ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4

Journal of Coastal Life Medicine

Captioning and Classification of Brain Tumor from MRI Images using Deep Learning Methods

Received: 24 February 2023, **Revised:** 20 March 2023, **Accepted:** 16 April 2023

Yashaswini S,
Assistant Professor, Cambridge Institute of Technology

Jayanthi M. G.
Associate Professor, Cambridge Institute of Technology

Jenitha Subhash
Assistant Professor, Cambridge Institute of Technology

Keywords
CNN, LSTM, Base64 Decoder

Abstract

Captioning is a fundamental task in deep learning. It refers to the process of generating textual description from given image based on its features. Deep learning method handles images using CNN and LSTM. CNN understands image contents. There are several promising methods like RCNN, LSTM, and Base64 Decoder etc. In this paper, different approaches to image captioning for medical datasets are discussed.

1. Introduction

Comprehending medical images and generating suitable description for brain MRI image is at nascent stage among image captioning applications. Generating the semantically meaningful descriptions is quite challenging in NLP. A typical Brain MRI caption pipeline comprises of an encoder like Convolutional Neural Networks (CNN) and decoder like Long Short Term Memory (LSTM).Traditional image captioning techniques cannot generate semantic captions. This research focuses on hybrid image captioning techniques that generate more meaningful and superior captions by assigning a label to every pixel in the image. Semantic segmentation identifies, classifies different parts of image, rather than assigning single label to the entire picture.

A deep learning method for classification of Medical MRI images is presented on the datasets of Brain Tumor Classification (MRI images). The earlier research concentrates on extracting robust features to learn the structure of MR images via convolutional and the fully connected layers are replaced with trained classifier to distinguish tumour and non tumor classes.

The decoder requires Pre-trained discriminator of a GAN to employs data augmentations and dropout to prevent overtraining of dataset. Onehotencoder is used to train the model. This method is applied to an MRI dataset consisting of MR images. The MRI images are segmented and color coded for easy recognition of tumor. The Brain MRI images are segmented into different regions like scalp, skull, Cerebral Spinal Fluid(CSF),Gray matter and white matter. The detailed understanding of brain structure helps in identification of different brain tumour types like meningioma, glioma, and pituitary tumour. cross-validation is best suited method to evaluate the performance as compared to state-of-art methods.

Semantic Segmentation identifies salient elements in medical scans especially abnormalities such as tumors. Tumor identification applications require high accuracy and low recall. There is a scope to automate less critical operations such as estimating the volume of organs from 3D semantic segmented scans. However Manual automation can provide good results.

The Section 2 introduces to the recent methods for Image Captioning. In Section 3 highlights architecture to assesses the performance of the proposed model.

Section 4 discusses different methodologies available to implement attention mechanisms. Section 5 shows experimental setup, results on different ideas, software and hardware platforms needed for implementing the model, which is followed by a conclusion in Section 6.

2. Methodology

The image embeddings are obtained by pre-trained InceptionV3. These embeddings generates caption using RNN. The system is trained by set of tumorous and non-tumorous images that were segmented manually by humans using CVAT tool, and system is expected to reuses human captions for test images from dataset .

2.1 Encoders & Decoders : -

CNN and RNN are predominant encoder-decoder models that work on two-dimensional image data. The encoder framework consists of Inception functions for text recognition tasks and the decoded vector generates pattern through the LSTM (Long Short Term Memory) and GRU (Gated Recurrent Unit) used for speech recognition and NLP tasks. The Generalized architecture of the Encoder-Decoder architecture is as shown in fig.2:

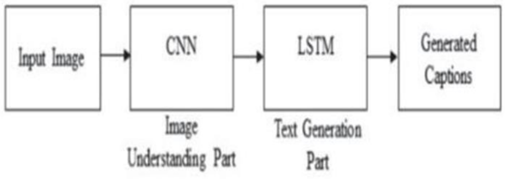


Figure 2. Encoder Decoder Generalized Architecture

2.2 Image Captioning

There are multiple sources of images without description text however humans are capable enough to understand but it is hard for machines to comprehend, thus requires text to understand.

Image captioning methods must identify objects, action, and relationships between the objects. The generated sentence must be semantically coherent. A good understanding of an image obtains features of a given image by employing either Conventional Machine Learning-Based or Deep- Learning Based Methods.

