A Comparative Assessment of CNN and Machine Learning Models on MRI Brain Tumor Images

Alexander Hernandez	Elie Bonkoungou	Gabriel Giron	Nouhad Rizk
aherna98@cougarnet.uh.edu	edbonkou@cougarnet.uh.edu	gjgiron@courgarnet.uh.edu	njrizk@uh.edu

ABSTRACT

Brain tumors, ranking as the 10th leading cause of mortality according to Cancer.net, constitute a critical health concern necessitating early detection and precise classification for effective intervention. This study draws upon a comprehensive compilation of diverse datasets, comprising 7,023 human brain MRI images sourced from figshare, SARTAJ dataset, and Br35H, categorized into glioma, meningioma, no tumor, and pituitary groups. The primary objective is to harness the collective power of these datasets to develop a robust predictive tool aimed at early brain tumor detection and lifesaving interventions. This diverse dataset, coupled with machine-learning methodologies, offers a pivotal resource for optimizing brain tumor diagnosis and treatment by facilitating early detection. To achieve this, this study will initially employ Support Vector Machines (SVM) and Random Forest Classifier (RFC) algorithms to enhance accurate classifications and diagnostic precision. Furthermore, this study will leverage Convolutional Neural Networks (CNNs) for multi-task classification, concurrently addressing tumor detection, classification, and localization within MRI scans, including considerations of malignancy, grade, and type. Moreover, four pre-trained models will be involved namely: VGG16, Xception, InceptionV3 and ResNet50. This study uses different techniques involving hyperparameter tuning, Dropouts and Early Stopping features to prevent models from overfitting. All transfer learning models achieved satisfactory levels of accuracy when classifying brain tumors with minimal loss. Statistical variables of significance like F1 scores and specificity are used to gauge further insight into how the models performed. Finally, this research aspires to advance brain tumor diagnosis and improve patient outcomes through the integration of deep learning techniques.

Keywords: MRI images, brain tumor, machine learning models, Tensorflow, Scikit-learn, Support Vector Machines, Random Forest Classifier, deep learning, figshare, SARTAH, Br35H, Convolutional Neural Networks, hyperparameter tuning, Dropouts, overfitting, transfer learning, VGG16, InceptionV3, Xception, ResNet50.

INTRODUCTION

According to the National Cancer Institute there is no single test that can diagnose cancer, but if there is suspicion of cancer, a doctor can order scans such as: lab tests, biopsy and imaging scans like MRIs [1]. However, this study aims to use the power of deep learning and machine learning models in order to provide aid to doctors in their examination and classification of brain tumors. The possibility for these tools to be trained on large amounts of images of existing tumors and make accurate predictions as to the type of brain tumor gives doctors a significant tool with which to help patients. This would lead to better outcomes for patients who are diagnosed early on and can be treated accordingly.

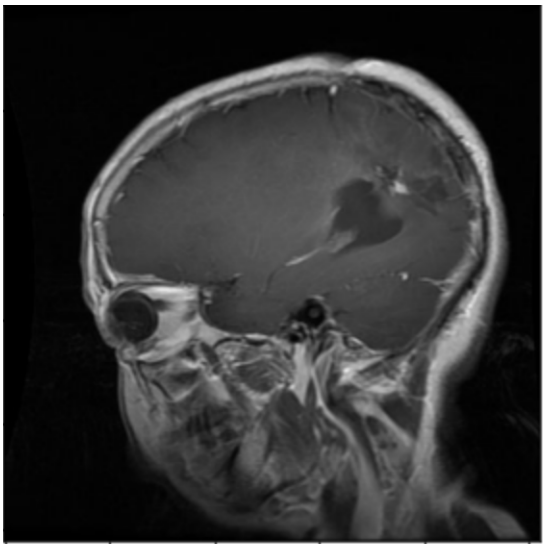

Fig 1. An example image from the dataset

RELATED WORK

The Classification of brain tumors in MR images using deep 'spatiospatial' models study compared various models, including ResNet Mixed Convolution, against others for classifying Low-Grade Glioma (LGG) and High-Grade Glioma (HGG) tumors. The pre-trained ResNet Mixed Convolution model exhibited strong performance, especially when dealing with dataset imbalance and computational efficiency. It achieved a macro F1-score of 0.9345 and a mean test accuracy of 96.98% [10]. This study's objective is to build upon and significantly advance the findings of prior research, striving for enhanced accuracy, efficiency, and applicability in the realm of MRI brain tumor detection and classification using CNN and machine learning models.

