Comprehensive Analysis based on One Hot Encoder for Brain Tumor Detection and Classification

Chinmay Daund
Department of Information Technology
Ramrao Adik Institute of Technology
Nerul, Navi Mumbai
chinmay.daund@gmail.com

Siddharath Bohra
Department of Information Technology
Ramrao Adik Institute of Technology
Nerul, Navi Mumbai
siddharth.bohra09@gmail.com

Neel Nimbalkar
Department of Information Technology
Ramrao Adik Institute of Technology
Nerul, Navi Mumbai
neel.nimbalkar@gmail.com

Sujata Oak
Department of Information Technology
Ramrao Adik Institute of Technology
Nerul, Navi Mumbai
sujata.oak@rait.ac.in

Abstract—**A tumor on the brain can be divided into two categories: benign and malignant. Prompt diagnosis and nurs- ing arrangement ushers to refined standard of life and ex- panded life span of these convalescents. One of the majorly effective and important course of action is to avail the Deep Neural Network (DNN). In this case the Convolutional Neural Network (CNN) was utilized to ascertain a tumor with images of the Magnetic Resonance Imaging (MRI) brain. Along with the issue of precision, we will be using the susceptibility, particularity and precision standards to as- sess network conduction. Because of the significance ofthe prognosis given by the physician, the precision of the phy- sician helping in diagnosing the tumor and medicating the sick skyrockets. In presence of the present methods will try to implement various algorithms to bring out maximum accuracy. We have implemented an algorithm called as One Hot Encoder which will classify the tumor into different types based on some key features and display the results accordingly with the type of tumor and accuracy of result. (Abstract)**

Keywords— brain tumor detection, DNN, CNN, deep learning

I. INTRODUCTION

Deep learning is a method of stratifying, combining, and fore- casting objects using a neural network trained with a large quantity of input. Deep learning produces numerous surfaces of neurons, trying to grasp the formal portrayal of large input in a layer wise manner. Deep Neural Network consists of three types of neural networks out of which we have used Convolutional Neural Network (CNN). It is a kind of deep neural network linearization as it takes input in size and pro- duces output with established size.

It is widely used in computer vision systems for object dis- cernment, stratification and semantic dissociation and is fea- sible for video and image clarifying.

Softwares likes Keras & Tensor Flow which is a highly pro- ductive interface for solving Deep Learning problems. Ten- sorFlow is an extensively utilized deep learning model de- rived from Github stars and forks and stack overflow func- tion. The majorly popular and generally availed visual connector for deep learning today is the Cameras. Keras is an advanced and prominent learning API, corresponded via Python.

II. LITERATURE SURVEY

Identification of the brain tumor takes a major role in the pro- posed pathway. In more recent times, the classification meth- odology obliges substantial utility of DL and ML algorithms. This segment introduces previous techniques and functions in differentiating brain tumor and ML-based isolation from MRI.

A.Sivaramakrishnan and Dr.M.Karnan (2013) brought for- ward an innovative and fruitful discovery of a brain tumor region from a brain image performed using Fuzzy C-means compilation and histogram. Histogram estimates were availed to determine the level of magnitude of the gray im- ages. Image decompression was performed using a partial analysis system used to reduce the wavelength coefficient. The outcome of the suggested Fuzzy C-means (FCM) algo- rithm, fruitfully and precisely extracted the tumor area from brain MRI images.

Jaskirat kaur et al (2012), characterized integrating images, classification algorithms and evaluated various kinds of im- age separation techniques. They also suggested a process of differentiating and measuring contrasting integration algo- rithms hinging on their compatibility in different methodol- ogy. They explain the many execution limitations on which compliance will be measured.

Roy et al (2012) determines the impacted region of a tumor for scrutiny. They demonstrated its effectiveness with several data sets of contrasting tumor measurement, magnitude and whereabouts. They have demonstrated that their algorithm can intuitively ascertain a brain tumor. MR images provide a superior outcome in contrast to different methods like CT and X-ray images. Pre-processing images involves transforming an RGB image to a grayscale image and transferring that im- age to a elevated throughput filter to separate the subsisting noise in the image.

