Development of Synergic Deep Learning model for Glioma Tumor Grading

Preprintt

Saransh Chopra
Cluster Innovation Centre
University of Delhi

Harshvir Sandhu
Cluster Innovation Centre
University of Delhi

Ishan Bansal
Cluster Innovation Centre
University of Delhi

Nirmal Yadav
Cluster Innovation Centre
University of Delhi
nirmaliitr25@gmail.com

Abstract

Computer-aided diagnosis using deep learning approaches has made tremendous improvements in medical imaging for automatically detecting tumor area, tumor type, and grade. These advancements, however, have been limited due to the fact that 1) medical images are often less in quantity, leading to overfitting, and 2) significant inter-class similarity and intra-class variation between the images. To tackle these issues, we propose a Synergic Deep Learning model [Zhang et al., 2018, 2019] with an AlexNet [Krizhevsky et al., 2012] backbone for the automatic grading of glioma tumors. The Synergic Deep Learning architecture enables two pre-trained models to mutually learn from each other, allowing them to perform better than vanilla pre-trained models. Our study uses 417 T1-weighted sagittal tumor Magnetic Resonance Imaging (MRI) slices obtained from the REMBRANDT [Scarpace et al., 2019] dataset. These 417 slices, obtained from 20 patients, are pre-processed and augmented before they are fed into the model, which then classifies the tumor into one of the three grades: oligodendroglioma, anaplastic glioma, and glioblastoma multiforme. The proposed architecture achieves a training accuracy of 98.36% and a testing accuracy of 92.85%. Finally, the proposed SDL model with AlexNet backbone outperforms popular pre-trained models in terms of testing accuracy, recall, specificity, sensitivity, and F1 score.

Keywords Synergic Deep Learning · Glioma Tumor Grading · REMBRANDT · AlexNet

1 Introduction

Glioma is the second most prevalent brain tumor in adults after meningiomas. The earliest documented accounts of gliomas were published by Berns in British scientific journals in 1800. Percival Bailey and Harvey Cushing established the foundation for the contemporary classification of gliomas in 1926 [Stoyanov and Dzhenkov, 2017]. As per a National Library of Medicine journal, there are six cases of gliomas diagnosed annually per 100,000 individuals in the United States [Mesfin and Al-Dhahir, 2020].

Gliomas are tumors that infiltrate the surrounding brain tissue diffusely. They are classified by the World Health Organization's malignancy scale into grades I to IV [Jiang and Uhrbom, 2012]. Pilocytic astrocytoma, the most common glioma in children, is a benign grade I lesion that has a slow proliferation rate and is relatively well-defined. Grade II tumors, such as astrocytoma, oligodendroglioma, and oligoastrocytoma, have a slow growth rate, are highly differentiated, and infiltrate the normal brain parenchyma diffusely, making them prone to malignant progression. Anaplastic astrocytoma, anaplastic oligodendroglioma, and anaplastic oligoastrocytoma are grade III tumors that exhibit higher cellular density and have atypia and mitotic cells. Glioblastoma and gliosarcoma are the most malignant grade IV tumors and are also the most

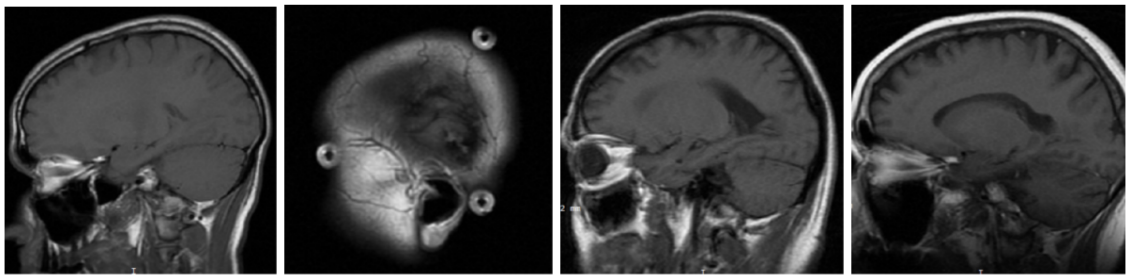


Figure 1: First 2 images (both grade 3 tumor) depicting intra-class variation and last 2 images (from left to right: grade 4 and grade 3) depicting inter-class similarity.

common gliomas. They exhibit microvascular proliferation and pseudo palisading necrosis in addition to grade III features..

