For MRI analysis, a qualified neurosurgeon is required. The absence of skilled doctors and a lack of information regarding tumors can make generating reports from MRI's extremely difficult and time-consuming. Because of the intricacies involved in brain tumors and their qualities, a manual examination can be prone to errors. Machine Learning based automated classification systems have consistently outperformed manual classification.

The study of the relationship between cancer imaging features and gene expression is known as radiogenomics. Biomarkers that determine the genetics of a disease without the use of an intrusive biopsy can be created using radiogenomics. A biomarker is a biological indicator of some state or condition. The presence or lack of biomarkers is important in avoiding intrusive biopsies because certain treatments for brain tumors are more successful in the presence or absence of a biomarker. The detection of biomarkers can ensure that patients receive the most effective treatment for their specific situation [2].

Low-grade gliomas (LLG) [2,3,4] are tumors that are thought to be formed from glial cells, have infiltrative development, and lack malignant histopathological characteristics. 1p/19q chromosome co-deletion is one of biomarkers that appear to be important for Low-grade gliomas. When 1p/19q co-deletion is discovered in low-grade gliomas, studies demonstrate that they respond better to chemotherapy and radiotherapy. The novelty and promising results of combining deep learning with radiogenomics are what make this study noteworthy. The detection of 1p/19q co-deletion using deep learning works better with T2 images than with T1 post contrast images[2].

In 2017 deep learning was firstly used by Akkus et al. [2] to predict 1p19q from LGG MR images, Image registration, tumor segmentation, and CNN-based 1p/19q status classification are the three primary steps of their method. When data augmentation is not performed, their multiscale CNNs overfit to the original training data. Chelghoum et al., 2020[4] used popular public networks, including AlexNet, VGG19, GoogleNet, etc. for 1p19q categorization through transfer learning.[5,6] According to their description, even with limited datasets, the results offered by transfer learning are robust. Abiwinanda et al. [8] used five different CNN designs, with the second design containing two convolutional layers, the Maxpool layer and ReLU layer, followed by hidden 64 neurons, achieving the highest accuracy.

Why are there just thousands of training examples, Maithra Raghu et al. [9] wondered. They looked upon transfer learning in small data settings. They discovered that there was a significant performance difference between transfer learning and training from scratch for a big model (ResNet), but not for a smaller model. For a little amount of data, the huge model built for ImageNet can have too many parameters. They discovered that transfer learning provides limited performance increases for the evaluated medical imaging tasks after a rigorous performance evaluation and examination of hidden representations of neural networks. Transfer learning had little effect on the performance of medical imaging tasks, and the model trained from the ground up was nearly as well as the ImageNet transfer model.

The following are our contributions:

- Using convolution stack, we create a dedicated convolutional neural network for detecting brain tumors on MRI images.
- During training, we use a customizable combination of dropout and Gaussian noise to reduce overfitting and increase performance.
- Our proposed model is compared to MobileNetV2, InceptionResNetV2 and VGG16 that have been fine-tuned through transfer learning.

2. Materials and Methods

We use the provided dataset to train our planned network. Meanwhile, we use a transfer learning technique to fine-tune MobileNetV2, InceptionResNetV2, and VGG16 on the same dataset. Then we compare the performance of our trained model with these models

2.1. **Experimental Data**

The Kaggle Open Datasets[10] provided the Brain MRI dataset that was utilized to evaluate the planned study. The dataset contains 253 brain MRI images in two folders: yes and no. There are 155 tumorous brain MRI images in the folder yes and there are 98 non-tumorous brain MRI images in the folder no.