PROPOSED METHODS

Dataset Preparation

The dataset that was used for the research was MRI scans from figshare, SARTAJ dataset, and Br35H, where the input image size was changed from varying height and width around 230 pixels to constant 150 by 150 pixels and was converted to grayscale. Although the sample size was quite large for the SVM model, it kept the same settings and was allowed a longer training time.

CNN model

The Tensorflow and Keras libraries allow the ability to create a sequential convolutional neural network model. The architecture for the CNN model followed a very simple and straightforward path; initially there is a Keras Conv2D layer, wherein there are 32 nodes. That convolutional layer uses a small kernel size of 3x3 pixels. Then, these kernels are passed through the ReLU activation function. After that, the output is fed into a MaxPooling2D layer with 2x2 pool size that reduces the number of outputs by a factor of four. The number of strides in this layer is 2, meaning that the pooling window will move two units at a time after every operation. This set of convolution, activation and pooling is then repeated another two times with an increasing number of nodes in the convolutional layers from 32, 64 to 128. The output of the third pooling layer is then flattened/reshaped and fed into a fully-connected Dense layer of 64 nodes and with activation function ReLU. The last layer of this fully-connected layer is a Dense layer of 4 nodes, since the CNN model is predicting 4 classes, with a softmax activation function.

Hyperparameter tuning was used on the CNN model using KerasTuner. Specifically, this study addressed finding the best learning rate. This model was trained with learning rate values of: 0.01, 0.001, or 0.0001 [4].

In an effort to prevent overfitting, an EarlyStopping function was used when the model is fitted on the training dataset, wherein if after two consecutive epochs no improvement in the accuracy is found, the callback function calls the EarlyStopping function and the model stops training. Additionally the model's best weights are restored and kept.

The model was trained on the Adam optimizer, the reason being that using an optimizer for a network will speed up the training speed. More specifically, the Adam optimizer was chosen over the others because it is a combination of momentum and RMSProp optimizers. Meaning it takes into account exponential decay average of past gradients and square gradients. It is suitable for complex data [2].

CNN model with Dropout

To improve the existing CNN model, Dropout layers come into play. The Dropout layer is a method that aims to control and reduce the likelihood of overfitting. In this study, following every MaxPooling layer, a 20% dropout layer was added. In this process, this means that 20 percent of individual nodes are excluded in various training runs as if they were not part of the model architecture at all [3]. For the training of this model, the best learning parameter from the basic CNN model was used instead of having to retrain the new CNN model with the hyperparameter tuning.

SVM: Support Vector Machine classifier

Support Vector Machine (SVM) models are supervised machine learning algorithms used for classification and regression. In this study, it is used to find a hyperplane that classifies MRI scans based on features of the image data. After the image size was reduced and color converted to 1 grayscale channel, the sample size remained large enough for standard SVMs to run at a reasonable time. To optimize this classifier, the LinearSVC function from the scikit-learn library is used, which is a variant of the SVC function that supports a one-vs-the-rest scheme classification [11]. Using a default loss function, the primal optimization problem, and not requiring kernels made the classification process run significantly faster.

RF: Random Forest classifier

In this study, the Random Forest Classifier (RFC) was also utilized for MRI brain tumor classification. Renowned for its effectiveness in handling complex datasets, RFC constructs multiple decision trees to improve prediction accuracy and robustness. It's particularly adept at feature selection and managing overfitting, making it a valuable tool in medical imaging analysis, especially when dealing with intricate patterns in MRI scans. The RFC's integration aims to enhance the overall precision and reliability of tumor classification that will be used as a comparison to the above models.

Transfer Learning Models

A pre-trained model is a saved network that was previously trained on a large dataset. This study repurposes it and uses it to classify four types of brain tumors. Specifically, all transfer learning models will be trained on the ImageNet dataset. Those acquired weights will then be used to address the issue of classification [5].

The following pre-trained models were used in this research:

VGG16: In the 2014 ImagNet challenge, a Visual Geometry group published a paper called "Very deep convolutional networks for large-scale image recognition" won 1st and 2nd place in object detection and classification [6]. That submission was the VGG-16 convolutional neural network. VGG-16 is a type of CNN model considered to be one of the best computer vision models to date. The 16 refers to the 16 layers that have weights. There are 13 convolutional layers, five MaxPooling layers, and 3 Dense layers that sum up to 21 layers. The creators specifically focused on convolutional layers of 3x3 filter with stride 1 and always used the same padding and maxpool layer of 2x2 filter of stride 2.

