Malignant Tumor Detection Performance Analysis using Convolutional Neural Networks and SVM Classifier Model

Avik Pathak, Suswan Biswas, Koushani Bhattacharyya, Aparna Biswas,

Arindam Chakraborty , Soham Kanti Bishnu

Department of Electronics and Communication Engineering

Institute of Engineering and Management, Kolkata

Abstract.

An unusual mass of tissue in which some cells multiply and grow uncontrollably is called brain tumor. It starts growing inside the skull and interposes with the regular functioning of the brain. Brain tumors can be detected at an early stage using MRI or CT-scanned images when it is small. Early detection is very necessary as the tumor can possibly result in cancer. Tumor detection and removal is one medical issue that still remains challenging in the field of biomedicine as it is prone to human errors. We have tried to understand the effectivity in successfully detecting a malignant tumor through the implementation of Convolutional Neural Network using Binary Cross-Entropy and Categorical Cross-Entropy as loss functions along with SVM Classifiers and Logistic Regression. We do hope that our performance analysis helps other researchers and ML engineers to see the constraints and advantages of using these models for accurate prediction.

Keywords. Tumor detection, Convolutional Neural Network, Binary Cross-Entropy, Categorical Cross-Entropy, SVM Classifiers, Logistic Regression.

1. INTRODUCTION

There has been and will be extensive research on the Tumour Malignancy analysis from MRI reports and they strive to develop a model that most accurately predicts the malignancy and the tendency of a given tumor to be malignant in the future. The research developed by Boucif Beddad, Kaddour Hachemi, and Sundarapandian Vaidyanathan in the paper named 'Design and implementation of a new cooperative approach to brain tumor identification from MRI images'[10] have quite extensively treaded on this field and has given us some really helpful insights to develop the analysis and implement few newer techniques to get some unique insights into the project. Although the proposed segmentation algorithm was not directly used to train our datasets, the modeling techniques we used were analogous to the model we wanted to implement.

2. METHODOLOGY

2.1. Deriving Raw Data:

The Datasets used in the analysis mainly consisted of MRI scan reports of 1000 images, both malignant and non-malignant.In SVC we reduced the pixels to 200 X 200 while in CNN , 64 X 64 image was used .

2.2. Data Preprocessing:

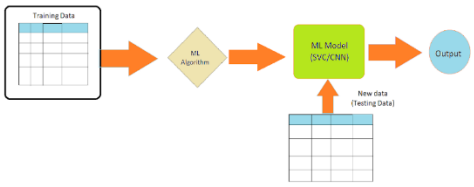
The raw data that we have first needs to be converted to some numeric values such that we can perform certain operations on it. To do this the very fundamental step that needs to be performed is Importing the required libraries for the operations.

2.3. Dataset Operations :

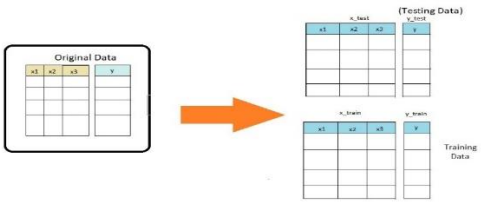
In machine learning, we have some raw data like images in this case which have a unique format of .jpg. We can't perform any operation on the data with this format as a machine understands only 1's and 0's, on a broader scale we may refer to them as elementary mathematical operations.

2.4. Splitting the Data:

To analyze the performance of an algorithm, this is an extremely important technique that is used to check the model's accuracy over an unknown dataset that is extracted from the raw data. original dataset. A conventional flow chart of implementing train_test_split over a model would look like this:



Generally, We refer to the raw training data as x_Train and The data label is stored in y_train. The testing data is referred to as x_test and the label of x_test is y_test.



2.5. Feature Scaling of the Data:

Feature Scaling is primarily used to remove the independent variables in the data and normalize the data such that we can apply mathematical methods to fit the data in an ML Algorithm.

3. OBSERVATIONS:

Coming to the most crucial part of our discussion, we have tried to observe how the Tumor detection model would perform as it is passed through SVC, simple Logistic Regression and finally, neural networks with a variation in loss and activation functions.

3.1. Logistic Regression Performance:

Evaluation

```
In [18]: print("Training Score",lg.score(pca_train,ytrain))
         print("Testing Score", lg.score(pca_test,ytest))

         Training Score 1.0
         Testing Score 0.9591836734693877
```

As we can see, the logistic regression model shows a training score of 1 which means the data perfectly fit the curve. The testing score however is not good enough (as would be proved by the other models).