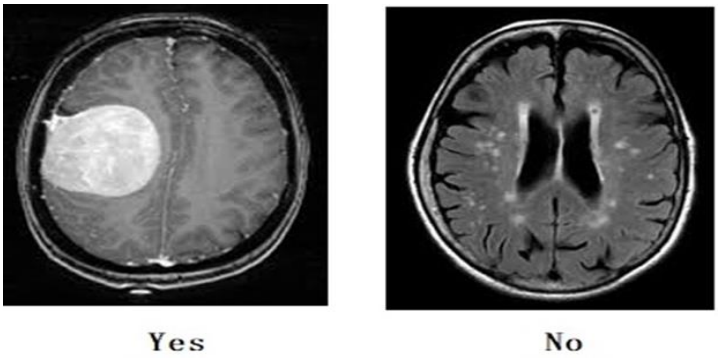

Figure 1. Brain MRI

In Figure 1. The left image Brain with Tumor and the right image is Brain without Tumor.

2.2. **Network Architectures**

In Figure 1, there are 14 layers in the model. convolutional kernel with smaller convolutions - 3 x 3 - were found to produce positive outcomes, as these smaller convolutions may capture some of the finer characteristics of the edges. This network's convolutional layers all employ 3 x 3 kernels. It is starting with 16 kernels per layer, the architecture progresses to 32 kernels per layer, 64 kernels per layer, and finally 128 kernels per layer.

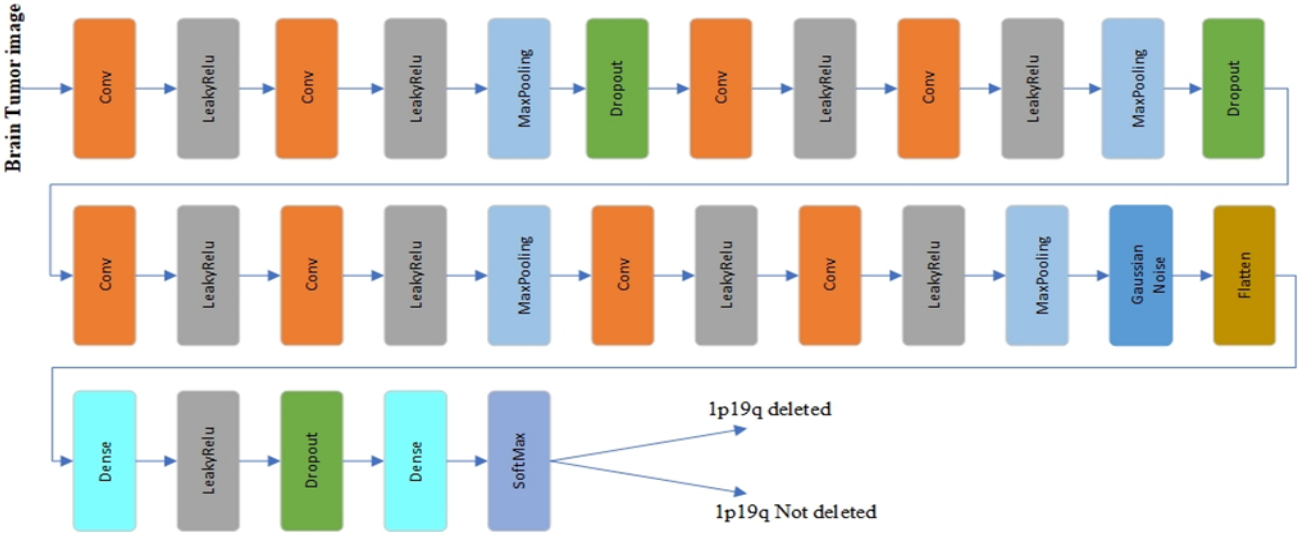

Figure 2 Network Architectures

This network depicted in the Figure2 is made up of convolution layers, pooling layers, dropout layers, LeakyRelu layer, Dense layers[11], Flatten layer, Softmax layer, with the input picture.

