Using Machine Learning Architecture to Detect Brain Tumors

Mike Sullivan

8/24/22022

Abstract

Magnetic Resonance Imaging (MRI) scans are the primary way medical professionals study the brain. Doctors are trained to identify irregular cell growth patterns and notify patients if they have or are at risk of cancer. Identifying tumors before they become malignant saves both resources and welfare of patients. In this paper, we apply a basic SKLearn Logistic Regression model, a custom Convolutional Neural Network model, and a transfer learning model based on the VGG16 model to automatically diagnose malignant or benign tumors in MRI scans. We found that the VGG16 model performed the best, with a validation accuracy of 91.6%. Logistic Regression had a validation accuracy of 82.3% and CNN an accuracy of 83.2%. These promising results suggest that automated machine learning models can be used as a MRI scan analyzer to diagnose patients more often and more quickly which may have the ability to identify tumors before they become malignant, saving lives.

1. Introduction

[3]Cancer as a disease stems from malignant tumors, which are unregulated cell growths as a result of cell mutations. These growths can be extremely problematic, as they can disrupt functions of the brain, including clogging vital fluid transfer and harming surrounding tissues. With roughly 5000 yearly diagnoses in adolescents, Brain Cancer is most commonly found in children. In order to diagnose cancer, doctors use MRI scans to pinpoint the location and size of a tumor, which can be used to predict the potential danger of a tumor. Not all tumors in the brain are cancerous, so irregular cell growths can often be determined as benign and are ignored completely. However, many of these low-danger cell growths can develop into higher-danger cancerous tumors. Identifying these low-danger tumors early is key to preventing cancer in the future.

We used three different models to detect brain tumors. Logistic Regression is a statistical technique used to give predictions towards outcomes based on one or more independent variables. In this experiment, the model uses the training data to predict a percent chance that any one of the test images contains a tumor. When testing the accuracy of a model, these predictions are normalized to a 0 or a 1 and compared with the test data.

A Convolutional Neural Network exists as a series of convolving layers, which reduce the base image through filters while maintaining values needed for pattern recognition. At the end of the model, the flattened information is input into a dense layer which can output a probability based on weights and biases that determine if the image does or does not contain a tumor.

VGG16 is a pretrained Convolutional Neural Network of 13 convolutional layers and 3 fully connected layers. We use transfer learning and pre-train the VGG16 convolutional architecture on the ImageNet database made up of millions of images, before fine-tuning the model on both malignant and benign MRI tumor scans for our task.

1

2. Background

[2]This paper compares the abilities of an Artificial Neural Network(ANN) and a Convolutional Neural Network(CNN) on a similar dataset, and achieved validation accuracies of 65% and 89%, respectively. This paper is clear in proving the capabilities of a CNN model in image classification, which can recognize spatial patterns in the pixels, while ANN loses such information when converting images to a single dimensional array of pixels.

[3]The medical community has great hopes for the capabilities of AI. This paper mentions the issue of image segmentation, how it is often undermined by human error, and that a Deep Learning process could prove a solution. Image segmentation is an encompassing term for the processes applied to reduce images to more manageable data. Particularly with larger, multi-dimensional images with voxeling and color patterns, relevant data can frequently be lost. If segmented images are being used to determine diagnoses, it is vital that these images are accurate and optimally generated. The paper cites the use of CNN models to be the most efficient towards the segmentation process, and furthermore uses this data to make accurate predictions about the image. Our goal is to measure the efficacy of different CNN models and compare them to different methods of machine learning.

3. Dataset

Our dataset was taken from Inspirit AI and consists of 253 images of 150x150 resolution. The dataset includes a separate 'labels' array of size 253, each value corresponding to its respective image in the dataset, denoting a '1' for images that contain a tumor and a '0' for healthy images. We employed an augmentation to the dataset by flipping all images on the Y-axis, doubling the sample size. While this augmentation does not provide the models with real data, the extra samples are beneficial toward recognizing different potential patterns. The original dataset consists of 98 healthy brains and 155 cancerous brains, which after augmentation equates to 196 healthy brains and 310 cancerous brains. This was then split by a test size of ⅓: 339 images for training, and 167 images for validation testing.
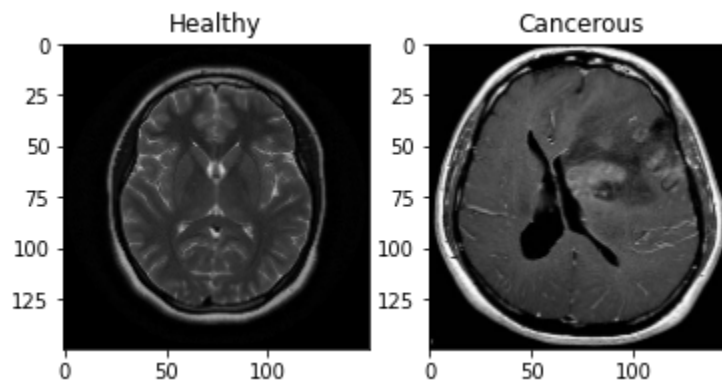