2.3 LSTM : -

LSTM architecture comprises of three cell states , namely input, output and forget gateway. The decoded vector obtained from LSTM is combined with CNN image encoder to unroll connections between the LSTM memory, consisting of an internal gates that regulates the information flow. The gates decide whether data has to be saved or discarded. The GRU is similar to LSTM, but does not contain cell state and transmits information using the secret state.

2.4 Generative Adversarial Networks(GAN): -

GANs give a new vision in the world of the neural network. It consists of the Generator and the Discriminator. The Generator's generates data with some noise derived from random distribution that closely resembles the original distributed data. The Discriminator model identifies whether the generated output is real or fake. In the image caption, the sentences generated must be close to human consensus. The GANs truth value used in discriminator helps to classify the result thus implying the quality of the output.The system Architecture is given below in the Fig.3.

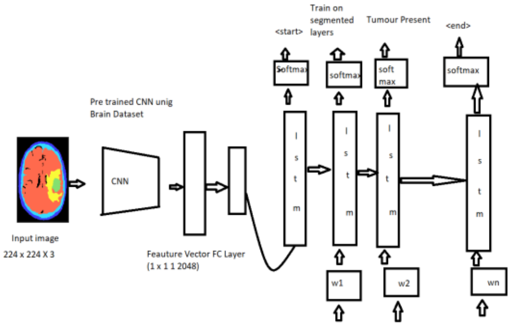


Figure 3. System architecture of proposed method

Methodologies

2.5 Dataset

The data for CNN model is adopted by Kaggle consisting around 3K brain MRI images, separated in 4 categories namely glioma tumors, meningioma tumors, pituitary tumors and no tumors. The dataset employ k-fold validation consisting of 2880 as training and 384 testing images. The Simple model consists of only two convolution layers and the

ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4
Journal of Coastal Life Medicine

Softmax layer trained for 30 epochs has achieved an accuracy of about 81%. The output returns probability of each MRI image belongs to any 4 classes.

2.6 Convolutional Neural Network

The convolutional neural layer is followed by ReLU activation function, to decreases the training time. The cross-channel normalization scales and adjusts the activations. The max Pool layer is down sampled to achieve spatial invariance by splitting the entire image into 2× 2 small rectangles.), softmax layer and classification layer. The Fully Connected layer (FC) is used to connect every neuron in a layer to every neuron in the following or preceding layer unit. Then, the FC layer is followed by a softmax layer to squash all the predicted classes between 0 and 1, and the total sum of these values is equal to 1 (100%).The output of this layer can be calculated as follows. The output of this layer can be calculated as follows

f(x) = max(0, x) …………(1)

y(z)j = ezjPk where k=1 ……………….(2)

H(p, q) = −X x (p(x) ∗ log(q(x)))…………(3)

The probability of each class can be calculated over k different classes as a function y (z). The classification layer based on cross-entropy loss estimates the classification loss and provides the final predicted categorical label for each input image as denoted by equation (3), where p is the target labels vector, and q (x) is the output vector from the softmax layer as shown in fig 4.

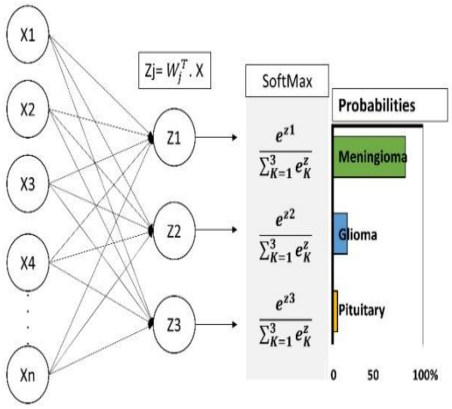

X1
X2
X3
X4
Xn
$Z_j = W_j^T . X$
Z1
Z2
Z3
SoftMax
$\frac{e^{z1}}{\sum_{K=1}^{3} e_K^z}$
$\frac{e^{z2}}{\sum_{K=1}^{3} e_K^z}$
$\frac{e^{z3}}{\sum_{K=1}^{3} e_K^z}$
Probabilities
Meningioma
Glioma
Pituitary
0
50
100%

Figure 4. Adjustment of weights in Soft-max layer

2.7 One-Hot Encoder

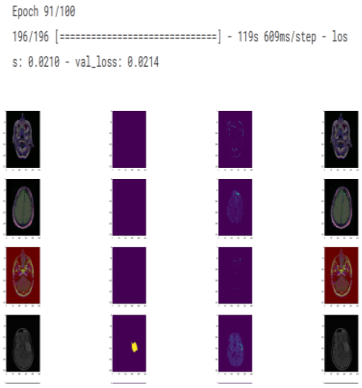
Label encoding has ordering trouble where numeric values may get misinterpreted hence it is addressed using ‘One-Hot Encoding’. This strategy transforms cost into a new column and assigns either 1 or 0. This approach is more flexible because it allows encoding as many category columns and chooses label to the columns using a prefix.