Type	Filter Shape	Input Size
InputLayer	N/A	256 x 256 x 1
Conv2D	16 x 3 x 3	254 x 254 x 16

LeakyReLU	N/A	254 x 254 x 16
Conv2D	16 x 3 x 3	252 x 252 x 16
LeakyReLU	N/A	252 x 252 x 16
MaxPooling	2 x 2	126 x 126 x 16
Dropout	N/A	126 x 126 x 16
Conv2D	32 x 3 x 3	124 x 124 x 32
LeakyReLU	N/A	124 x 124 x 32
Conv2D	32 x 3 x 3	122 x 122 x 32
LeakyReLU	N/A	122 x 122 x 32
MaxPooling	2 x 2	61 x 61 x 32
Dropout	N/A	61 x 61 x 32
Conv2D	64 x 3 x 3	59 x 59 x 64
LeakyReLU	N/A	59 x 59 x 64
Conv2D	64 x 3 x 3	57 x 57 x 64
LeakyReLU	N/A	57 x 57 x 64
MaxPooling	2 x 2	28 x 28 x 64
Conv2D	128 x 3 x 3	26 x 26 x 128
LeakyReLU	N/A	26 x 26 x 128
Conv2D	128 x 3 x 3	24 x 24 x 128
LeakyReLU	N/A	24 x 24 x 128
MaxPooling	5 x 5	4 x 4 x 128
GaussianNoise	N/A	4 x 4 x 128
Flatten	N/A	2048
Dense	N/A	1024
LeakyReLU	N/A	1024
Dropout N/A	1024	1024
Dense N/A	2	2
Softmax N/A	2	2

Table 1

Table 1 specifies the network activities utilized by each layer, as well as the size of the convolution kernel and the size of the input.

- All images of brain tumors that are fed into the network are scaled to 256 x 256.
- This network uses 3 x 3 kernels for all convolutional layers. The architecture starts with 16 kernels per layer, then 32, 64, and finally 128 kernels per layer.
- Because negative values are kept and saturation concerns are avoided when employing tanh, LeakyReLU was chosen as the activation function.
- This model employs 2 x 2 maxpooling at first, then 7 x 7 maxpooling later. The number of neurons in a layer is reduced when it is dense.
- Dense(fully connect) is used twice just before the softmax layer to reduce the number of neurons to two, reflecting the binary prediction of either 'co-deletion' or 'no co-deletion.'.
- Gaussian Noise was purposely supplied to the training data to minimize overfitting which can think of it as a form of random data augmentation. For corrosion processes with genuine inputs, Gaussian noise (GS) is a natural choice.During training, the error is reduced, and at the same time, the interference items generated by noise are punished to achieve the purpose of reducing the square of the weight. We set standard deviation of the noise distribution is 0.5.
- The goal of Flatten is to one-dimensionalize multi-dimensional input, which is accomplished by transitioning from the convolutional to fully connect layers.

- During the forward and backward propogation phases, Dropout avoids neurons at random. The amount of neurons that aren't updating is determined by the Dropout value. The dropout rate was set to 0.3.
- The probability of each of the binary outcomes - 'co-deletion' and 'no co-deletion' - is included in the output layer using a softmax classifier.

2.3. **Hyperparameters**

Several hyperparameter values were explored over this study.

Learning rate

The learning rate is the amount of time we spend moving in a particular direction in order to find the global minima. Starting with a greater learning rate usually works fine because initial weight values are rather random. We typically grow closer and closer to either the global or local minima as we proceed through the training phase. Because we don't want to overshoot the minima, annealing the learning rate is a typical method. To put it another way, as the training phase progresses, we begin to take smaller and smaller moves in a specific direction. When there is no change in the loss value, we will continue to reduce the learning rate by the square root of 0.1 until it reaches a reduction of 0.5e-6.

EarlyStopping

Overfitting models to training data can be prevented or limited by early stopping techniques. Also, when the findings are static, early halting procedures prevent needless computations. If there hasn't been a change of at least 0.001 in 10 (epochs), the model provided here will terminate training.

Batch Size

The batch size specifies how many photos are handled during forward propagation to produce a loss value for backpropagation. Batch size is typically set to a power of two and is restricted by available memory. Furthermore, while a bigger batch size allows for faster training, weights update less frequently and may not deliver the greatest outcomes.

Number of Epochs

When training your model, the number of epochs denotes the number of times the complete training dataset is iterated over. Validation determines how well the model generalizes to new data at the end of each epoch.