B. Sathya et al (2011), put forward four congregating algorithms; k means, refined k means, c means and refined c means algorithm. They performed an exploratory scrutiny of a huge database that includes a smorgasbord of images and perused the outcome utilizing different limitations.

III. PROPOSED METHODOLOGY

Deep Learning concept has been used since we are having unstructured data. Frameworks like Keras and TensorFlow are used. Dataset has been taken from Kaggle. we have di- vided our dataset into two parts i.e Training image dataset and Testing image dataset and One Hot Encoder is used for Data Pre-Processing. The model will preprocess these images for better image quality. Pre-processing is required as it gives im- age input improvement which amplifies some pertinent im- age attributes to supplemental processing. using appropriate algorithm, Train the model with training image, and saving the model with weights. We have built a sequential model. It permits us to create a layer wise paradigm. Different layers like Conv2D, Activation, axPooling2D, Flatten and Dropout is created. After training our model with the training dataset, we are getting the accuracy of the predicted model. We are testing our model with the testing dataset and finally observ- ing the model with testing image to get final prediction.

IV. CNN (CONVOLUTIONAL NEURAL NETWORK)

It is an intricate paradigm that is employed to designate the tumors and is quite elaborate, therefore it requires outstand- ing analytical strength. It is also a frequently implemented neural network which is broached on specific image stratifi- cation complications. This set-up encompasses of "Seven Hidden Layers" :

1. Zero Padding
2. Max pooling
3. Convolutional 2D
4. Batch Normalization
5. Flattening Layer
6. Dense Layer
7. Relu

The Database of the tumors is generated. It predominantly comprises of meningioma, pituitary and no tumor. The im- ages of the tumor are transfigured from RGB to Gray Scale for additional activity, as it aids in achieving superior accu- racy. Post the part of image pre- processing, availing CNN a Deep Learning Algorithm training of this paradigm is com- pleted.

The libraries made use of are Matplotlib, Tensorflow, Keras and OpenCV. The dataset was allocated in training set and substantiation set. Next the training data was produced by conveying specific variables like size.

Convolutional Neural network was availed to train the model. There is also an input layer and a convolutional network (2D). Max pooling layers are utilized, accompanied by a flattening layer.
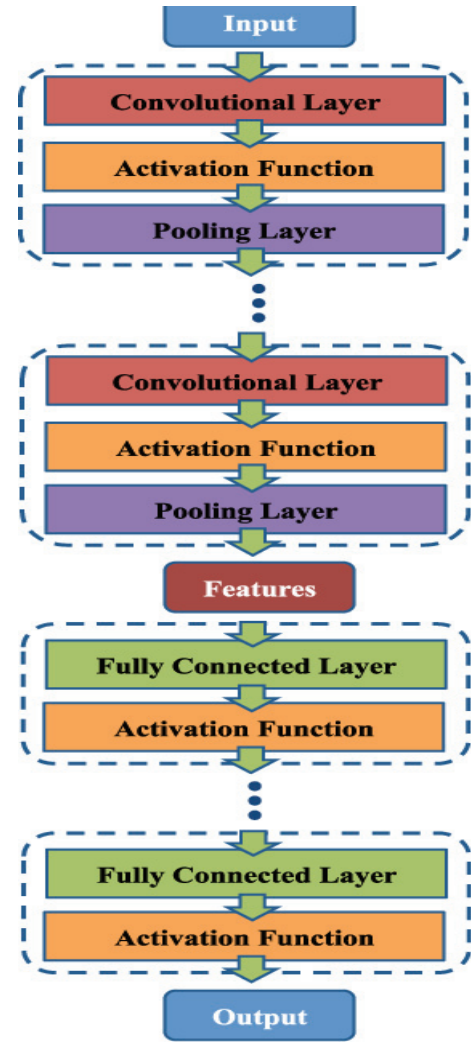

Fig. 1. System architecture

Convolutional neural networks are differentiated from additional neural networks by their more advanced execution with visuals or auditory wave data.