Imaging is crucial for the diagnosis, surveillance, characterization, and therapeutic monitoring of intracranial tumors. The American Cancer Society recommends Magnetic resonance imaging (MRI) and computed tomography (CT) scans as the most commonly used techniques to detect brain tumors. For gliomas, MRI is particularly valuable, and conventional MRI protocols using T1-weighted, T2-weighted, and gadolinium-enhanced sequences play a significant clinical role. These protocols provide high-resolution structural information in multiple planes, enabling better tissue characterization compared to CT [Upadhyay and Waldman, 2011]. Medical imaging techniques have greatly benefited from the use of Artificial Intelligence (AI). In the field of diagnostic imaging, there has been a significant increase in the number of publications on AI, from approximately 100-150 per year in 2007-2008 to 1000-1100 per year in 2017-2018 [Tang, 2020]. Researchers have utilized AI to automatically identify complex patterns in imaging data and provide quantitative assessments of radiographic features. AI has also been applied in radiation oncology to various image modalities used in different stages of treatment, such as tumor delineation and treatment assessment. However, AI advancements in biomedical images have been limited due to two primary reasons. Firstly, medical images are often limited in quantity, leading to overfitting of models or high variance. Secondly, there is significant inter-class similarity and intra-class variation between images, making it difficult for the model to classify them accurately.

The Synergic Deep Learning model was first proposed for the task of Skin Lesion Classification in Dermoscopy Images. The proposed SDL model used ResNet50 [He et al., 2016] as its backbone and achieved an accuracy of 85.75%, average precision of 66.4%, and AUC of 82.6% on the ISIC 2016 [Gutman et al., 2016] skin classification dataset. The accuracy achieved by the model was greater than the accuracy of ResNet50, ResNet152 [He et al., 2016], and the other top 5 performing models on the leaderboard. The SDL model was even able to beat a joint segmentation-classification model, depicting how strong it was. The same authors presented another paper on the SDL model, and generalized the model even further to an SDLn model. Their study used the ImageCLEF-2015 [García Seco de Herrera et al., 2015], ImageCLEF-2016 [García Seco de Herrera et al., 2016], ISIC-2016, and ISIC-2017 [Codella et al., 2018] datasets and outperformed the existing state-of-the-art models, including ResNet50 and ResNet502. However, the paper emphasized the need to use Synergic Deep Learning Models for medical image classification and analysis. The limited quantity of datasets in medical image analysis, along with the significant intra-class variation and inter-class similarity, present an even greater challenge in classifying medical images. The Synergic Deep Learning (SDL) model allows for the simultaneous learning of multiple image pairs and utilizes multiple DCNN components without sharing parameters, enabling the model to benefit from an ensemble of multiple networks. The model can be trained end-to-end using classification errors from DCNNs and synergic errors from each pair of DCNNs. If one DCNN correctly classifies an image, the synergic error generated by the other DCNN serves as an additional force to update the model.

The SDL model has been employed for several diverse datasets, including a study on Diabetic Retinopathy [Kathiresan et al., 2020]. The paper focused on the classification of DR fundus images on the basis of severity level using a deep learning model. They coupled the SDL model with histogram-based segmentation (to extract useful regions from the image) on the MESSIDOR [Decencière et al., 2014] dataset and achieved an accuracy of 99.28%, a sensitivity of 98%, and a specificity of 99%. To ensure the goodness of the proposed model, they also performed a CT analysis, where their model required a minimum CT of 15.21 seconds to

classify the fundus images, followed by AlexNet [Krizhevsky et al., 2012] and then VGG19 [Simonyan and Zisserman, 2015] models.