Number of Epochs

When training your model, the number of epochs denotes the number of times the complete training dataset is iterated over. Validation determines how well the model generalizes to new data at the end of each epoch.

2.4. **Experiments**

For the experiments, we used the Keras Python module with TensorFlow as the backend.

2.4.1. **Data Preparation**

Data preparation steps are included deleting a third class, standardizing the data, and implementing cross-validation [12], shuffle the training data. Because this is a small dataset, there were insufficient examples to train the neural network. In addition, data augmentation was useful in addressing the data imbalance issue.

The image is preprocessed before being processed into the proposed structure. The original MR image is scaled to 225 x 225 1 pixels in the first step. Image augmentation techniques such as flipping, mirroring, and rotating are used to generate redundant data for the network, which is frequently used to avoid network overfitting and improve system resilience.

2.4.2. **Proposed Workflow**

We build evaluate and train our model to improve performance and using cross-validation in model training is depicted in In Figure3.

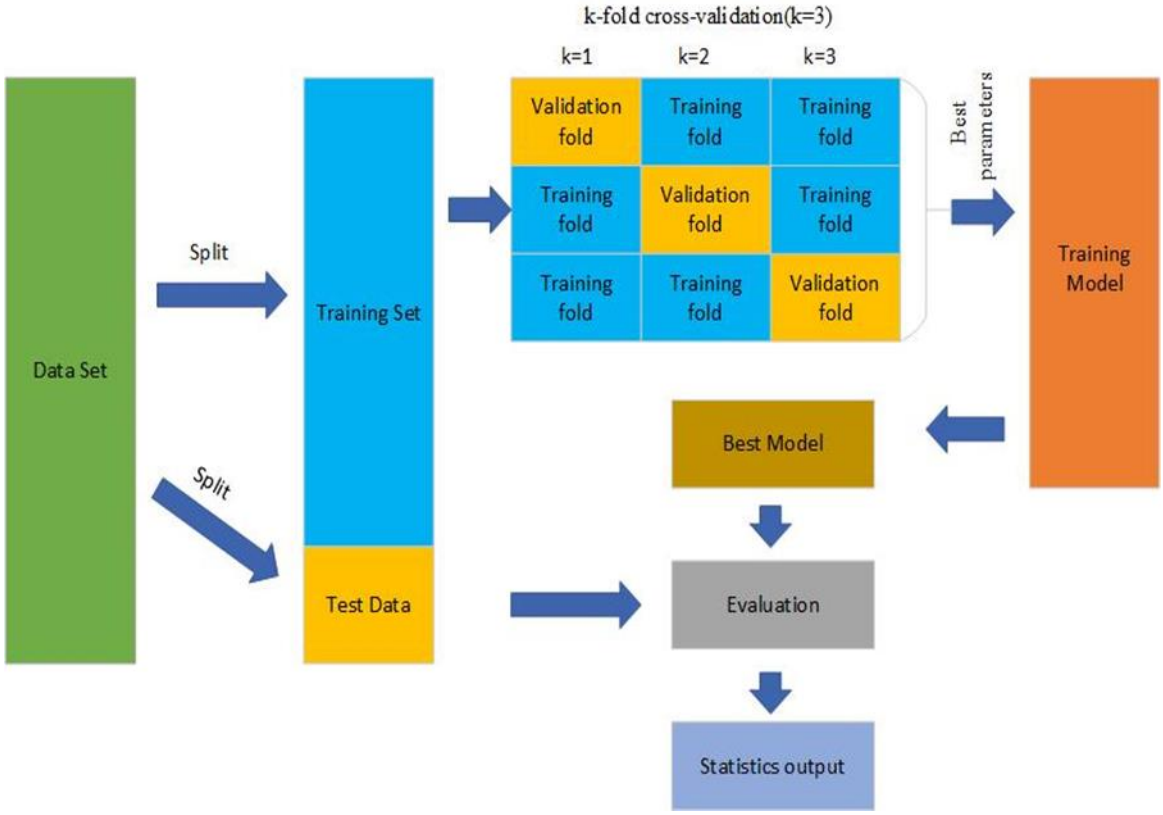

Figure 3 Proposed Workflow

- We divided the data into training and testing datasets at random and built the model using the training set, and estimating its accuracy using the test set.
- Then, we acquire the best quality model by fine-tuning the model via 3-fold cross-validation to enhance the estimate's accuracy.
- On the test set, evaluate the model's expected accuracy.
- Output evaluation statistics include Precision, Recall, F1-score and Confusion matrix.