The convolutional layer is the topmost layer of a convolutional network. While convolutional layers can be succeeded by extra convolutional layers or pooling layers, the completely-intertwined layer is the terminal layer. With every layer, the CNN boosts in its intricacy, recognizing substantial fragments of the photo.

Previous layers emphasize on straightforward characteristics, like shades and corners. As the photo details advances through the layers of the CNN, it begins to perceive bigger components or structures of the entity till it eventually recognizes the planned entity.

Convolutional neural networks reinforce visual identification and computer vision errands. Computer vision is a scope of artificial intelligence that allows systems and devices to acquire relevant data from visuals, media and different sources, and depending on those sources, it can take relevant measures. This capacity to dispense suggestions differentiates it from image identification endeavors.
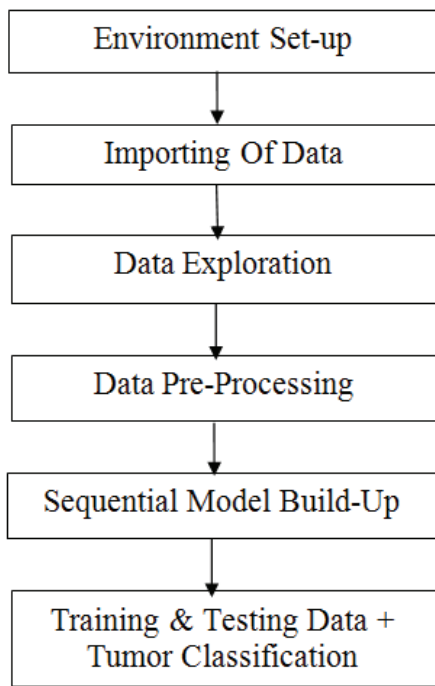
Fig. 2. Workflow of System

In the subsisting system, the RGB MR figure is transformed to a gray scale figure and further utilized as a filter to extract noise from MR brain images; Extracted noise for further pro- cessing as elevated precision is desired. Then the fringes are found on the sifted images using the canny edge identifica- tion. An image frame is required to separate the image. Then a watershed separation is performed to determine the co-or- dinates of the tumor in the brain. Separation is the action of segmenting an image into many parts. The purpose of the split is to turn the image depiction into anything that is less complicated to the result of the watershed splitting is image label.

In the image label, every other entity is recognized and have dissimilar pixel values, each initial entity pixels will contain a value of 1, each second entity pixels will contain a value of 2 and so on. Primary congregation is a form of integration. This algorithm has a repetitive undertaking, which, in turn, in a fixed number of collections, tries to assimilate points as col- lective centers, which are actually the identical points of measure for each component. Also allocate data for each specimen to a collection that provides data within a lowest interval between that collection. In the simplest form of this procedure, foremost the collection centers are arbitrarily cho- sen. Points are awarded to collective centers in accordance to the level of resemblance, and thus new collections are ob- tained.

The existing systems have reported errors and fallacies in their results where they have mistakenly identified and re- ported. In some cases, some fatty areas in the photos are ac- cidently diagnosed as a tumor, or the tumor may not be diag- nosed by a medical expert; The most precise diagnosis de- pends entirely on the expertise of the physician.

Environment Setup

The Environment Setup's established by importing necessary packages like matplotlib, shuffle, listdir, time, cv2, numpy.

Loading of Input

We have imported dataset from Kaggle. Importing of data and data exploration is done in this step.

Preprocessing Split the data into two parts

In this step, we have divided our dataset into two parts i.e Training image dataset and Testing image dataset and One Hot Encoder is used for Data Pre-Processing.

Building a Deep learning Model

In this step, using appropriate algorithm we train the model with training image, and saving the model with weights. We have built a sequential model. It permits us to develop a layer wise paradigm.

Model Evaluation

In this step, we are getting Accuracy for the predicted model. Finally observing the model with testing image to get final prediction.

Environment Setup has been done by importing necessary packages required to build our model

Listdir- Used for getting a list of all the files and references in the directories

time- This function returns the number of seconds passed since epoch

cv2- Loads an image from the specified file

Tensorflow- TensorFlow is a unconfined software library and an open source for information and segmented system across multiple functions.

shuffle- The shuffle() method takes a sequence (list, string, or tuple and data) and reorganize the order of the items.