2.8 Base64 Decoder

Base64 encoding converts bytes in binary or text data to ASCII characters. Decoding Base64 string is exactly opposite to that of encoding. Each character in the string is changed to its Base64 decimal value. The decimal values obtained are converted into their binary equivalents by truncating first two bits of the binary numbers and the sets of 6 bits are combined to form one large string of binary digits. The obtained large string of binary digits is further divided into groups of 8 bits.

3. Result:

A Brain tumor is one of the aggressive diseases, among children and adults. The commonly found Brain tumors account for Central Nervous System(CNS) tumors. This methodology attempts to identify and classify Brain Tumors as Benign Tumor, Malignant Tumor, and Pituitary Tumor by applying Deep learning techniques using the semantically segmented brain MRI Images as input as shown in fig 5. A manual examination can be error-prone due to the level of complexities involved in brain tumors and their properties.



```
Epoch 91/100
196/196 [==============================] - 119s 609ms/step - los
s: 0.0210 - val_loss: 0.0214
```

Figure 5. results obtained for semantically segmented Brain MRI Images

ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4

Journal of Coastal Life Medicine

The results shows the accuracy of 40 % and loss of 25% as the approach directly uses the Brain MRI Images as shown in fig 6.

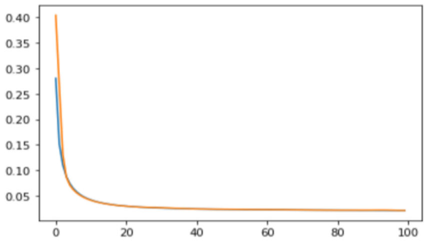


Figure 6. The Accuracy and loss of brain MRI images

The results of UNet approach shows an evident gain in accuracy and decrease in loss and predominantly able to detect the tumor by considering input image of 240x240 were divided into 144 patches of size 20 x 20.The training and validation accuracy is as shown in fig.7.

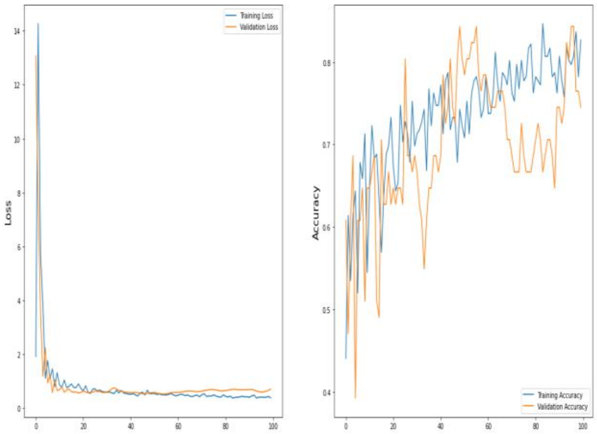
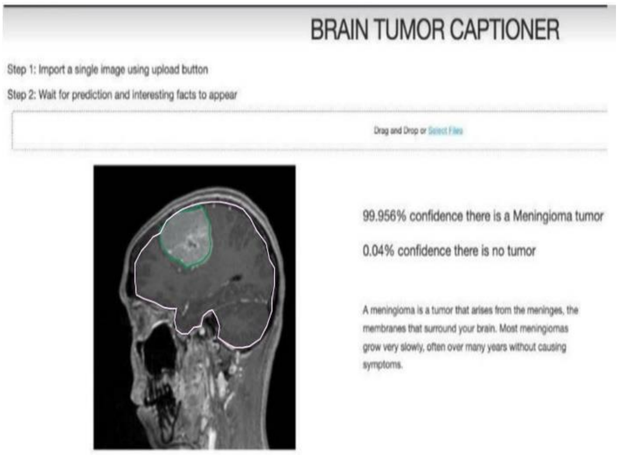


Figure 7. The Training stage, validation stage loss and accuracy

The testing phase includes using an un trained Brain MRI images that has to be trained and then it has to be processed further to identify the type of brain tumor and also it has to provide the caption and more information regarding the tumor location and type as shown in the fig.8 it classifies it has meningioma cancer with the probability of 99.9 the segmented output clearly shows the location of tumor and it is colored as green. The Fig.9 shown the presence of pituitary tumor and the fig.10 shows the presence of glioma cancer with the probability of 64.8 and fig 11 denoted the no tumor cell. Our approach not only identifies the exact location of tumor through segmentation but also specifies the probability of classification and also provides more information regarding the type of cancer as shown in the screenshots below.