2.4.3. Evaluation method

Confusion Matrix

A confusion matrix is a method of analyzing a classification algorithm's performance. If you have an unbalanced amount of observations in each class or if your dataset has more than two classes, classification accuracy alone can be misleading.

	1p19q deleted	1p19q not deleted
1p19q deleted	True Positive	False Positive
1p19q not deleted	False Negative	True Negative

Table 2 Confusion Matrix

In Table 2 we can clearly see the number of correct identifications and the number of incorrect identifications for each category.

Precision

The percentage of 1p/19q co-deleted patients properly predicted by the model based on the total number of patients with 1p/19q co-deletion is referred to as Precision.

Precision = True Positive / (True Positive + False Negative)

Recall

The percentage of 1p/19q non-deleted patients detected by the model to all 1p/19q non-deleted patients is used to calculate specificity.

Recall = True Negative / (True Negative + False Positive)

F1 score

The F1 score's purpose is to combine the precision and recall measurements into a single number. It is important metrics for class imbalance problem, due to an imbalance in the number of brain and non-brain tumors in this brain MRI dataset the F1 score was created to operate effectively with data that is unbalanced. Its formula is as follows:

F1score = 2 * (precision * recall) / (precision + recall)

3. **Results**

The harmonic mean of precision and recall is calculated using the f1-score. The scores for each class indicate how accurate the classifier was in classifying the data points in that class in comparison to all other classes. The number of samples of the true response that fall into that class constitutes the support.

Fold	Train/Test	precision	recall	f1-score	support
Fold-1	Train	0.9850	0.9562	0.9704	274
	Test	0.9517	0.9841	0.9598	477
Fold-2	Train	0.9881	0.9679	0.9241	274
	Test	0.9851	0.9635	0.9742	477
Fold-3	Train	0.9886	0.9526	0.9703	274
	Test	0.9517	0.9841	0.9598	477

Table 3. Classification performance

We training the model use three-fold cross validation. In terms of F1-score, precision, recall, Table 3 demonstrates that model achieves good values.

type	precision	recall	f1-score	support
1p19q deleted	0.9881	0.9635	0.9742	125
1p19q not deleted	1.0	0.9444	0.9375	39
avg / total	0.9746	0.9634	0.9637	164

Table 4 Statistics for test set

In the test set we employed 164 photos, 125 of which are 1p19q deleted and 39 of which are 1p19q not deleted, the suggested architecture received an f1-score of 0.9637 in Table 4

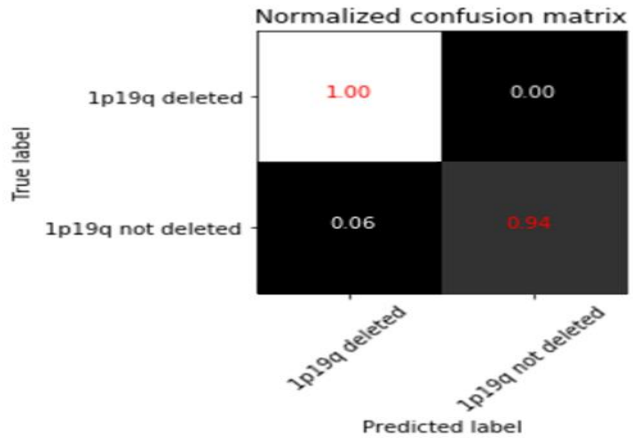

Figure 5 Confusion matrix

The accompanying Figure 5 shows the confusion matrix for the classification of 1p19q status on test set. We can be certain that all 125 1p19q deleted pictures were detected accurately.
We compare MobileNetV2, InceptionResNetV2, VGG16 which we fine-tuned using transfer learning approach.