Numpy - A Python programming language library, which ap- pends sustenance for huge, various-sided matrices and arrays, as well as a huge assembly of advanced mathematical options to work on these components.

matplotlib – is a library which builds an plotting region quantity, edits certain lines in the area, embellishes the area with labels.

Sequential Model

Sequential is the easiest way to build a model in Keras. Se- quential paradigms of machine learning models add or sub- tract data sequences. It allows us to create a layer wise para- digm. We are utilizing the 'add ()' function to input layers to our paradigm. Our foremost layer is Conv2D layers. These are the layers of convolution that interact with the inserted images, which are being considered as 2-dimensional matri- ces. Activating the activation layer function. Using the acti- vation function which we will use for our first 2 layers is ReLU, or Rectified Linear Activation. This activation func- tion's

demonstrated to perform efficiently on neural net- works. Between the Conv2D surfaces and the dense surface, there exists a 'Flatten' layer. Flatten acts as a link connecting convolution and dense layers. A hidden layers is nothing but number of hidden patterns associated between input(x) and output(y).

Zero Padding=> symmetrically adding zeroes to the input matrix

Convolutional 2D=> Calculates dot product of pixel value **Batch Normalization**=> Improves performance, stability of CNN

Maxpool => A Sample based discretization process for input data.

Flattern => Pooled features mapped into single column **Dense**=> To connect the pooled features into connected layer.

Dataset

The dataset comprises 2 class kinds: yes and no comprising 253 Brain MRI images. The yes file contains 155 Brain MRI images which are a kind of tumor. The no file contains 98 Brain MRI images that are not tumors.

Loading of Input-Data importing, Exploration, Preprocessing We have loaded the images by taking in necessary parameters like format, type and size of the images and put all the images into a directory which will further be exported into the model as our dataset. We have converted the format of the image and grey scaled it so that further it will be easier for the model to predict the tumor. The original image is cropped. Any small noise present in the image is tried to reduce which will clearly help to focus on the tumor.

Working of Model

All the necessary modules required for the detection are imported. We have used One Hot Encoder which is used for Data Pre-Processing. It is a customary way to preprocess the features of the categories of deep learning models. This type of coding creates a new binary feature for each probable category and allots 1 value to every test feature relevant to its foremost category. So it will classify the tumor based on its type (meningioma, glioma, pituitary and no tumor). We have built a sequential model. It permits us to create a layer wise model. So accordingly, we will get the accuracy and type of the tumor a patient has.

One hot encoding is described as the important procedure of transforming the absolute feature variables to be supplied to machine and deep learning algorithms which further enhance forecast along with stratification precision of a paradigm. One Hot Encoding is a regular method of pre-processing absolute characteristics for machine learning paradigms. This kind of encoding generates a pristine binary characteristic for every attainable criteria and allots an estimate of 1 to the characteristic of every sample that correlates to its initial criteria.

One hot encoding is a hugely important block of the characteristic constructing way in training for learning methodologies. To take an example, we have our features like

colours and the headers are "red," "green," and "blue," we can encipher every one of these headers as a three-part binary vector. The absolute data whilst clarifying, has to be transformed into a binary configuration.

V. EXPERIMENTAL RESULTS

```
[ ] img = Image.open('/content/Brain Tumor Data sets/Testing/no_tumor/image(1).jpg')
    x = np.array(img.resize(dim))
    x = x.reshape(1,150,150,3)
    answ = model.predict_on_batch(x)
    classification = np.where(answ == np.amax(answ))[1][0]
    imshow(img)
    print(str(answ[0][classification]*100) + '% Confidence This Is ' + names(classification))

    99.9996900558717% Confidence This Is no_tumor
```
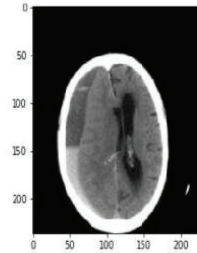

