

Brain Tumor Detection: Model and Analysis

Abhiral Dubey¹, Atharva Pangerkar², Tejas Pawar³, Santhosh Pillai⁴, Prof.R.S. Bhoyar⁵
^{1,2,3,4,5}*Rajiv Gandhi Institute of Technology, Mumbai*

Abstract - There is a high need of machine learning techniques in the medical field. Many medical applications work on the deep learning and machine learning algorithms. In this project we will going to detect the tumor of an MRI image. Detected output region will provide the information about the patient is having a tumor or not. Currently and before the standard process of detecting the tumor is based on the basic doctor check-up or examination, but this process has many drawbacks as it is not applicable in the crucial time when there are lots of MRI reports. It will be quite hectic for the medical staff. Therefore, we have come up with a solution with this project. The project will be based on the two steps first the detection of the MRI image. With the help of this techniques, we will be able to locate the tumor very effectively. The second step is of classification. Tumor classification purpose we will going to use three classifiers in order to make our system and program very efficient, at the output of this classifiers there won't be any doubt regarding the classification of the MRI image. Here the detection will be going to detect the tumor with various image pre-processing techniques. And for the classification we are using SVM, logistic regression, CNN classifiers in order to classify the MRI images. The principal focus will be to learn program using three classifiers and identifying which one is most efficient one.

Index Terms - Image detection, Image classification, Machine learning, Image pre-processing techniques, Classifiers (SVM, LG, CNN), Python, Modules or Libraries.

I. INTRODUCTION

The growth of the artificial intelligence and machine learning has been increasing rapidly in the current time. Manual processes of handling different techniques are getting replaced by the machines but still there are few sectors where the manual process of working is happening. Our project is some of the evolved step towards the research of automatic tumor detection. The abnormality in the cells causes tumor. The main aim of our project is to detect and classify the brain tumor depending on the provided MRI

images data set and getting high accuracy out of the classifiers. Another need or the working of the project is it will be going to give the accuracy score depending on all the three classifiers so that we will be able to identify which of the respective classifier can give the efficient output. Now for the image classification purpose, deep learning, and machine learning .Several classifiers are available such as support vector machine, random forest, k nearest neighbor, convolution neural network etc. Out of these classifiers the project will work on three specific classifiers that is support vision machine, logistic regression, and convolution neural network. In the brain tumor detection section will be going to involve many of the image processing techniques. The classifiers will be going to work specifically for the training and the testing modes. The output of all the classifiers and depending on their accuracy the program will tell us out of the three classifiers which of the classifiers will give the appropriate and highly efficient accuracy result.

II. LITERATURE SURVEY

[1] Research on image classification model based on deep convolution neural network by Mingyuan Xin , Yong Wang published that Deep convolution neural networks are apply to spot scaling, translation, and other forms of distortion-invariant images. And to keep away the feature extraction, feature detection layer is served by the convolutional network to learn from training data.

[2] Application of Logistic regression in risk management by Arun Velu tells the purpose of this study was to develop a predictive model that will help create awareness on how to handle customers and reduce customer churn. When developing a predictive model charts a lot of care should be taken to produce the right result.

[3] Brain Tumor Detection Using Anisotropic Filtering, SVM Classifier and Morphological

Operation from MRI Images published by M. O. Rashid, M. A. Mamun, M. A. Hossain works on to detect the tumor noise removal using Anisotropic filtering, segmentation using SVM and morphological operations are performed.

[4] An article written by Sobhangi Sarkar , Avinash Kumar , Sabyasachi Chakraborty , Satyabrata Aich , Jong-Seong Sim & Hee-Cheol Kim proposed the work related to the CNN based Approach for the Detection of Brain Tumor Using MRI Scans, they works on The implementing a deep learning architecture by leveraging 2D convolutional neural networks for the classification of the different types of brain tumor from MR image slices.

[5] Sushma L. K.P. Lakshmi had issued her work on Analysing of Convolution Neural Network (CNN) for Image Classification by applying Different Models which proposed a deep learning CNN rely on Keras and TensorFlow and it is deployed python for image classification.