Model	Precision	Recall	F1 score
Ours	0.9746	0.9634	0.9637
MobileNetV2	0.9667	0.9355	0.9497
InceptionResNetV2	0.96429	0.87097	0.9153
VGG16	0.961538	0.806452	0.8763

Table 5 Performance comparison

From Table 5 We discovered that the Precision of our model and the remaining models is very close. In terms of Rcall, Our model outperforms MobileNetV by 2% and much higher than InceptionResNetV2 and VGG16 . On the F1 score side, our model outperforms MobileNetV2 by over 2% and InceptionResNetV2 and VGG16 by over 5%. As the comprehensive indicator of Precision and Recall, F1 score can see the model's performance in recognition more clearly.

4. Discussion

We provide a reliable and non-invasive approach for predicting 1p/19q chromosomal arm deletion in this work. Having a sufficient amount of data sets is a significant difficulty when applying deep learning approaches to medical imaging. Despite the fact that the initial data quantity was limited, artificially enhancing data increased the volume of our data. With larger patient populations and more varied data, it's possible that additional performance gains will be gained.
As large convolution kernels are inefficient in terms of cost. We are reducing the amount of irrelevant features conceivable by restricting the number of parameters. This drives the deep learning algorithm to learn traits that are common to a variety of scenarios, allowing it to generalize more effectively. Smaller odd-sized kernel filters would be preferable. However, 1x1 is removed from the list of possible ideal filter sizes since the features recovered would be fine-grained and local, with no information from nearby pixels. Furthermore, it does not extract any useful features. Through experiments, we found that although VggNet also uses a 3x3 convolution kernel, it is prone to overfitting due to the complexity of the network and dataset is small. As a result, Vgg16 categorization the precision and recall of 1p/19q chromosomal arm deletion are not very good.
Because of the deep architecture of current networks like GoolgeNet and ResNet, feature maps from these networks frequently have a very large receptive field. However studies [13] reveal that the network gathers information from a considerably narrower portion of the receptive field, which is

referred to as the valid receptive field in this research. In this experiment we found that the recall rate was not high by using InceptionResNetV2 and Vgg16. As a result, a large receptive field does not increase the performance of medical images with samll datasets considerably.
We discovered that MobileNetV2 is significantly higher than InceptionResNetV2 and VGG16 In the fields of Precision, Recall and F1 score.
We discovered that MobileNetV2 is significantly higher than InceptionResNetV2 and VGG16 In the fields of Precision, Recall and F1 score. It employs Depth-Wise Separable Convolutions and divides an ordinary 3x3 convolution into two convolutions, which is the same as the 3x3 convolution we employ.

Since medical imaging data is scarce, transfer learning approaches are used to fine-tune medical imaging models employing various public popular models (e.g. VggNet, GoolgeNet, etc.) generated using large public ImageNet datasets. However, these models create a large number of characteristics that are unrelated to medical imaging, jeopardizing the accuracy of medical diagnosis [14].Our model does not involve transfer learning, and the parameters it generates are specific to the medical imaging dataset that was used. As a result, the reliability of brain tumor diagnosis has substantially improved. Small dataset can result in substantial training errors.
The model's capacity to learn mapping rules from the input space can be increased by adding Gaussian noise during training, as can the model's generalization ability and fault tolerance. Adding noise means that the network cannot remember training samples because they are changing all the time, resulting in smaller network weights and a more robust network with lower generalization error. Because new samples are selected from the domain adjacent known samples, the structure of the input space is smoothed. This smoothing may make it easier for the network to learn the mapping function, resulting in better and faster learning. After adding Gaussian noise to our model training, we can see a big improvement in performance.

5. Conclusion

The results of our CNN technique for predicting 1p/19q codeletion status non-invasively are promising. We create a brain tumor detection model that does not rely on transfer learning. Our network structure employs deep convolution stack strategy when training with Gaussian noise, reducing overfitting and improving performance. Compared to transfer learning models, our model gives more accurate findings.