Figure 1. Healthy Brain and Cancerous Brain

4. Methodology / Models

The dataset was imported into the notebook and loaded into a numpy ndarray. Prior to augmentation, the dataset consisted of a four dimensional array of 253 images with three color channels

each. The dataset and labels were then split into train and test arrays. For the Logistic Regression model, we applied a flatten method to reduce each image to a one dimensional array.

The first model we applied to the dataset was a basic SKLearn Logistic Regression with default parameters and a maximum iteration of 1000. After creating the model, we fit it to the flattened training data and applied it to the test data to generate a predictions array. We then compared the predicted labels with true labels to determine metrics and created a confusion matrix and ROC curve accordingly.

To create the CNN model, we applied a self-defined classifier method using a Keras Sequential model. We fit the model to the dataset and instructed a 100 epoch training period, over which the accuracy history is recorded and plotted. After training, we applied it to test data to generate a predictions array. Because Sequential predicts probability values rather than labels, we normalized the predictions at a threshold of 50% in order to compare them to test labels. Using the binary predictions, we generated metrics with the appropriate confusion matrix and ROC curve.

The CNN model begins with a single convolutional input layer of 32 filters and a kernel size of 3x3 units, followed by a 'relu' activation function and a max pooling window of 2x2. We then follow with six identical layers of 64 filters and the same kernel size, activation function, and max pooling window. Finally, the model ends with two Dense layers of units 128 and 64 and an output layer with output activation function 'sigmoid.'
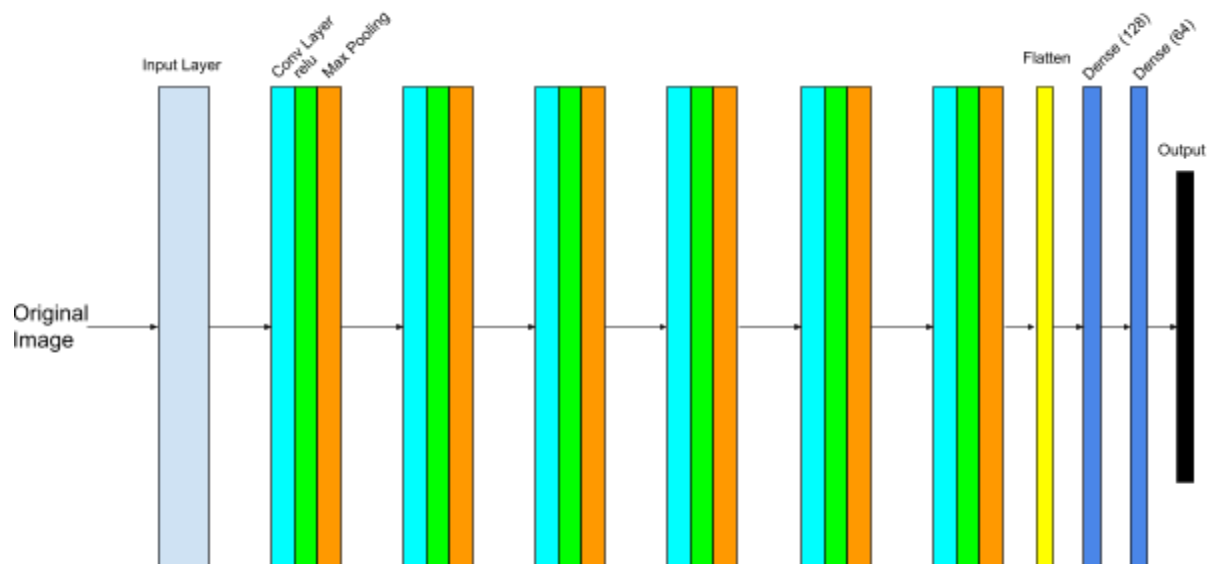