Xception: Xception by Google stands for Extreme version of Inception. With a modified depthwise separable convolution, it is even better than InceptionV3. Compared with conventional convolution, there is no convolution across all channels. This means the number of connections is fewer and the model is lighter. Xception outperforms VGGNet, ResNet and InceptionV3. It is 71 layers deep and has 22,855,952 trainable parameters [8].

InceptionV3: InceptionV3 is an image recognition model that has been shown to attain a greater than 78.1% accuracy on the ImageNet dataset. The model is made up of convolutions, average pooling, max pooling, concatenations, dropouts and fully connected layers. Batch normalization is used extensively throughout the model. The loss is computed using the softmax activation function [7].

ResNet50: ResNet50 is a variant of the ResNet model with 48 convolutional layers along with 1 MaxPool layer and 1 average pool layer. Residual networks, like ResNet, use the concept of skipping connections through which it resolves the problem of vanishing gradient. In skip connections, traditional layers are stacked one after the other, but they include the original input to the output of the convolutional block. The uses for the ResNet architecture include image classification, object localization and object detection [9].

RESULTS AND DISCUSSION

Model	Accuracy	Weighted F1 Score
CNN	0.97	0.97
CNN with Dropout	0.92	0.92
VGG16	0.94	0.94
InceptionV3	0.91	0.92
Xception	0.96	0.96
ResNet50	0.79	0.77

TABLE 1: ACCURACY, WEIGHTED F1 SCORE ON THE TESTING DATA FOR THE DIFFERENT CNN MODELS TESTED

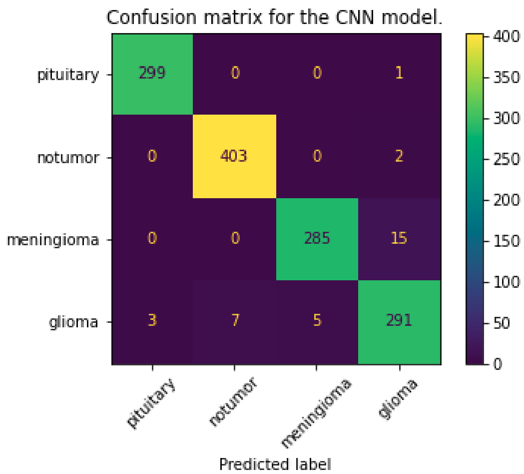

Fig. 2. Confusion Matrix for CNN model

CNN Model:

It was observed in the training phase of the model that it overfitted quite quickly. Specifically, it converged to 100% accuracy by the 12th epoch. We see based on the chart that this CNN model achieved an accuracy of 97% with an F1 score just as good. This means that both precision and recall were high, therefore the model did not generate too many erroneous annotations and it successfully found annotations it should have found.

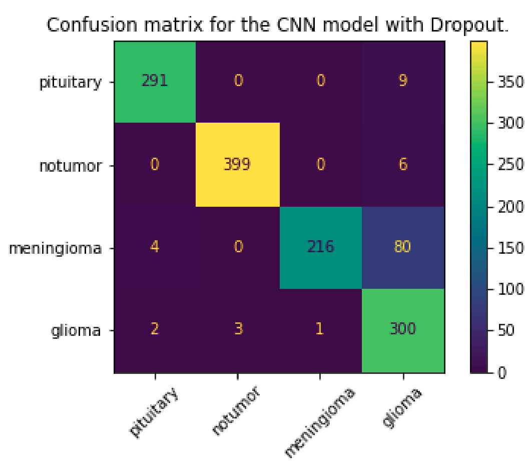

Fig. 3. Confusion matrix for CNN with Dropout

Based on the confusion matrix for this model we see that it predicted 15 cases of gliomas when in fact the observed classification was meningiomas. It also erroneously predicted gliomas when they were not.

CNN with Dropout

In the case of the CNN model with three dropout layers of 20% after every maxpool layer, we see that the accuracy and weighted F1 score is 92%. Again, this suggests a good level of precision and recall from the model. However, this differs from the CNN model because it makes even more mistakes on the same exact issue of predicting a brain tumor was a glioma when in fact it was a meningioma. A total of 80 misclassifications happened as a result. Moreover, it seemed like the ability to accurately predict gliomas diminished as a result of the added dropout layers since it mistook gliomas for the other classifications.