6. Reference

1.Amin, J., Sharif, M., Haldorai, A. et al. Brain tumor detection and classification using machine learning: a comprehensive survey. Complex Intell. Syst. (2021). https://doi.org/10.1007/s40747-021-00563-y
2.Akkus, Z., Ali, I., Sedlář, J. et al. Predicting Deletion of Chromosomal Arms 1p/19q in Low-Grade Gliomas from MR Images Using Machine Intelligence. J Digit Imaging 30, 469–476 (2017). https://doi.org/10.1007/s10278-017-9984-3
3.Bhattacharya, D., Sinha, N. & Saini, J. Determining chromosomal arms 1p/19q co-deletion status in low graded glioma by cross correlation-periodogram pattern analysis. Sci Rep 11, 23866 (2021). https://doi.org/10.1038/s41598-021-03078-1
4.Lombardi G, Barresi V, Castellano A, Tabouret E, Pasqualetti F, Salvalaggio A, Cerretti G, Caccese M, Padovan M, Zagonel V, Ius T. Clinical Management of Diffuse Low-Grade Gliomas. Cancers (Basel). (2020) Oct 16;12(10):3008. doi: 10.3390/cancers12103008.
5.Alanazi MF, Ali MU, Hussain Chelghoum R., Ikhlef A., Hameurlaine A., Jacquir S.Transfer learning using convolutional neural network architectures for brain tumor classification from MRI images (2020), 10.1007/978-3-030-49161-1_17

6.SJ, Zafar A, Mohatram M, Irfan M, AlRuwaili R, Alruwaili M, Ali NH, Albarrak AM. Brain Tumor/Mass Classification Framework Using Magnetic-Resonance-Imaging-Based Isolated and Developed Transfer Deep-Learning Model. Sensors (Basel). 2022 Jan 4;22(1):372. doi: 10.3390/s22010372. PMID: 35009911; PMCID: PMC8749789.
7.Swati ZNK, Zhao Q, Kabir M, Ali F, Ali Z, Ahmed S, Lu J. Brain tumor classification for MR images using transfer learning and fine-tuning. Comput Med Imaging Graph. 2019 Jul;75:34-46. doi: 10.1016/j.compmedimag.2019.05.001.
8.Abiwinanda N., Hanif M., Hesaputra S.T., Handayani A., Mengko T.R.Brain tumor classification using convolutional neural network World congress on medical physics and biomedical engineering 2018, Springer, Singapore (2019), pp. 183-189
9.Raghu, M.; Zhang, C.; Kleinberg, J.; Bengio, S. Transfusion: Understanding Transfer Learning for Medical Imaging, NeurIPS 2019. Available online: https://arxiv.org/abs/1902.07208 (accessed on 19 2 2022).
10.Brain MRI Images for Brain Tumor Detection, https://www.kaggle.com/navoneel/brain-mri-images-for-brain-tumor-detection
11.Keras API reference / Layers API / Core layers / Dense layer, https://keras.io/api/layers/core_layers/dense/
12.Ashalata Panigrahi, Manas R. Patra, Chapter 6 - Network Intrusion Detection Model Based on Fuzzy-Rough Classifiers,Editor(s): Pijush Samui, Sanjiban Sekhar, Valentina E. Balas,Handbook of Neural Computation,Academic Press,2017,Pages 109-125,ISBN 9780128113189, https://doi.org/10.1016/B978-0-12-811318-9.00006-5.
13.Chao Peng, Xiangyu Zhang, Gang Yu, Guiming Luo, Jian Sun,Large Kernel Matters ——Improve Semantic Segmentation by Global Convolutional Network, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 4353-4361https://doi.org/10.48550/arXiv.1703.02719
14.Liu, J.; Deng, F.; Yuan, G.; Lin, X.; Song, H.; Wang, Y. An Explainable Convolutional Neural Networks for Automatic Segmentation of the Left Ventricle in Cardiac MRI. In Proceedings of the CECNet 2021, Beijing, China, 18–21 November 2021. https://doi.org/10.3233/FAIA210416.