Figure 2. Architecture of Self-Defined CNN Model

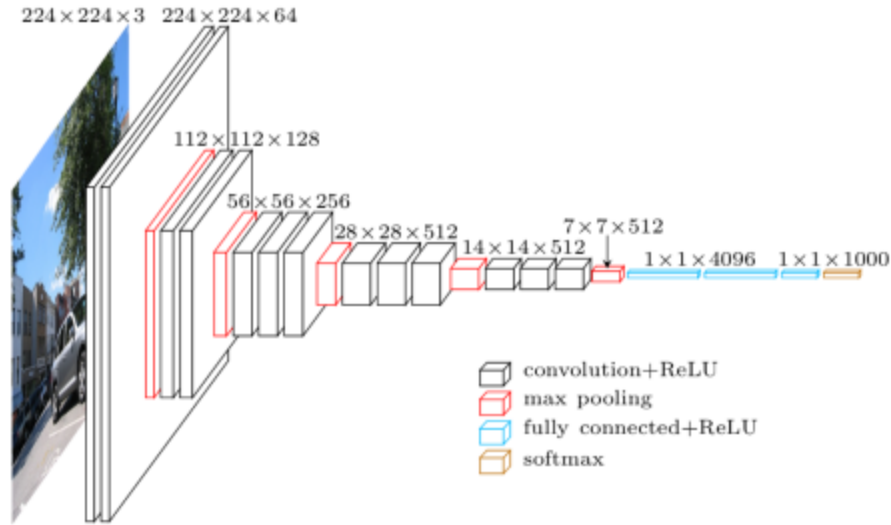The VGG16 model is a pre-defined CNN with the architecture shown below:

Figure 3. [4]Architecture of VGG16

For each layer, we set the 'trainable' property to 'True,' allowing the weights and biases to change to fit our dataset. Similar to the CNN model, we chose to add two extra dense layers of units 128 and 64, plus an output layer of 'sigmoid' activation function.

5. Results and Discussion

Each model was given the dataset and trained for 100 epochs. The first model, Logistic Regression, was trained with a flattened version of the dataset, in which the 3 dimensional array for each image became one long single dimensional vector. The model achieved a validation accuracy of 82.6%.

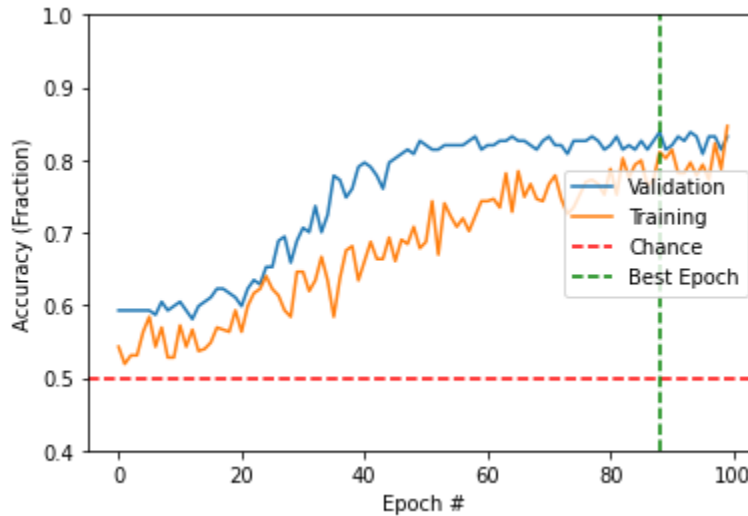Comparatively, the CNN model achieved a validation accuracy of 83.2%. The accuracy over time is shown below:

Figure 4. Accuracy History Plot for CNN model.

