Detecting Brain Tumors with Brain MRI Images with Image Augmentation.

AI Technology and Systems

www.ai-techsystems.com

Prashant Arora

Machine Learning Intern

at

***Abstract*— Artificial Intelligence in today's time is being used for various tasks in different fields from healthcare to technology. This paper is one such of use case of Artificial intelligence in detecting brain tumors from Brain MRI images. The main problem for such classification is the requirement of enough dataset , hence we will be using Data augmentation techniques to increase dataset size. The last step will be to compare training accuracy using the F1-score calculated on two different models with same architecture , just with different datasets .**

Keywords—Data Augmentation , classification , Artificial Intelligence .

I. INTRODUCTION (*HEADING 1*)

Object detection or Object Recognition is a basic machine learning task in artificial intelligence. Starting Classification problems from detecting dogs , cats or different animals to detection problems in healthcare. There are less problems in former as there are multiple datasets available , but in the later case there is no such large dataset available. Hence , another task that begins is of Data Augmentation to increase the datasets. Here we will be explaining ways of data augmentation , acquiring big datasets and different results acquired while training.

The increase of tumor in side brain leads to brain tumor. Brain Tumor can be identified with various numbers of symptoms like mood changing , seizures , difficulty in walking or listening , stiffness in mascular movement ,etc. If detected in starting stages (primary stage) , brain tumor can easily be curated or removed , but in final stage it leaves effects in body after removal also. This is because of the part of the brain , it occurs .

Brain tumor can be identified using :

1. MRI scans
2. CT scans
3. Ultra sound etc.

II. METHODOLOGY

1. Collecting Dataset

The first step was to search for datasets available on web , having enough MRI scans which may be suitable for training purpose. Thanks to kaggle , someone has created a dataset on kaggle having the structure the of dataset as shown in figure below.

Folder Name	*Number of files Present*
Yes	155
No	98

The folder named as “yes” contains 155 positive images having brain tumor , and the other “no” folder contains the negative samples.

2. GENERATING NEW DATA

After Collecting Data , it was seen that the data was not suitable for training. So now we needed to generate more data from it , hence used augmentation techniques for doing it.

3. AUGMENTING DATA

A Deep learning Framework , Keras provides a class specially for preprocessing images to generate new data fromavailable dataset, named as ImageDataGenerator.

DataImageGenerator class generates batches of tensor image data with real-time data augmentation. These batches created can be looped over to retrieve those augmented tensor images.

We performed Data augmentation in two sets , in which we increased data-set size to 10 times and 20 times of original data-set

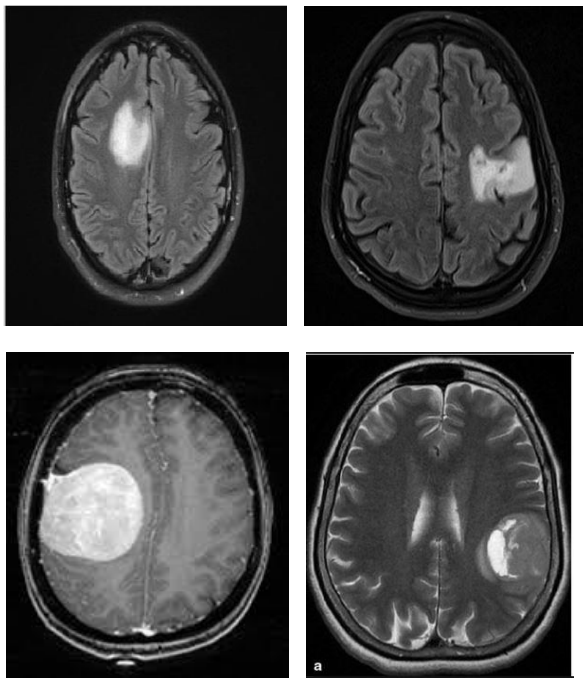
4. PREPARING DATA FOR TRAINING

respectively. The Images were all of different size , so firstly we resized all the images to a fixed size of (228,228,3) .