Fig. 3. An image after importing gave us 99.99% confidence that the brain has no tumor.

```
[ ] img = Image.open('/content/Brain Tumor Data sets/Testing/meningioma_tumor/image(10).jpg')
    x = np.array(img.resize(dim))
    x = x.reshape(1,150,150,3)
    answ = model.predict_on_batch(x)
    classification = np.where(answ == np.amax(answ))[1][0]
    imshow(img)
    print(str(answ[0][classification]*100) + '% Confidence This Is ' + names(classification))

    99.741530418396% Confidence This Is meningioma_tumor
```
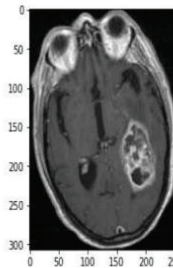

Fig. 4. An image after importing in our model gave us 99.74% Confidence that the brain has a tumor and it is meningioma tumor.

```
[1] img = Image.open('/content/Brain Tumor Data sets/Testing/pituitary_tumor/image(13).jpg')
    x = np.array(img.resize(dim))
    x = x.reshape(1,150,150,3)
    answ = model.predict_on_batch(x)
    classification = np.where(answ == np.amax(answ))[1][0]
    imshow(img)
    print(str(answ[0][classification]*100) + '% Confidence This Is ' + names(classification))

    100.0% Confidence This Is pituitary_tumor
```
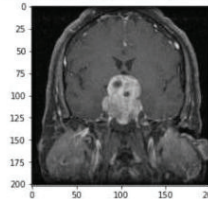

Fig. 5. An image after importing in our model gave us 100% Confidence that the brain has a tumor and it is pituitary tumor.

VI.   CONCLUSION

In this venture, a latest procedure is established by the amalgamation of One Hot Encoder algorithm and CNN to ascertain tumors in brain images. CNN can detect the tumor. CNN is very helpful in selecting a default feature in diagnostic imaging.

The images collected at the centers were written by doctors, and the tumor examination was divided into two general classes and one patient. Because of the im- portance of the diagnosis given by the physician, the ac- curacy of the physician assisted in diagnosing the tumor and treating the patient increases the medical accuracy of the proposed method.

VII.   FUTURE SCOPE

As the segregation of medical imaging plays an ex- tremely dominating role in the field of imaging-based surgery. By building three-dimensional (3D) corporeal paradigms for every convalescent, priming, drafting, and machine-assisted directives are developed. This proce- dure can be applied to the ovaries, breast, lungs and skin cells. In ensuing work, in order to refine the precision of the current task division, we plan to investigate the se- lected classification system by merging with additional classifiers with steps for choosing features.

ACKNOWLEDGMENT

REFERENCES

[1] M. Soltaninejad, et al, "Automated brain tumor de- tection and segmentation using superpixel-based extremely randomized trees in FLAIR MRI", International journal of computer assisted radiology and surgery, 12(2), pp. 183-203, 2017.

[2] R. Bayot and T. Gonalves, "A survey on object clas- sification using convolutional neural networks", 2015.

[3] A. joshi, D. H. Shah, et al, "Survey of brain tumor detection techniques through MRI images", International Journal of Research in Engineering and Technology, 10, pp. 586- 594, 2013.

[4] JP. Poonam, "Review of image processing tech- niques for automatic detection of tumor in human brain", In- ternational Journal of Computer Science and Mobile Compu- ting, 2(11), pp. 117-122, 2013.

[5] M. Karuna and A. Joshi, "Automatic detection and severity analysis of brain tumors using GUI in matlab", Inter- national Journal of Research in Engineering and Technology, 10, pp. 586-594, 2013.

[6] H. Cecotti and A. Graeser, "Convolutional neural network with embedded Fourier transform for EEG classifi- cation, Pattern Recognition",19th International Conference on. IEEE, ICPR, pp. 1-14, 2008.

[7] KS. Aboody, A. Brown, et al, "Neural stem cells dis- play extensive tropism for pathology in adult brain Evidence from intracranial gliomas", Proceedings of the National Academy of Sciences, 97 (23), pp. 12846- 12851, 2000.