III. BACKGROUND

In the year Nov, 1984, the first ever brain tumor was detected and operated. Since from the very beginning onwards the brain tumor is detected with the help of magnetic resonance imaging that is MRI. This MRI image reveal that there is tumor in the brain or not. The MRI use magnetic field instead of x rays to produce an image which will reveal the complete structure of the detected object. Earlier doctors used to study more and more about MRI in order to identify and make a judgement on whether the MRI image is having a tumor or not and it was a terrified job. The detection of the tumor at the very initial stage is quite difficult in order to classify the detected portion into to healthy or the infected portion. But later on, some of the highly technology got establish such as Machine learning, Deep learning, Artificial intelligence, and Computer-Aided Diagnosis (CAD) in medical science. Now because of these techniques the MRI image can not only be detected to specific region without any more consideration but also, we can classify the detected tumor into the respective tumor type. But still sometimes there is no hundred percent accuracy in it and the system might fail to detect and classify the table into proper and real section. It is needed to improve the technology and this project will going to help in some way as the chances of error is less as every part is used and verify at regular intervals. This

new technology has helped the medical science fraternity immensely. The process will be going to reduce the time as well as increases the chances of getting diagnosed at a very early stage. And because of all this points the overall view cost off the system should not go higher.

IV. OBJECTIVE

In order to have proper image classification the image size and dimensions are expected to be same for tumor detection is by using various image processing techniques. The classifiers will work efficient on the basis of the dataset we will provide as we will make sure that all the main elements of the classifiers are there in the program or not. Correctly classifying the different types of tumors. On the basis of the execution time, training score and testing score we will check the efficiency of classifiers that which one is working efficiently. Also, the image pre-processing techniques will help to detect the tumor for the classification purpose and the validation graphs will allow us the understand the working of respective classifier. Creating this project will help as it is using various techniques, which will try to detect the tumor in early stage and increase survival rate of patient. It can automatically learn to perform any task just by going through the training data i.e., there no need for prior knowledge. The overall purpose of this project is to get the maximum amount of accuracy out of these classifiers used or which of the classifier will give the maximum accuracy in the testing mode respectively.

V. BLOCK DIAGRAM

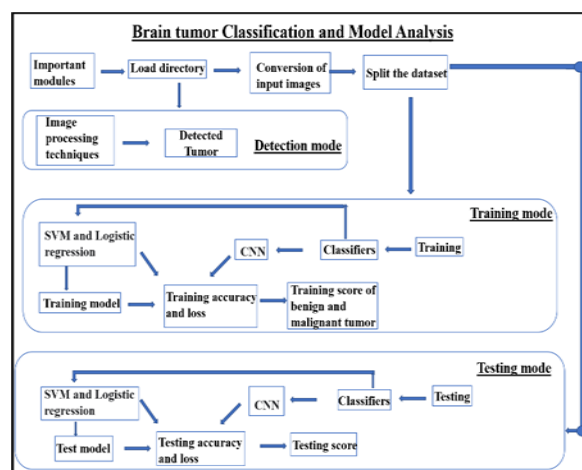


Fig I: Block Diagram

In this block diagram there are three sections: -

1. Detection mode.
2. Training mode.
3. Testing mode, respectively.

The starting few blocks will going to import the models and directory of the database. After providing the directory the data will be get split in two types of tumors. Detection module has several image preprocessing techniques which will help in the detection of the tumor. The training mode will have three classifiers: -

1. Support Vector Machine (SVM).
 2. Logistic Regression (LG).
 3. Convolution Neural Networks (CNN),
- in order to train the data set. Here the training model will get constructed. Similarly, the final section is of testing mode where the actual testing will happen on the basis of the training mode. The testing mode accuracy score will state that which of the above classifiers give good accuracy on the provided dataset

VI. FLOWCHART

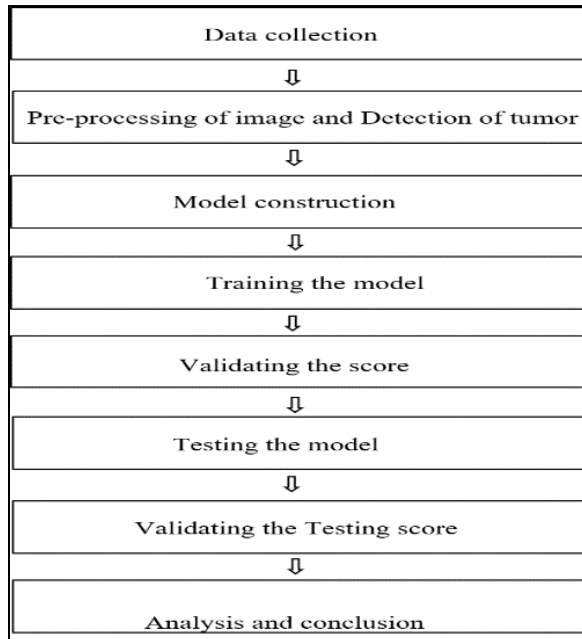


Fig II: Flow Chart

The flow of the project will begin from the tumor detection process. After the detection of the tumor the program will move to the classification part where the model is been constructed for the training of the data set. That training part will train the data set on the basis of three classifiers. After the training is done