BRAIN TUMOR CAPTIONER

Step 1: Import a single image using upload button

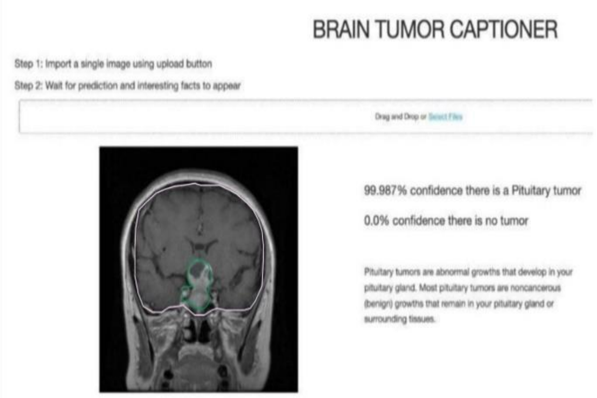
Step 2: Wait for prediction and interesting facts to appear

99.956% confidence there is a Meningioma tumor

0.04% confidence there is no tumor

A meningioma is a tumor that arises from the meninges, the membranes that surround your brain. Most meningiomas grow very slowly, often over many years without causing symptoms.

Figure 8. classification of tumor as meningioma with captions and probability



BRAIN TUMOR CAPTIONER

Step 1: Import a single image using upload button

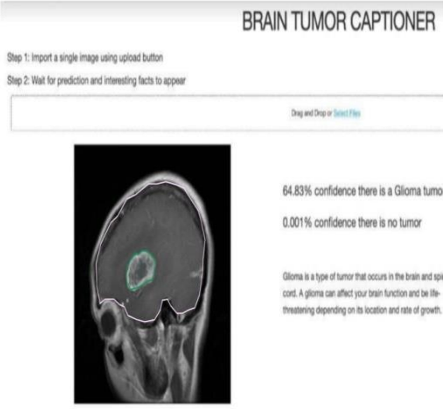
Step 2: Wait for prediction and interesting facts to appear

99.987% confidence there is a Pituitary tumor

0.0% confidence there is no tumor

Pituitary tumors are abnormal growths that develop in your pituitary gland. Most pituitary tumors are noncancerous (benign) growths that remain in your pituitary gland or surrounding tissues.

Figure 9. classification of tumor as pituitary with captions and probability



BRAIN TUMOR CAPTIONER

Step 1: Import a single image using upload button

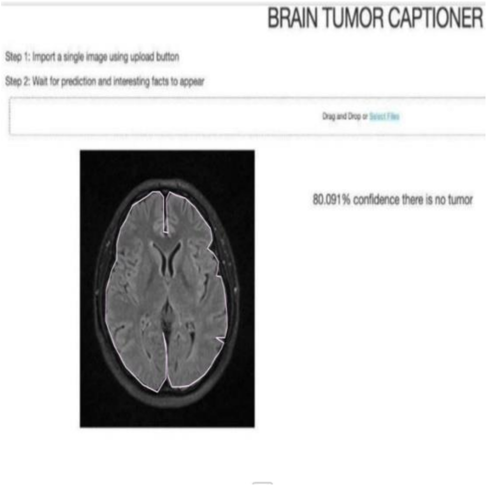
Step 2: Wait for prediction and interesting facts to appear

64.83% confidence there is a Glioma tumo

0.001% confidence there is no tumor

Glioma is a type of tumor that occurs in the brain and spi cord. A glioma can affect your brain function and be life-threatening depending on its location and rate of growth.

Figure 10. classification of tumor as glioma with captions and probability



BRAIN TUMOR CAPTIONER

Step 1: Import a single image using upload button

Step 2: Wait for prediction and interesting facts to appear

Drag and Drop or Select Files

80.091% confidence there is no tumor

Figure .11. No tumor image with probability

4. Discussion

A. Visual Relationship Detection

Annotation can attain greater heights with deep learning. Image captioning applications such as Image indexing, editing applications, social media are in trend however its scope could be extended for identification of abnormalities in medical scans.