Focussing particularly on grading glioma tumors using Deep Learning, a paper published in the AJNR [Gutta et al., 2021] trained a joint segmentation-classification pipeline using Convolutional Neural Networks and achieved an average accuracy of 87% on a dataset of 237 patients. Their model outperformed the methods considering radiomic features alone and also the highest performing model (gradient boosting) of that time. [Xiao et al., 2019] involved integrating radiomics features with high-level deep learning features to construct a more comprehensive representation of medical images. The features were extracted using a fine tuned VGG-16 model, and the BRATS 2018 [Menze et al., 2015, Bakas et al., 2017a, 2019, 2017b,c] dataset was used, which includes 285 subjects from multiple institutions. The extracted features were then utilized to train three classifiers, namely Logistic Regression (LR), Support Vector Machine (SVM), and Linear Discriminant Analysis (LDA). The best results were achieved by combining radiomics feature extraction with deep learning feature extraction. Recursive feature elimination (RFE) algorithm was used for feature selection. The proposed method achieved an average accuracy of 89.1% and an average AUC of 93.4%. Another study [Babu and Sourirajan, 2017] used the Rembrandt dataset for the detection of brain tumors, specifically for tumor grading. To recognize tumors from brain MRI images, the Tetrolet Transform (TT) [Krommweh, 2010] and Support Vector Machine (SVM) Classifier were employed. The Tetrolet transform was used to decompose the MRI brain tumor image at a predefined level, and the resulting image features were then classified using SVM. The fifth level of decomposition with SVM-based classification resulted in an accuracy level of 98.8%. Overall, the Tetrolet transform was found to be a useful technique for recognizing tumors in MRI brain images.

Even after the success of the SDL model, it has never been applied to tumor MRI slices, especially for the purpose of grading glioma tumors. We propose a novel Synergic Deep Learning model with an AlexNet backbone for classifying the grade of a glioma tumor. The pre-trained AlexNet models aim to prevent overfitting caused due to the low number of MRI slices and the overall Synergic Deep Learning architecture allows two AlexNets to mutually learn from each other. Each DCNN is trained separately and the extracted feature vector for an image is fed into the synergic layer, which predicts if the two images passed through two DCNNs belong to the same class. The error produced in the synergic layer is back propagated to the DCNNs, allowing them to learn from each other.

Moreover, this paper makes the following contributions -

- A Synergic Deep Learning model with an AlexNet backbone fine-tuned for T1-weighted sagittal tumor MRI slices.
- The best SDL model (at $\lambda = 3$) with a training accuracy of 98.36%, a testing accuracy of 92.86%, an average precision of 91.75%, average recall of 94.07%, average specificity of 96.19%, average sensitivity of 94.07%, and an average F1-score of 92.79%.
- Comparison with popular state-of-the-art pre-trained models - VGG19 [Simonyan and Zisserman, 2015], AlexNet [Krizhevsky et al., 2012], ResNet50 [He et al., 2016], and ResNet152 [He et al., 2016], with results showing that the proposed model outperforms the pre-trained models listed above for grading glioma tumors.
- A first-of-its-kind open-source implementation of the Synergic Deep Learning model.
- Discussions on the REMBRANDT dataset, stability interval of the synergic hyperparameter, and the use of Gaussian filters for preprocessing brain MRI slices.

2 Proposed work

2.1 Data pre-processing

From the dataset, we selected the T1-weighted sagittal MRI slices by iterating through the metadata of each DICOM image. This was done to address the fact that a large number of axial images contained no information about the tumor, but we could see the tumor in almost all the sagittal images. The relative paths of these sagittal images were stored in a CSV file, and the images were then moved to a folder named as their grade value (II, III, and IV).

The Synergic Deep Learning model requires a pair of images as its input with 3 labels: 1) the label of the first image, 2) the label of the second image, 3) the synergic label conveying if both the images belong to the

same class (1) or not (0). Images belonging to each grade were randomly shuffled and the first 50% of images belonging to a particular grade were paired with each other, giving them a synergic label of 1. The remaining 50% of the images were paired randomly with an image of another grade, giving them a synergic label of 0. A new CSV file was prepared to track the pair of images, their respective labels, and the synergic label. This split also resulted in a loss of images; the last set of images could not be paired up with an image belonging to a different grade, as all the remaining images had the same grading. Finally, the testing images were not paired because both the DCNNs were tested independently of each other. Given that the MRI slices have only 1 channel and the pre-trained models accept images with 3 channels, every image was stacked depthwise to make the resulting image have 3 channels where each channel held the same information. The DICOM images are represented as a pixel array of integer-type data; hence, the images were also converted into data types compatible with PyTorch [Paszke et al., 2019] and the pre-trained model.