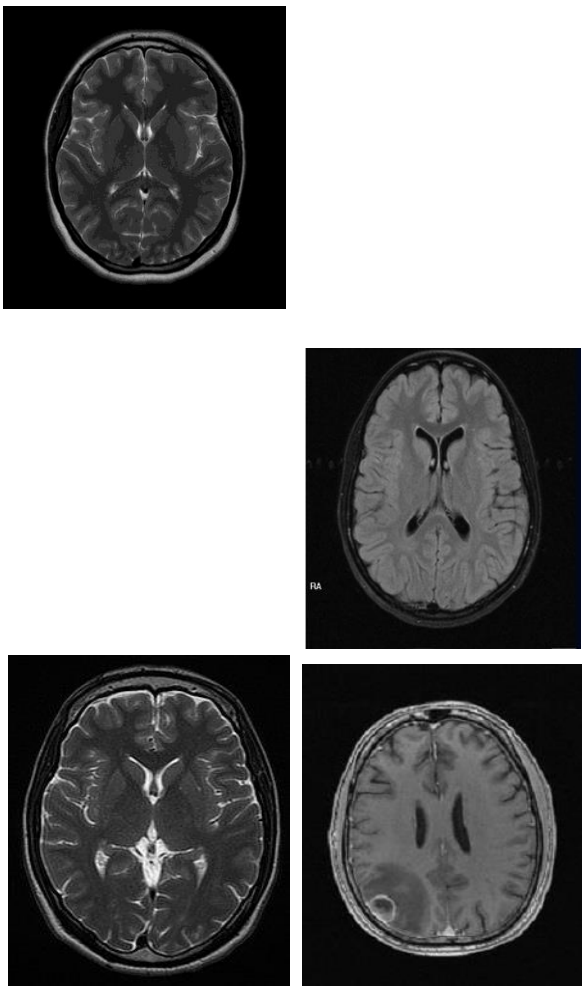
Now we have 2 folders for 10x and 20x images having yes and no sub-folders in them , which contained images respectively to there names.

Now its time for training and testing.

Some samples of positive images are :



Some samples of negative images are :

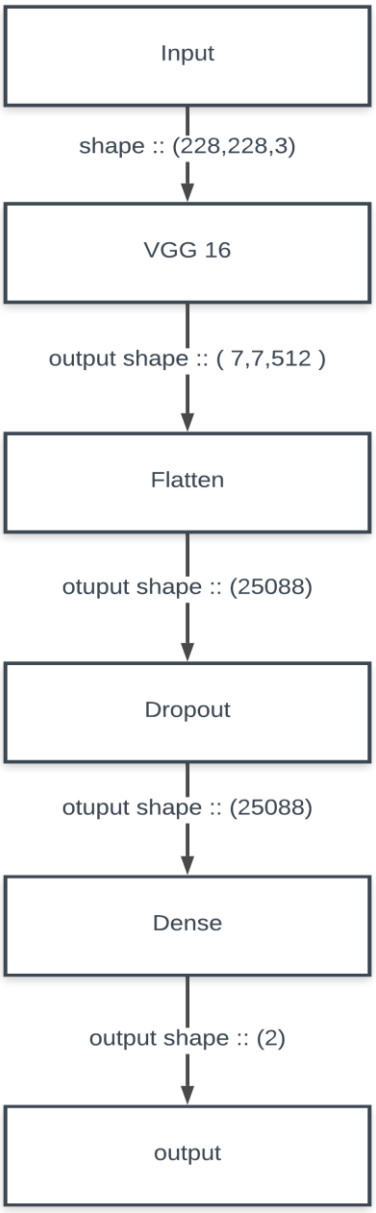


III. Model

In our model , we have used a VGG16 model ,easily accesible from keras framework to work as base layer for our model. VGG-16 is convolutional neural network . It has shown some good result on various large image dataset availabe. Along with VGG-16 to be as our base layer , we flatten the output from this layer and passes them through dropout layer to drop some unuseful data and then finall through a Dense layer with sigmoid activation to get final output.

The final output consists of 2 values , indicating probability to which the input belongs to.

Below figure shows the architecture of model used in this paper.



IV. RESULTS

F1 Score was used to compare the model.

The **F**1 **score** is the harmonic average of the precision and recall, where an **F**1 **score** reaches its best value at 1 (perfect precision and recall) and worst at 0.

The following table shows the results achieved by this network trained with different augmentation but with same number of iterations as 15.

augmentation	*f1-score*	*accuracy*	*loss*
10x	0.83	0.9486	0.9658
20x	0.84	0.9506	0.7468

REFERENCES

1. https://www.kaggle.com/navoneel/brain-mri-images-for-brain-tumor-detection
2. www.google.com
3. https://keras.io/preprocessing/image/
4. www.lucidchart.com