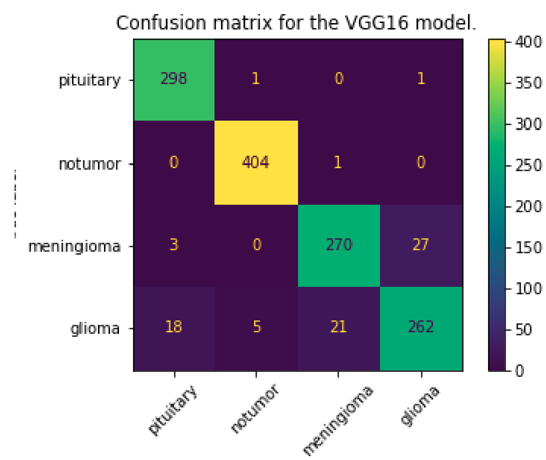

Fig. 4. Confusion matrix for VGG16 model

VGG16

We see that for VGG16, there was an improvement over the CNN with dropout model, but not by much. Instead of having most of its misclassifications in one area, this model tended to over predict the presence of other tumors when it fact it was gliomas. In a way, this is not preferred over the other model as misclassifying brain tumors from across the board shows that the model did not train as well.

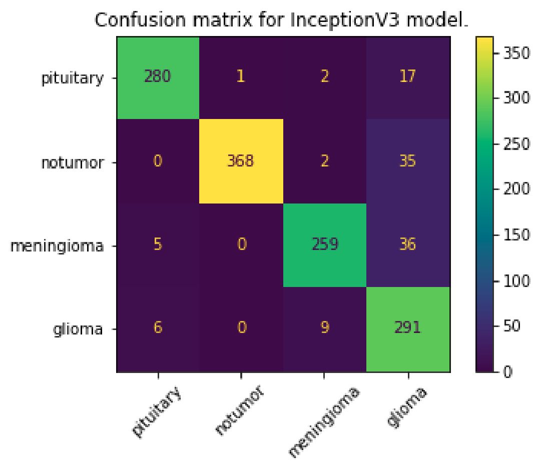

Fig. 5. Confusion matrix for InceptionV3 model

InceptionV3

For the InceptionV3 model we see it had an accuracy 91% and a weighted F1 score of 92%. These are very good results, but upon seeing the confusion matrix, we get a better picture of its faults. This model is very similar in performance to the CNN with dropout model, but to a much larger degree. It was over predicting gliomas and mistaking them for all the other brain tumors.

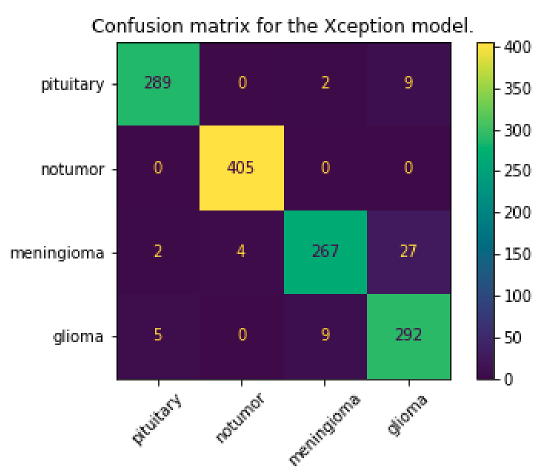

Fig. 6. Confusion matrix for the Xception model

Xception

The Xception model performed very nicely with a 96% accuracy and just as equally nice F1 score of 96%. This is evident by the confusion matrix as it performed similar to the best model, the CNN model. The distinctions, however, are the CNN model being more uniform in its errors. This Xception model had tiny errors all across the board.