After pairing images, the dataset was split in a ratio of 9:1 for the training and testing phase, with no validation phase due to the low number of images. Every MRI slice was resized to 128X128 to allow the model to converge faster. Furthermore, the images were augmented to avoid overfitting. The augmentation process included randomly rotating the images in the range of ± 10 degrees and applying Gaussian filters of kernel sizes 3x3, 5x5, and 7x7.

2.2 Synergic Deep Learning model

The Synergic Deep Learning model constitutes of 2 independent DCNNs and a synergic layer. The model accepts a pair of images as its input with 3 labels: 1) the label of the first image, 2) the label of the second image, 3) the synergic label conveying if both the images belong to the same class (1) or not (0). These 2 images are fed independently to the two DCNN components with their respective labels. During forward propagation, the feature vector of the pair of images is obtained from the second last layer of each DCNN and is concatenated for the synergic layer. This concatenated feature vector is fed to the synergic layer, which then predicts if the images belong to the same class or not. Figure 2 shows a thematic description of the Synergic Deep Learning model.

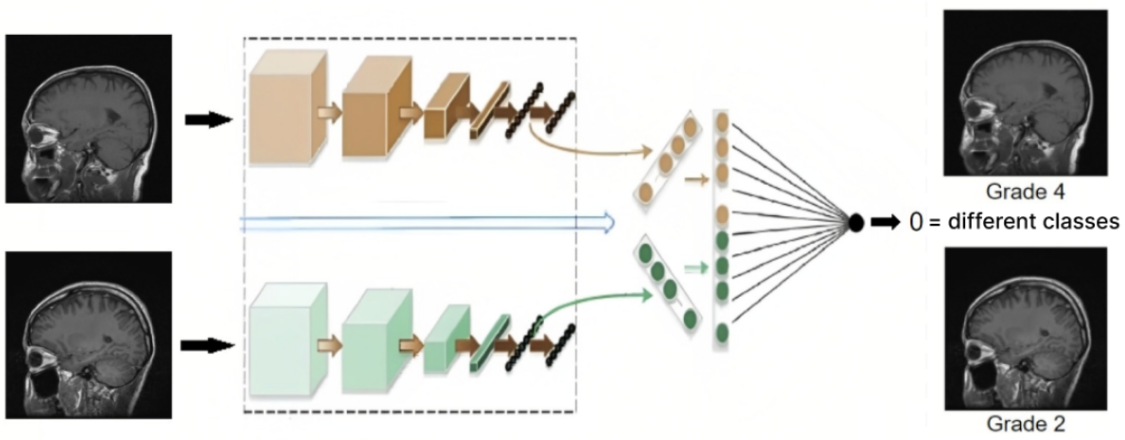


Figure 2: Diagrammatic representation of the proposed Synergic Deep Learning model.

For a data pair Z_A,Z_B with labels y_A,y_B where $A \neq B$, the synergic label y_s is defined as:

$$y_s = \begin{cases} 1 & y_a = y_b \\ 0 & y_a \neq y_b \end{cases}$$

We use 2 AlexNets as the DCNN components with their classification layer replaced with a single layer of neurons followed by the prediction layer of 3 neurons. We also define a single neuron layer as the synergic layer with 1 output neuron and sigmoid activation, to predict the synergic label. The synergic layer of the SDL model minimizes the following binary cross entropy loss:

$$Loss_{SDL} = y_s \log \hat{y} + (1 - y_s) \log (1 - \hat{y})$$

and the DCNN components are trained to minimize the following Cross Entropy (CE):

$$Loss_{DCNN} = \sum_{i=1}^{n} \sum_{c=1}^{3} y_{ic} \log p_{ic}$$

Where, n is the batch size, c is the number of classes, and p is the predicted probability of the DCNN component. Figure 3 shows the architecture of our data pipeline and model with the double headed arrows in the model section depicting both forward and back propagation.