B. Image Caption Generation Using Multi-Level Semantic Context Information

The traditional captioning method only converts the extracted image features to text description. However In recent years, image captioning is moved towards comprehensive deep image-captioning. Image captioning methods detects objects and the relationship, but it must also determine the context information between objects in scene environment.

5. Conclusion:

The image captioning generation using deep learning techniques approach generates the accurate description text of the visual image scene using the CNN, RNN, and GAN model with LSTM. The different datasets of variety of tumor and no tumor dataset are considered for evaluating the system. We also discussed the limitations of the automatic segmentation by machine and emphasize of comparing machine generated results with manual annotation with CVAT tool. In the end, evaluation metrics like probability derived from confusion matrix helps to assess our model. The system delivered around 89% accuracy for Brain tumors classification

C. Gray Level Segmentation

The semantic segmentation assigns gray level intensity with a particular label by using the Split and Merge algorithm. The algorithm splits an image into sub-regions until there is no change in color intensity assigns a label and then merges adjacent sub-regions with the same label.

D. Fully Convolutional Network (U-Net)

FCN down sample the input image through a series of convolutions (encoder),then up samples either through bilinear interpolation or a series of transpose-convolutions(decoder). The U-Net is an upgraded version of FCN architecture that skip connections from the output of convolution blocks to the corresponding input of the transposed-convolution block at the same level as explained in fig 1.

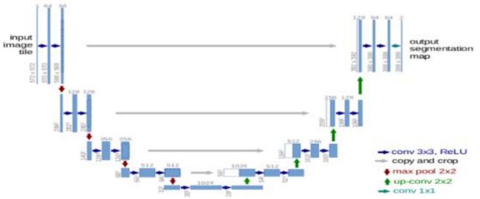


Figure 1 The UNet Architecture for Brain MRI Images

This skip connection consists of gradients that flow and provides better information regarding image size. Information from upper layers helps the model for better classification, whereas deeper layers can help the model to segment/localize.

on Brain MRI images available in Kaggle.

The image captioning is used for generic application like natural image scenes using models for reinforcement and unsupervised learning. Integration of textual cues with visual information will improve the image captioning task to empirically evaluate attention impacts across a range of tasks. We also observe that ResNet and visual transformers are definitely capable of encoding a better feature vector for image scenes.

Generation-based methods can be used to get novel captions for medical image. However, these methods fail to detect exact objects and attributes and their relationships. Also, there is no exact means to measure accuracy of the machine generated captions which in turn rely on a powerful and sophisticated language generation model. Existing methods show their performances on the datasets where images are collected from the same domain. Therefore, working on open domain dataset will be an interesting avenue for research in this area.

References

[1] Omkar Sargar, Shakti Kinger.: Image Captioning Methods and Metrics 2021 International Conference on Emerging Smart Computing and Informatics (ESCI) AISSMS Institute of Information Technology, Pune,India. Mar 5-7, 2021

[2] Vikram Mullachery, Vishal Motwani.: Image Captioning arXiv:1805.09137v1 [cs.CV] 13 May 2018.

[3] MD. ZAKIR HOSSAIN, FERDOUS SOHEL, MOHD FAIRUZ SHIRATUDDIN, HAMID LAGA.: A Comprehensive Survey of Deep Learning for Image Captioning. arXiv:1810.04020v2 [cs.CV] 1Tarun Wadhwa1

[4] Tarun Wadhwa, Harleen Virk. Jagannath AgBoro, Savita Borol.: Image Captioning using Deep Learning.

[5] Murk Chohan, Adil Khan, Muhammad Saleem Mahar Saif Hassan, Abdul Ghafoor, Mehmood Khan.: Image Captioning using Deep Learning: A Systematic Literature Review. Vol. 11, No. 5, 2020

[6] Peng Tian, Hongwei Mo * and Laihao Jiang.: Image Caption Generation Using Multi-Level Semantic Context Information. Symmetry 2021, 13, 1184. https://doi.org/10.3390/sym13071184.

[7] Himanshu Sharma,Manmohan Agrahari, Sujeet Kumar Singh, Mohd Firoj, Ravi Kumar Mishra: Image Captioning: A Comprehensive Survey. 2020 International Conference on Power Electronics & IoT Applications in Renewable Energy and its Control (PARC) GLA University, Mathura, UP, India. Feb 28-29, 2022