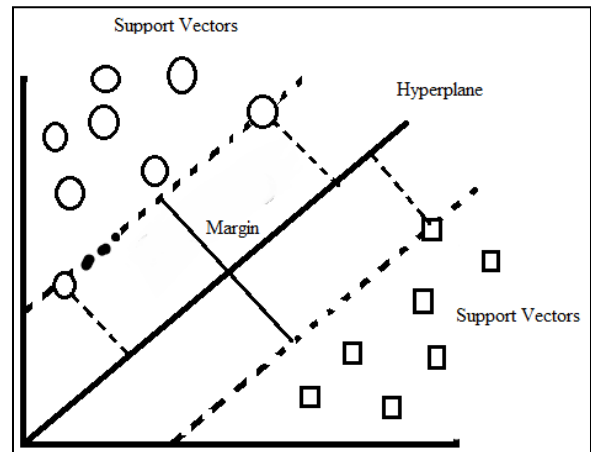
validation process will happen where the accuracy and the loss of the training will be notice. On the basis of the training and testing of the data-set will happened , ultimately the project and program will able to learn itself and give the actual accuracy score.

VII. SUPPORT VECTOR MACHINE (SVM)

Support vector machine is one of the classifier which we are using in the project. The reason behind using SVM is used to: -

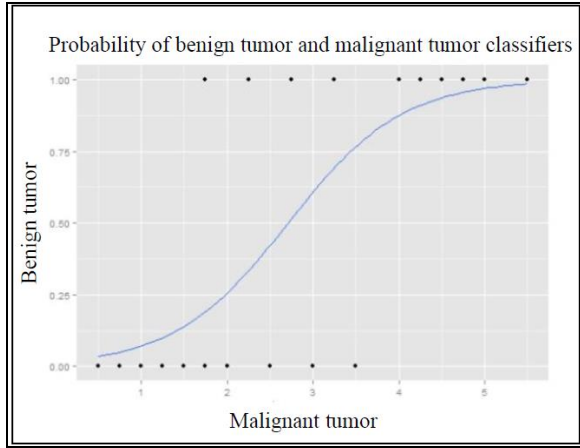
1. Solve classification and
2. Regression problem,

as it creates hyperplane between the positive and the negative points of the image of the provided data set. Now on the basis of that hyperplane it creates two separate margin lines which will have some distance and this line will be parallel to the hyperplane. Also, the SVM gives generalized model that is with less error.



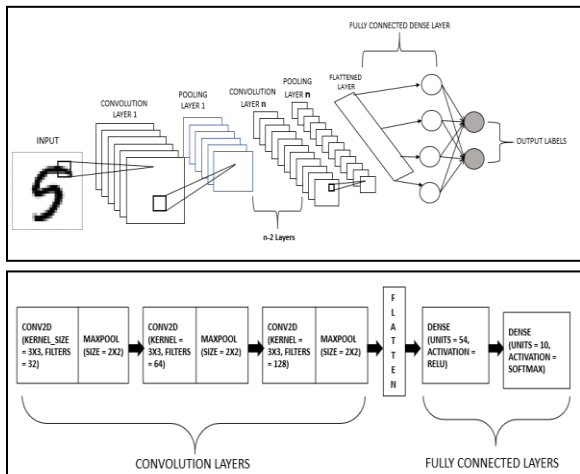
VIII. LOGISTIC REGRESSION (LG)

The others classifier used in this project is logistic regression which is highly applicable in the binary classification as this project is based on binary classification logistic regression will provide better and efficient working. the logistic regression working is somewhat same to the support vector machine that is both of them work on the basis of plane structure of plane graphs. The logistic regression will have a sigmoid function that is a curve line which will going to address and cover all the respective points which could not be get covered by a straight line that is in case of linear regression. Now because of this structure there is a high chance of true output.



IX. CONVOLUTION NEURAL NETWORKS (CNN)

The last classifier used in this project is convolution neural network. As compared to the earlier to classifiers CNN has good response the image classification that is its accuracy is highest among the rest classifiers. Now for the CNN the data base is needed to Be in a standard form that is all the images within the database should be same, CNN reduces the image dimension, except all the importance parts of the image for the classification purpose. The various technique used for the convolution neural networks are the kernel that is it is a filter for the feature detection, stride, padding, pooling, flatten.