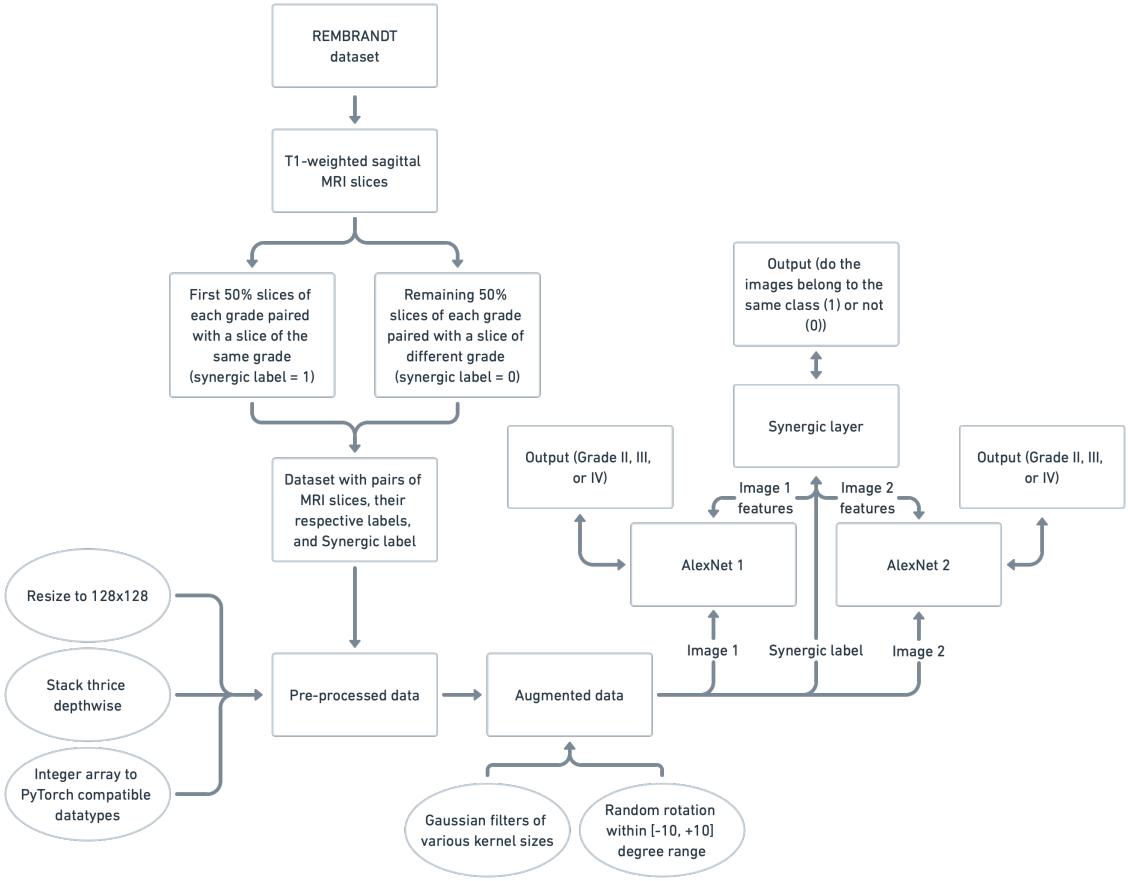


Figure 3: Architecture of our data pipeline and model.

2.2.1 Forward propagation

We define f_i as the features extracted by $DCNN_i$ from X_i where X_i is a batch of images, with labels y_i. P_1,P_2 are the class probability distributions predicted by $DCNN_1$, $DCNN_2$. Similarly P_{SDL} is the probability predicted by the synergic layer. The loss for each DCNN and the synergic layer is computed [Zhang et al., 2019] as follows:

$$f_{SDL} \leftarrow f_1 \cdot f_2$$

$$P_{SDL} \leftarrow SDL(f_{SDL})$$

$$loss_{DCNN_1} \leftarrow CE(P_1, y_1)$$

$$loss_{DCNN_2} \leftarrow CE(P_2, y_2)$$

$$loss_{SDL} \leftarrow BCE(P_{SDL}, y_{SDL})$$

2.2.2 Back propagation

We define θ_1, θ_2 as the parameters of $DCNN_1$, $DCNN_2$ respectively and θ_{SDL} as the parameters of the synergic layer. λ or the synergic hyperparameter indicates the factor by which the synergic error is backpropagated to the individual DCNNs and η represents the learning rate. The model parameters are updated [Zhang et al., 2019] as follows:

$$\theta_1 \leftarrow \theta_1 - \eta * (\frac{\partial loss_{DCNN_1}}{\partial \theta_1} + \lambda * \frac{\partial loss_{SDL}}{\partial \theta_{SDL}})$$

$$\theta_2 \leftarrow \theta_2 - \eta * (\frac{\partial loss_{DCNN_2}}{\partial \theta_2} + \lambda * \frac{\partial loss_{SDL}}{\partial \theta_{SDL}})$$

$$\theta_{SDL} \leftarrow \theta_{SDL} - \eta * \frac{\partial loss_{SDL}}{\partial \theta_{SDL}}$$