3.2. Support Vector Machine(SVM) Performance Analysis:

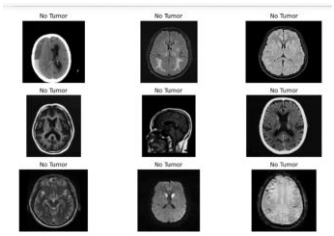
The model fits the data quite well and generates praiseworthy output, both in testing data and unknown data.

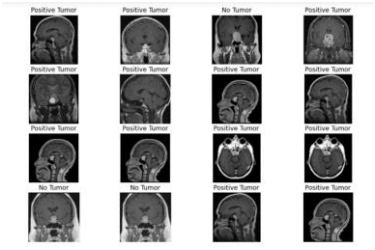
```
In [19]: print("Training Score",sv.score(pca_train,ytrain))
         print("Testing Score", sv.score(pca_test,ytest))

         Training Score 0.9938587512794268
         Testing Score 0.963265306122449
```

The training score, in this case, seems to be less than linear regression but the testing score improves from the previous model. Thus, when using known data, both models work perfectly.

For the known Set of “No Tumor” and “Malignant Tumour”

The result is:





As we can see, in this model, if we take a probability of success observing this graph. It would roughly come around to 1316 which is quite good. Yet, can we do better is what all

ML engineers should be asking themselves. Here comes our last study, CNN (Convolutional Neural network).

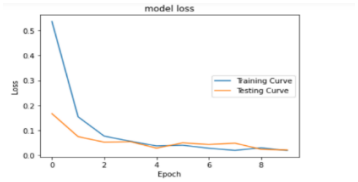
3.3. Convolutional Neural Network:

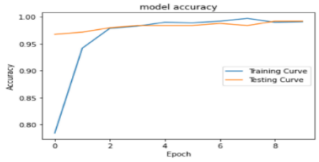
In CNN, accuracy score comes to be:

```
score, acc = model.evaluate(xtest, ytest,verbose=1, batch_size= 16)
print('test accuracy:', acc)

predict_x=model.predict(xtest)
classes_x=np.argmax(predict_x,axis=1)
print(classes_x)
```

```
16/16 [==============================] - 0s 16ms/step - loss: 0.0206 - accuracy: 0.9918
test accuracy: 0.9918367266654968
```

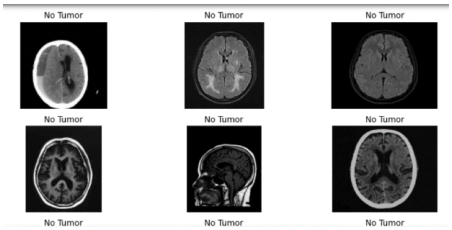
An extremely high-test score over the training data.The curves of accuracy vs epochs and loss vs epochs

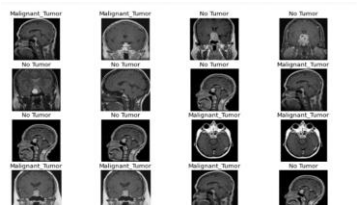




Let's now arrive at the observation of its performance with the unknown dataset.

With the 'No Tumor " dataset and "Malignant Tumor" dataset :

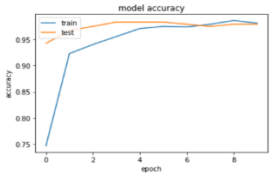


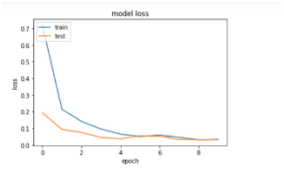


The data works fine.Quite interestingly, we find here that the probability of success over this data is just 816 which is just 0.5 which is really bad.Still, an interesting observation as it is. SVC remains to be leading in terms of success over unknown data, although it trails the neural network in terms of its accuracy score.

3.4. Categorical Cross Entropy:

The Accuracy vs Epoch Curve and Loss vs Epoch Curve :





Accuracy with increasing epoch values is farely stable .Loss of the function is really less with increasing epochs.The probability of success comes out to be 816 again which is similar to the binary cross entropy result. It is pertinent to see that the accuracy score of the model over the training dataset was 0.979 which was less than that of binary cross-entropy.

4. FUTURE SCOPE

Improved models with higher accuracy may be developed using different machine-learning models.Our existing model does not take into account the age, gender, and prior medical history of a patient for detection. Developing a universal model to detect all kinds of tumors in its earliest stage will ensure the earliest detection of disorders which in turn will reduce the chances of spreading malignancy to other parts of the body.