4

The CNN and Logistic Regression model performed very similarly. Because the Logistic Regression model performed so well comparatively, we can assume the structure of the CNN model could use further optimization.
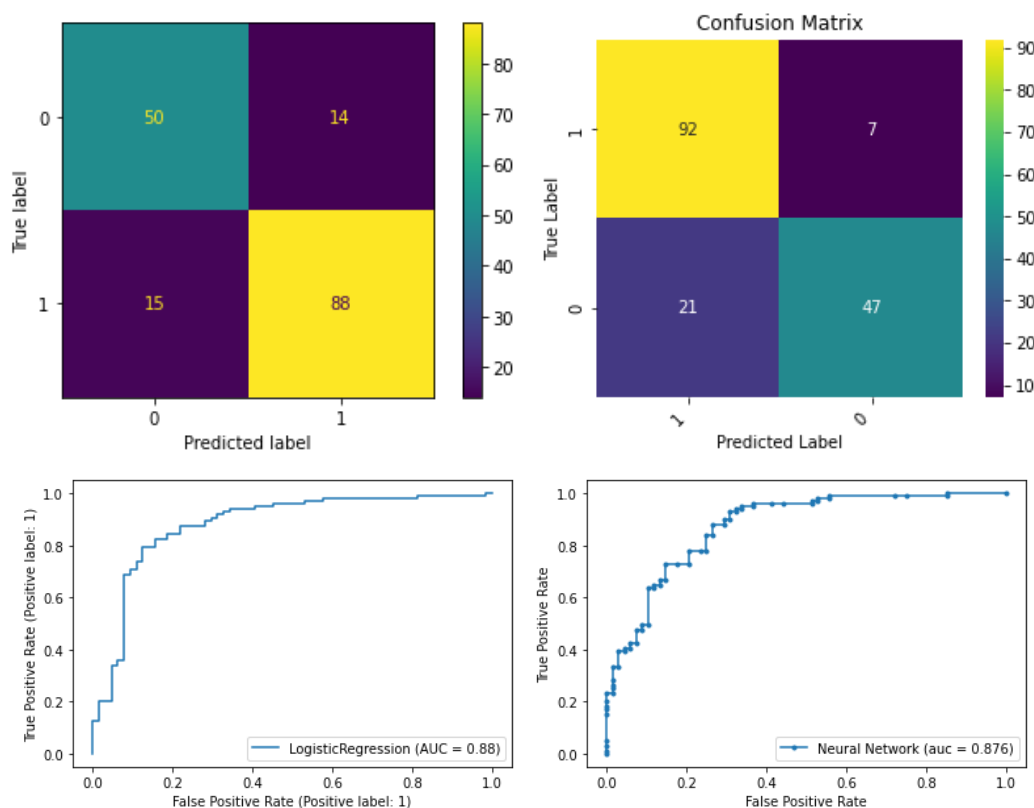

Figure 5. Confusion Matrices and ROC Curves for Logistic Regression(Left) and CNN(Right)

One major feature about the Confusion Matrices is the disparities between the two in their predictions of incorrect results. In this case, CNN saw a significant affinity towards predicting false negatives, whereas Logistic Regression was relatively equal in both predictions.

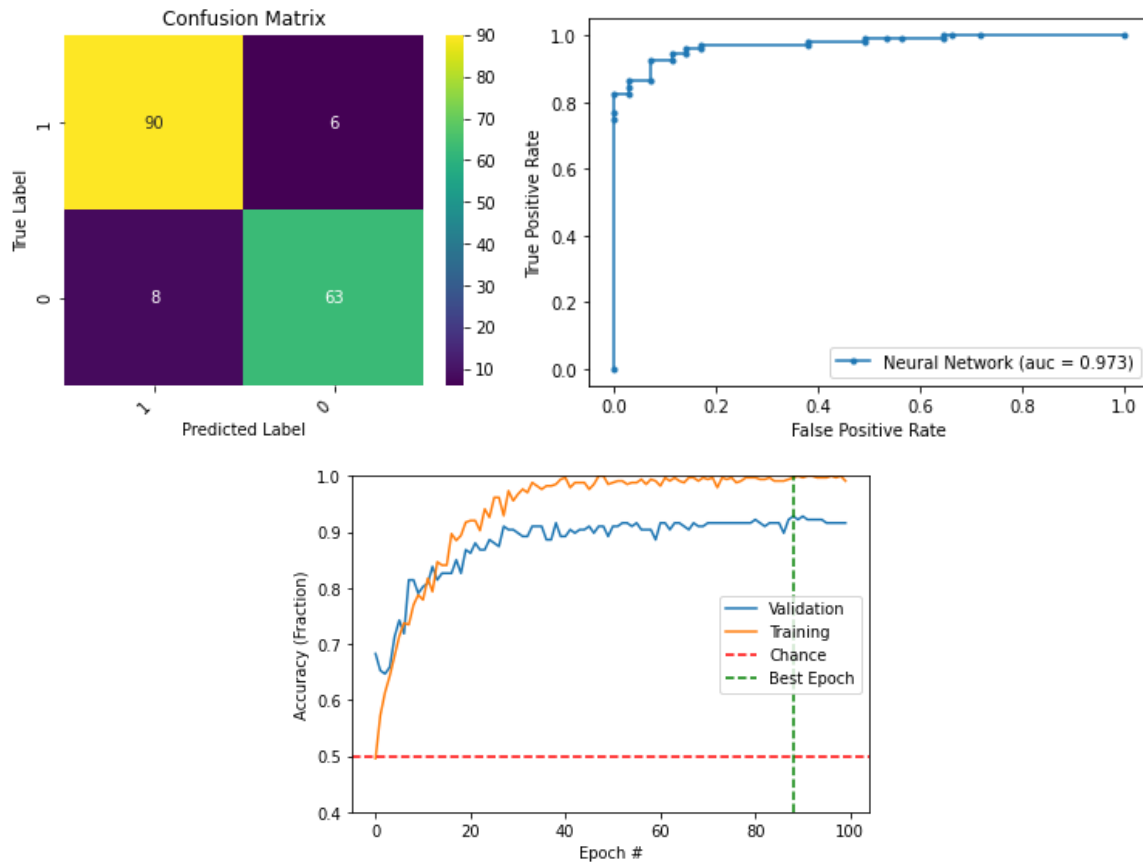The best performing model by far was the VGG16 model, with 91.6% validation accuracy.