3 Experiments

3.1 Experimental setting

We ran our model for 250 epochs using the mini-batch Stochastic Gradient Descent optimizer. The learning rate was initially set to 0.0001 and a learning rate scheduler of the form -

$$\eta = \frac{\eta}{1 + (10^{-4} * epoch)}$$

was used. We recorded the individual accuracies and losses for each DCNN and the synergic layer. Before running the model, the MRI slices were resized to 128x128 and augmented with a random rotation within the range [-10, +10] degrees. Additionally, we also varied the synergic hyperparameter and recorded the maximum training and testing accuracies obtained from $\lambda = 3$ to $\lambda = 8$. The dataset was divided into testing and training with a ratio of 9:1 and a batch size of 8 was used. We also experimented by adding a Gaussian filter to the SDL model and varied the size of the Gaussian kernel (from 3x3 to 7x7). The other pre-trained models were independently fine-tuned and trained on the same dataset using the mini-batch SGD optimizer [Li et al., 2014].

3.2 Dataset

This paper uses the freely (with restricted license) available REMBRANDT dataset obtained through The Cancer Imaging Archive (TCIA) [Clark et al., 2013]. The dataset houses MRI slices of 127 patients stored in the DICOM format and the typical folder structure used for storing DCM MRI slices. Along with several other variables, the dataset includes glioma grading information for 89 patients, which is described in table 1a.

From this dataset, we selected all the T1-weighted sagittal MRI slices (by iterating through the metadata of every slice) for our study which were fed to the model as 2D images. The selected 417 slices are described in table 1b.

Grade	Patients
2	41
3	25
4	23
Total	**89**

(a) Original dataset.

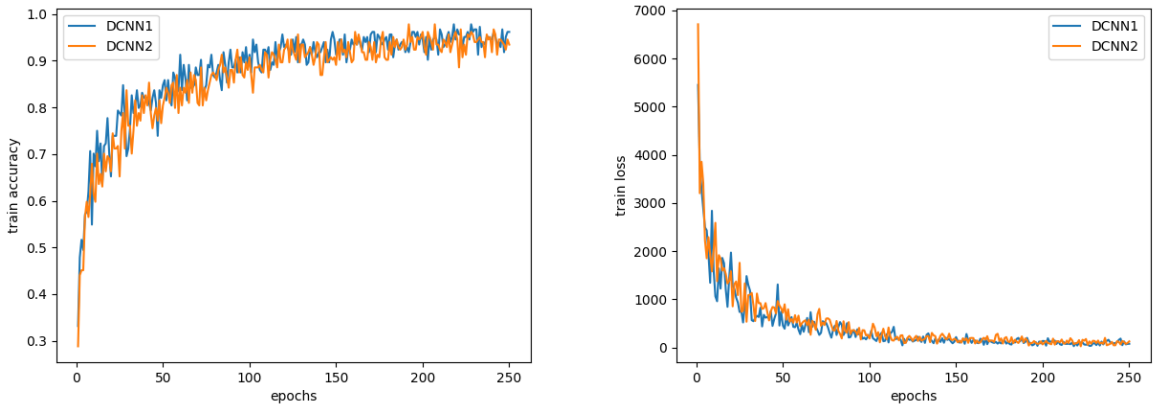
Grade	Patients	Slices
2	7	128
3	6	108
4	7	181
Total	**20**	**417**

(b) Selected dataset.

Table 1: Original REMBRANDT dataset and selected dataset.

3.3 Results

The model achieved the highest testing accuracy with the synergic hyperparameter λ set to 3 and the highest training accuracy with the synergic hyperparameter λ set to 7. Overall, the best model ($\lambda = 3$) achieved a training accuracy of 98.36% and a testing accuracy of 92.85%. Figure 4a shows the training accuracies of both the DCNNs from the SDL model with $\lambda = 3$. It can be observed that the accuracies of both the DCNNs converge with one of them attaining the value of 98.36%. Figure 4b shows the loss values for the same training phase, and it can be observed that both the loss functions converge as well. Furthermore, figure 5 shows the loss and accuracy of the synergic layer, the layer that was responsible for classifying the pair of images passed to the DCNNs as 1 (same class) or 0 (different class). It can be observed that the synergic layer quickly attains an accuracy of 100% and the loss function first fluctuates but soon converges exactly to 0.