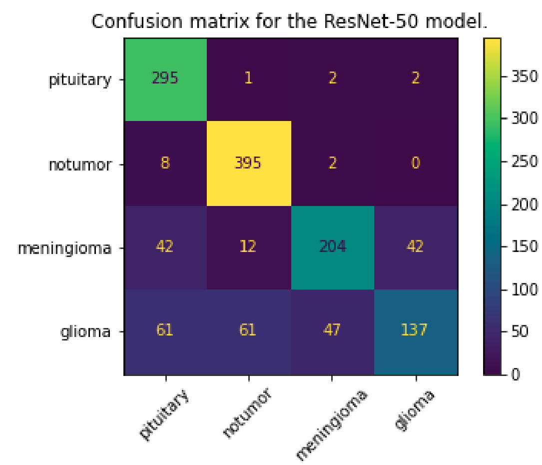

Fig. 7. Confusion matrix for the ResNet50 model

ResNet50

ResNet50's performance was the worst out of the CNN/pretrained models because of issues with the EarlyStopping function. Around 88% training accuracy it did not improve in accuracy after two consecutive epochs, so the model stopped training. This explains the issues at classifying brain tumors as shown in the confusion matrix.

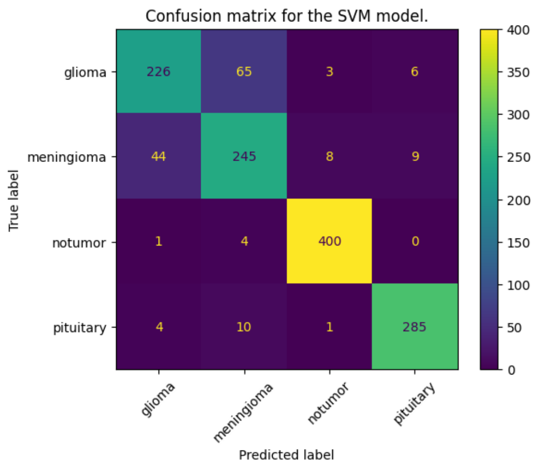

Fig. 8. Confusion matrix for the SVM model

SVM: Support Vector Model

Although it was originally expected for the SVM model to perform poorly, it did provide great results for identifying MRI scans with no tumor and pituitary cancer tumors. The Glioma and Meningioma tumors proved to be hard to detect, and the lack of hypertuning for the SVM pronounced this factor further on the resulting accuracy of the model:

Class	Recall	Accuracy
glioma	0.75	0.79
meningioma	0.80	0.78
no-tumor	0.99	0.98
pituitary	0.95	0.95
TOTAL	0.87	**0.88**

As the SVM works by creating a hyperplane to classify the data, it can be concluded that the glioma and meningioma tumors hold great similarity visually and have similar features. Consequently, the dividing line for the used data set is too narrow for these specific tumors. Changing the regularization parameters for the model would most likely improve the accuracy, but the similarity of the results to the CNN model and RF classifier suggests that a future study for classifying glioma and meningioma specifically would prove to be more accurate and effective.

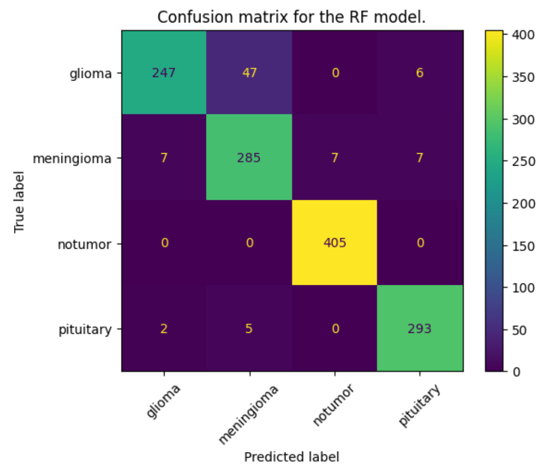

Fig. 9. Confusion matrix for the RF model

Random Forest Classifier

Upon examining the Random Forest Classifier's performance, as evidenced by the provided classification report and confusion matrix, we observe a strong predictive capability. The model exhibits a commendable overall accuracy of 94%, with a macro average F1-score of 0.93. This indicates a balanced precision-recall trade-off across the board, suggesting the model's adeptness at correctly identifying the various classes of brain tumors with a reasonable balance between false positives and false negatives. However, from the confusion matrix, we note certain areas where the model could be refined. It shows particular strength in identifying class 2 tumors, with perfect recall. In contrast, class 1 tumors have lower precision, which could indicate a propensity for this class to be confused with others. Such insights are crucial for further tuning the model to improve its diagnostic precision, especially in distinguishing between tumor types that share subtle imaging characteristics. The Random Forest Classifier's performance, with its inherent robustness and handling of feature correlations, underlines its potential as a reliable classifier in the multiclass setting of brain tumor MRI analysis.