X. TUMOR DETECTION

The initial process of tumor detection would be based on the module importing in the program. These modules will be NumPy, mat-plot lib, OpenCV. Here

the NumPy will going to help in the image related part as it will act as an array. With the help of OpenCV we can get standard and various number of outputs in intelligent real time machine learning and image processing outputs. Then there will be various steps that is checking the size of the image for the detection purpose, image preprocessing techniques such as black and white image, filtering the image, apply Edge detection on the image in order to detect and specific location and then providing thresholding on the image with the specified threshold value, applying morphological operations in it for finding out and eliminating white and black noises, and contours specification.

XI. TUMOR CLASSIFICATION

For image classification we have imported required models such as: -

1. NumPy.
2. pandas.
3. sklearn.
4. TensorFlow etc.

This module will be going to help the data set in various other processes. For the SVM and LG part the database is getting split into two parts. Checking the size and shape of the database as it is required to maintain a standard size and shape of all the images of the database, verifying the conversion of images step. Then providing the split data set for the training section. For the CNN part it will train its program and provide various CNN algorithm in it that is convolution 2D kernel, Max pooling, dense. After all the classifier's training model is constructed, they will be proceeding to execute and train the program. After successful training path testing part will begin where checking of the other database will take place. Based on the testing output or score it will be classified that which of the respective classifier has given a very effective and highly appropriate accuracy.

XII. WORKING

The working of the project has been divided into three major modes. But before there are few steps which are needed to be performed. the working of the program will initialize then all the important models are getting installed into the system. after the module or library is getting installed the path of the data set is needed to be

provided. Now the detection step will process. In the detection step the image will be passed through various image preprocessing techniques. In the detection step we will be able to detect the tumor within an MRI image successfully. Then comes the training mode. For the training mode there are several things which are needed, that is conversion of the input images where we will be going to convert all the images within the data set into a standard format so that there is no problem or any glitch during the data set training. After the conversion of the data set the details, it will get split. As it is a binary classification the data set will be of either

1. Benign tumor.
2. Malignant tumor.

Now the training mode will begin. Training mode will have three classifiers which will be going to classify the data set and clean our data set for those classifiers, we will be going to provide different training aspect and construct a model for it. After successful training of all the three classifiers we will get training accuracy and loss value. After this we will get training score for the provided split data set. Finally, the testing mode begin by considering the training mode techniques, which means that here also we will be going to use the classifiers and test the model which we have trained it earlier. Now in this testing mode we will going to test the model but by using different database because we are testing the model and classify database based on the training mode and implementing this on the testing mode new database. For the testing we will get testing accuracy and loss and finally we will get the testing score for the respective tumor type. Now on the basis of the testing score we can analyze which of the following classified has worked really efficient while classifying the database based on the provided type.

XIII. PYTHON

Python is a very compatible language in terms of our project. Python helps in statistical mathematical programming language. It can satisfy all the important parts, aspects which will be required in any of the artificial intelligence and machine learning. It provides various libraries which will be helpful for the classification purpose. the project works completely on the Python programming which will going to include other module or libraries respectively. due to the following reasons:

1. Python code is more compact and readable than MATLAB.
2. The python software is free and available even with minimum system requirements.
3. It is an open source and also provides more graphic packages and data sets (with TensorFlow backend 2.3.0 version)

XIV. TENSORFLOW, NUMPY AND IMPORTANT MODULES

TensorFlow is a library which is used for computing, and mostly it is used in machine learning. In this project TensorFlow will play important role, where we will be needing it to compute the data based on the different training aspects. Also, with the help of TensorFlow we will be able to predict the data for the unknown time intervals. Also, number is an important library for the machine learning as it is used for the mathematical operations as well as logical operations. NumPy arrays will work very efficient as they have equivalent elements in it, which will help the program not to involve the unspecified elements. Now this Python with NumPy will provide us the data of the images into array format which would be helpful to know the specific location or to eliminate the specific region. The matplotlib library will going to help in expressing the output images or graph in a very appropriate manner, for the output part we have also used open CV. Sklearn is used for the various other purposes such as classification, decoration where we will be going to predict some part, preprocessing that is the extraction of the features and normalization of the image or the data set. OpenCV library is imported in this project as it provides real-time computer vision which means that it can be used for the image preprocessing and various other operations like importing the image, writing on the output image, providing filtering on the image, changing the image quality, and converting the real image pixel value into to specific array.