Figure 7. Confusion Matrix, ROC Curve, Accuracy History for VGG16

The metrics for all three models can be seen below:
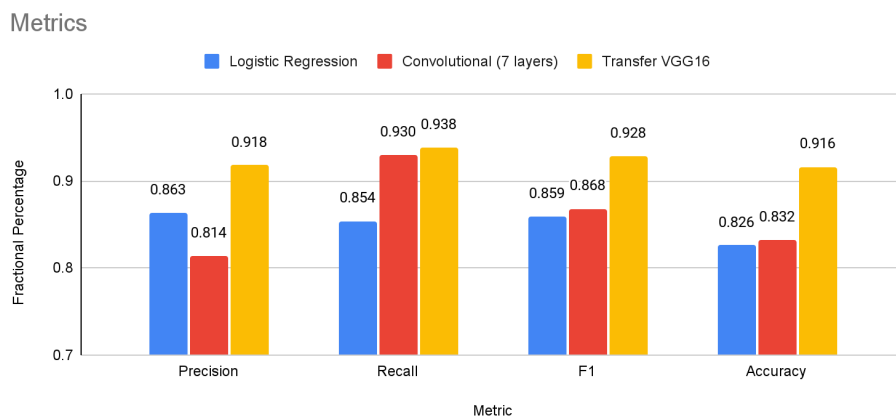

Figure 8. Metrics

The VGG16 model has clearly the highest metrics of the three in all four metrics with an accuracy of 0.916 and an F1-score of 0.928.

6. Conclusions

The purpose of this paper is to create a streamlined method of detecting the presence of tumors, so that doctors can put greater resources toward identifying the degree of the tumor and making precautions towards any risk of cancer. The end goal was to develop a model that could do so with the most favorable metrics. To do this, we tested the abilities of a Logistic Regression Model, a self-defined CNN model, and a modified VGG16 model. Of the three models, VGG16 managed to perform the best with 91.6% validation accuracy. Its higher complexity architecture proved most efficient toward identifying the dataset.

The results of this paper left much to be considered. The metrics of a self-defined CNN model and a VGG16 model prove that convolution is the future of image segmentation analysis. It would be beneficial to further develop new, more optimized models, as well as look to employ something like VGG16, with its robust training set, toward a specialized medical environment. Furthermore, in terms of perfecting image segmentation, optimal image preprocessing is worth exploring. Our dataset was relatively small and fundamental, and experimenting with higher quantities of higher definition images with greater relevant data should result in a relative increase in performance and reliability.

Acknowledgments

References

[1] P Gokila Brindha *et al* 2021 *IOP Conf. Ser.: Mater. Sci. Eng.* 1055 012115
https://iopscience.iop.org/article/10.1088/1757-899X/1055/1/012115

[2] Akkus, Z., Galimzianova, A., Hoogi, A. *et al.* Deep Learning for Brain MRI Segmentation: State of the Art and Future Directions. *J Digit Imaging* 30, 449–459 (2017).
https://doi.org/10.1007/s10278-017-9983-4

[3] Reference Articles:
https://www.hopkinsmedicine.org/health/conditions-and-diseases/brain-tumor
https://www.ibm.com/topics/logistic-regression#:~:text=Logistic%20regression%20estimates%20the%20probability,bounded%20between%20200%20and%201.
https://insightsimaging.springeropen.com/articles/10.1007/s13244-018-0639-9

[4] VGG16 architecture:
https://miro.medium.com/max/470/1*3-TqqkRQ4rWLOMX-gvkYwA.png