CONCLUSIONS

In this comparative study of CNN and machine learning models for MRI brain tumor image classification, diverse array of models were meticulously evaluated, including CNNs with dropout layers, pre-trained models like VGG16, InceptionV3, Xception, ResNet50, and traditional machine learning algorithms such as Random Forest Classifier (RFC) and Support Vector Machines (SVM). CNN models have demonstrated remarkable accuracy, with the best-performing CNN achieving a high accuracy of 97% and an F1 score to match. The RFC also showed strong performance, with an overall accuracy of 94% and a macro average F1-score of 0.93. while the SVM results are pending for a complete comparative analysis. Adding to this, the SVM model registered a solid accuracy of 88% with a macro average F1-score of 0.87, showcasing its efficacy in classifying brain tumors with a high degree of precision, particularly in identifying non-tumor instances with 99% recall. While this performance can be improved, these results signify the models' proficiency in correctly identifying tumor classes, minimizing false positives and negatives. Despite the complex nature of brain tumor classification and the inherent risk of overfitting, our approach has yielded promising outcomes. The integration of these models can potentially revolutionize the diagnostic process, enhancing early detection and providing doctors with a powerful tool to improve patient care and prognosis.

CONTRIBUTIONS

Alexander Hernandez: Alexander was responsible for the implementation of the CNN, CNN with Dropout, VGG16, InceptionV3, Xception and ResNet50 models. He led the effort to resize the size of the images down to 150x150 in order for the team to be able to run more models and have more comparisons available to study. He was responsible for working on all of the confusion matrices displayed. Additionally, he worked on the conclusions drawn from the models mentioned beforehand.

Elie Bonkoungou: Elie was instrumental in developing the Random Forest models and contextualizing the study within the current scientific literature. His research provided a background that guided the study's direction. Elie analyzed related works and played a role in driving the study's conclusions highlighting the potential impact of our findings in medical diagnostics.

Gabriel Giron: Gabriel was responsible for the implementation of the SVM model and providing computing power for local training of the machine learning models of the study. He managed and combined the code for each model into a single file and originally helped Elie with the KNN model that was discarded for the project due to its low efficiency. He also helped with formatting and selecting discussion methods that allowed file sharing.

REFERENCES
[1] *Tests and procedures used to diagnose cancer*. National Cancer Institute. (n.d.). https://www.cancer.gov/about-cancer/diagnosis-staging/diagnosis
[2] Hoang, N. (2020, June 14). *Full Review on Optimizing Neural Network training with optimizer*. Medium. https://towardsdatascience.com/full-review-on-optimizing-neural-network-training-with-optimizer-9c1acc4dbe78
[3] *What is the dropout layer?*. Data Basecamp. (2023, August 2). https://databasecamp.de/en/ml/dropout-layer-en#:~:text=The%20dropout%20layer%20is%20a,the%20network%20architecture%20at%20all.
[4]*Introduction to the keras tuner : Tensorflow Core*. TensorFlow. (n.d.). https://www.tensorflow.org/tutorials/keras/keras_tuner
[5] *Transfer learning and fine-tuning : Tensorflow Core*. TensorFlow. (n.d.-b). https://www.tensorflow.org/tutorials/images/transfer_learning
[6] Learning, G. (2021, September 23). *Everything you need to know about VGG16*. Medium. https://medium.com/@mygreatlearning/everything-you-need-to-know-about-vgg16-7315defb5918
[7] Google. (n.d.). *Advanced guide to inception V3 | cloud TPU | google cloud*. Google. https://cloud.google.com/tpu/docs/inception-v3-advanced#:~:text=Inception%20v3%20is%20an%20image,multiple%20researchers%20over%20the%20years.
[8] Tsang, S.-H. (2019, March 20). *Review: Xception - with depthwise separable convolution, better than inception-V3 (image...* Medium. https://towardsdatascience.com/review-xception-with-depthwise-separable-convolution-better-than-inception-v3-image-dc967dd42568
[9] Sapireddy, S. R. (2023, July 1). *ResNet-50: Introduction*. Medium. https://srsapireddy.medium.com/resnet-50-introduction-b5435fdba66f
[10] Scientific Reports volume 12, Article number: 1505 (2022). *Classification of brain tumours in MR images using deep spatiospatial models. https://www.nature.com/articles/s41598-022-05572-6*
[11] *Scikit-Learn.LinearSVC* (2023). *https://scikit-learn.org/stable/modules/generated/sklearn.svm.LinearSVC.html*