(a) Training accuracy vs epochs for two DCNNs.

(b) Training loss vs epoch for two DCNNs.

Figure 4: Training phase of the DCNNs.

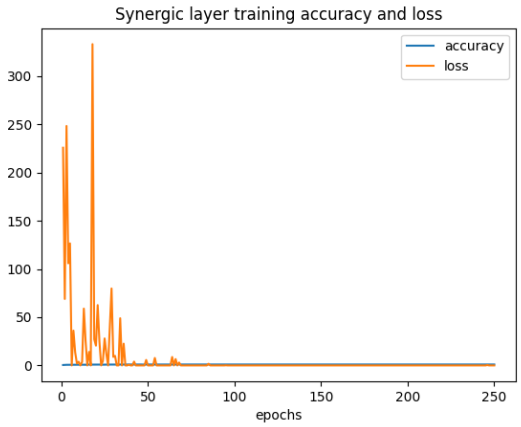
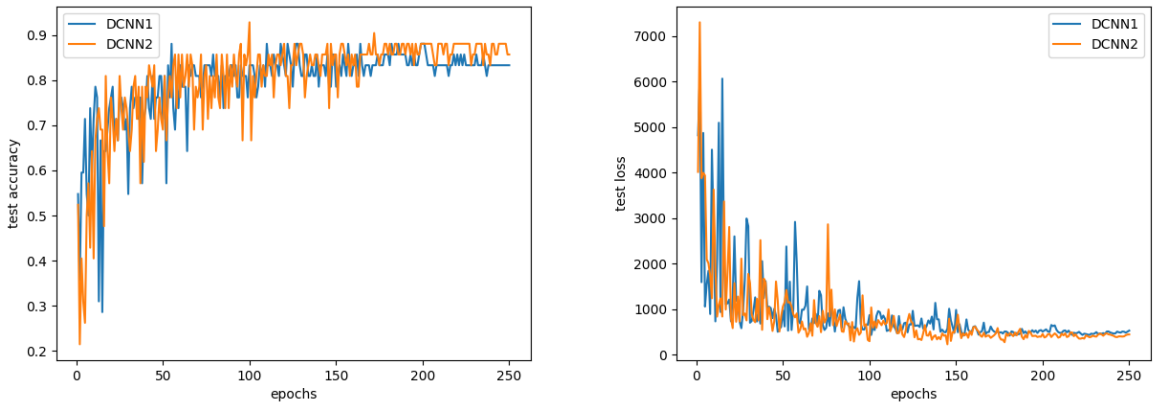


Figure 5: Training phase of the synergic layer.

The testing accuracies and losses of the DCNN components of the SDL model fluctuate but converges at the same time, as shown in figure 6a and figure 6b. The DCNN components from the best SDL model achieved a testing accuracy of 92.85%. It can also be seen the that the best testing accuracy was not achieved on the last epoch, rather it was achieved in an intermediate epoch.

Further, we varied the value of the synergic hyperparameter λ from 3 to 8 and recorded the highest testing and training accuracies achieved by the SDL model. We observed (figure 7a) that as λ increases, the training accuracy varies only by a small factor, whereas the variation in the testing accuracy is noticeable. The SDL