XV. VS CODE AND JUPYTER NOTEBOOK (IDE)

For the working of the program, we have used vs code as in our program editor and python as our main base. The efficiency and the environment of the visual studio codes is quite better as compared to the rest program editors. Visual studio code is light on any

system so it will help to continue your program on any system requirement. Plus, there are many shortcuts which we have implemented in vs code which really makes our work more straightforward. And for the main presentation part we have use Jupyter Notebook as in our integrated development environment that is (IDE) as it is an open-source application which is easily create by any of the user or could be edited easily. Also, the Jupyter Notebook helps in various numerical simulations visualizing the data machine learning deep learning and analyzing the statistical outputs.

XVI.IMPORTANT POINTS AND FUTURE SCOPE

It is considered as the best ml technique for image classification for to high accuracy. Image pre-processing required is much less compared to other algorithms. It reduces images to a form which is easier to process without losing features which are critical for a good prediction by applying relevant filters and reusability. It can automatically learn to perform any task just by going through the training data i.e., there no need for prior knowledge. We can design a user-friendly interface in hospitals which permit doctors to smoothly control the effect of tumor and propose treatment for it. Right now, this is just a software-based system, but by using the RDBMS we can easily get the access of large number of user's database as it will gives much more efficiency in the field work such as Clinics or Hospitals. Also, with the help of any frontend method we can develop a small application by using same python program in backend.

XVII.CONCLUSION

The project and the paper try to impose that the brain tumor can be detected and classify into the respective tumor type by using the machine learning algorithms. The classifiers have helped to classify the data set into the respective tumor type also the detection part will be going to perform the preprocessing technique successfully in order to detect the tumor. All the models or libraries have performed well and given the appropriate results. The program will only be going to detect the MRI image as it is developed for it, it won't be working on the X-ray images as conversion is done separately for the MRI images only. Now because of this project the medical staff, hospital, and clinics will

gets very beneficial as it is a less time-consuming technique with good results, and as discussed before the complete project is working on real-time there is no chance of any data manipulation and no additional hardware installation is required so the cost of this technology will be negligible as programs are completely software based.

ACKNOWLEDGMENT

The process of learning the artificial intelligence and machine learning has given a lot of experience to me and my group. New things have come in this experience. I would like to express my respect and gratitude to our project guide prof. Ragini Bhojar, for providing us the help and guidance which help us throughout this project.

REFERENCE

- [1] Chaganti, Sai & Nanda, Ipseeta & Pandi, Koteswara & Prudhvith, Tavva & Kumar, Niraj. (2020). Image Classification using SVM. 1-5. 10.1109/ICCSEA49143.2020.9132851.
- [2] Alzubaidi, L., Zhang, J., Humaidi, A.J. et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. J Big Data 8, 53 (2021).
- [3] Human Activity Recognition using CNN Neha Nilesh Jadhav Sarnaik feb 2020
- [4] Read this paper if you want to learn logistic regression Antônio Alves Tôrres FernandesI, Dalson Britto Figueiredo FilhoI, Enivaldo Carvalho da RochaI, Willber da Silva NascimentoI I Postgraduate Program in Political Science, Federal University of Pernambuco, Recife, PE, Brazil Accepted in May 16, 2020
- [5] Boateng, Ernest Yeboah & Abaye, Daniel. (2019). A Review of the Logistic Regression Model with Emphasis on
- [6] Medical Research. Journal of Data Analysis and Information Processing. 07. 190-207. 10.4236/jdaip.2019.74012. A comprehensive survey on support
- [7] Yang, Li & Shami, Abdallah. (2020). On Hyperparameter Optimization of Machine Learning Algorithms: Theory and Practice.
- [8] Seetha, J. & Selvakumar, Raja. (2018). Brain Tumor Classification Using Convolutional

Neural Networks. Biomedical and Pharmacology Journal. 11. 1457-1461. 10.13005/bpj/1511.

- [9] Sarkar, Sobhangi & Kumar, Avinash & Chakraborty, Sabyasachi & Aich, Satyabrata & Sim, Jong-Seong & Kim, Hee-Cheol. (2020). A CNN based Approach for the Detection of Brain Tumor Using MRI Scans. Test Engineering and Management. 83. 16580 – 16586.
- [10] Nadeem, Waqas & Ghamdi, Mohammed & Hussain, Muzammil & Khan, Muhammad & Khan, Khalid & Almotiri, Sultan & Butt, Suhail. (2020). Brain Tumor Analysis Empowered with Deep Learning: A Review, Taxonomy, and Future Challenges. Brain Sciences. 10. 1-33. 10.3390/brainsci10020118.