5. CONCLUSION

In this paper we discussed the design of a model which shows if there are any traces of brain tumors using Magnetic Resonance Imaging (MRI) scans. The model performs segmentation of MRI scan images and pinpoints if there are any traces of tumor in the brain. From comparative analysis, it is clear that Deep Learning techniques and algorithms have great power and ability to handle large amounts of data.

6. REFERENCES

[1] Jahnavi Athalur., Application of Edge Detection for Brain Tumor Detection

[2] Dr. Navneet Singh Sikarwar., Image Processing Techniques for Brain Tumor Detection: A Review

[3] Hao Dong, Guang Yang, Fangde Liu, Yuanhan Mo & Yike Guo., Automatic Brain Tumor Detection and Segmentation Using U-Net Based Fully Convolutional Networks

[4] JavariaAmin, MuhammadSharif, MudassarRaza, TanzilaSaba, Muhammad AlmasAnjum ., Brain tumor detection using statistical and machine learning method

[5] Pankaj Sapra, Rupinderpal Singh, Shivani Khurana., Brain Tumor Detection Using Neural Network

[6] Agarap, Abien Fred. (2018). Deep Learning using Rectified Linear Units (ReLU)

[7] Huang, Shujun et al. “Applications of Support Vector Machine (SVM) Learning in Cancer Genomics.” Cancer genomics & proteomics vol. 15,1 (2018): 41-51. doi:10.21873/cgp.20063

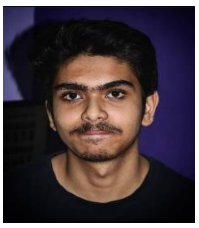
[8] Sperandei, Sandro. “Understanding logistic regression analysis.” Biochemia medica vol. 24,1 12-8. 15 Feb. 2014, doi:10.11613/BM.2014.003

[9] Heny Pratiwi et al 2020 J. Phys.: Conf. Ser. 1471 012010

[10] “Design and implementation of a new cooperative approach to brain tumour identification from MRI images” Boucif Beddad, Kaddour Hachemi and Sundarapandian Vaidyanathan 10.1504/IJCAT.2019.097113 [doi]
[11] Maria Nazir, Sadia Shakil, Khurram Khurshid,
Role of deep learning in brain tumor detection and classification (2015 to 2020): A review, Computerized Medical Imaging and Graphics, Volume 91,2021,101940, ISSN 0895-6111
[12] Bhagya Patil, Vishwanath Barkpalli. A Perspective View of Cotton Leaf Image Classification Using Machine Learning Algorithms Using WEKA, 13 May 2021,
PREPRINT (Version 1) available at Research Square [https://doi.org/10.21203/rs.3.rs-502091/v1]
[13] KRajesh Babu et al 2021 J. Phys.: Conf. Ser. 1804 012174
[14] Computational Intelligence in Pattern Recognition,2020, Volume 1120,ISBN : 978-981-15-2448-6

Biographies

Avik Pathak received the bachelor's degree in electronics and communication engineering from Institute of Engineering and Management in 2024 His research areas include deep learning, Optimizing Machine Learning/AI models ,



Suswan Biswas received the bachelor's degree in electronics and communication engineering from Institute of Engineering and Management in 2024 His research areas include mobile security and development, android manifest, mobile application development.

Koushani Bhattacharya received the bachelor's degree in electronics and communication engineering from Institute of Engineering and Management in 2024 Her research areas include blockchain and web3, web application development.

Aparna Biswas is presently a Scientific Officer in the faculty of Electronics and communication engineering at Institute of Engineering and Management(IEM) In Kolkata,India.

Mr. Arindam Chakraborty is currently an assistant professor at the Institute of Engineering and Management, Kolkata. He obtained his M. Tech. in Microelectronics and VLSI Design from the Jadavpur University – Kolkata, India. Mr. Chakraborty has published a number of papers in research Journals and chapters in books, and participated in different reputed conferences on solid state lighting, VLSI and IOT. His areas of interest include Solid State Lighting, Image Processing, Embedded System Design and IOT..

Soham Kanti Bishnu is presently is currently an Assistant Professor in the faculty of Electronics and communication engineering at Institute of Engineering & Management (IEM) in Kolkata, India. Prior to his recent appointment at IEM, he served as a researcher in CSIR-Central Glass and Ceramic Research Institute (CSIR-CGCRI).