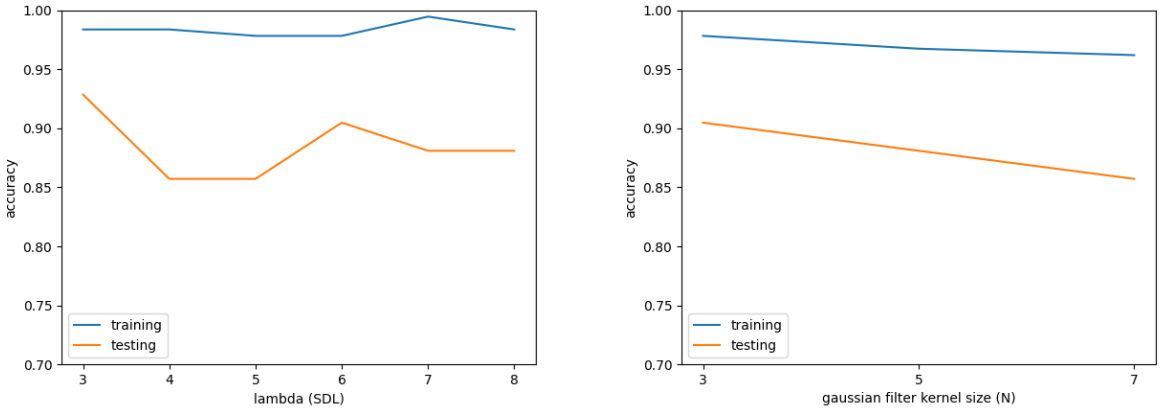


(a) Testing accuracy vs epochs for two DCNNs. (b) Testing loss vs epoch for two DCNNs.

Figure 6: Training phase of the SDL model.

model achieved the highest training accuracy of 99.45% at $\lambda = 7$ and the highest testing accuracy of 92.85% at $\lambda = 3$; hence, we consider the SDL model at $\lambda = 3$ to be the best model throughout the results.

Lastly, we added an extra data pre-processing step and applied a Gaussian filter to the incoming brain MRI scans. Figure 7b shows how adding a Gaussian filter deteriorates both the training and the testing accuracy when λ is kept constant (3). We experimented with Gaussian Filters of kernel sizes 3x3, 5x5, and 7x7, and the accuracy went down as the kernel size increased. The accuracy also saw a drop from just adding a Gaussian filter; hence, the highest accuracy achieved by the SDL model did not include the use of a Gaussian Filter.



(a) Accuracy vs lambda for SDL model. (b) Accuracy vs Gaussian filter size for SDL model.

Figure 7: Varying synergic parameter and Gaussian filter size.

Finally, table 2 compares the SDL model with other pre-trained models and models available in the literature. Further, table 3 lists down the defined metrics for each grade of the tumor. We define precision, recall, sensitivity, specificity, and F1 score as following:

$$Precision = \frac{True\ positives}{True\ positives + False\ positives}$$

$$Recall = \frac{True\ positive}{True\ positive + False\ negatives}$$

$$Specificity = \frac{True\ negatives}{True\ negatives + False\ positives}$$

$$Sensitivity = \frac{True\ positive}{True\ positives + False\ negatives}$$

$$F1\ score = \frac{2 * precision * recall}{precision + recall}$$

Table 2: Comparing the proposed SDL model with popular pre-trained models.

Model	Training accuracy	Testing accuracy	Precision	Recall	Specificity	Sensitivity	F1 score
Proposed SDL model	98.36	**92.86**	91.75	**94.07**	**96.19**	**94.07**	**92.79**
VGG19 [Simonyan and Zisserman, 2015]	**100**	85.71	85.19	81.11	92.66	81.11	81.99
AlexNet [Krizhevsky et al., 2012]	**100**	80.95	81.84	78.89	89.88	78.89	79.86
ResNet50 [He et al., 2016]	**100**	89.88	**92.92**	87.04	95.83	87.04	88.75
ResNet152 [He et al., 2016]	**100**	90.48	91.11	89.44	95.99	89.44	89.76

Table 3: Classwise classification report.

Grade	Accuracy	Precision	Recall	Specificity	Sensitivity	F1 score
II	90.90	90.91	90.91	96.77	90.91	90.91
III	**100**	88.89	**100**	**97.06**	**100**	**94.12**
IV	91.30	**95.45**	91.30	94.74	91.30	93.33

Figure 8 shows a diagrammatic representation of 6 pairs of MRI slices, their respective input target labels, the input synergic label, and the predictions given by each DCNN component of the Synergic Deep Learning model. The output of the 2 DCNNs has been represented as a tuple - ($DCNN_{1_{output}}$, $DCNN_{2_{output}}$).

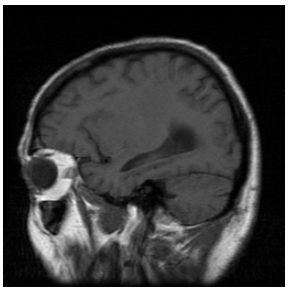
3.4 Hyperparameters and platform

The model was trained for 250 epochs with the data divided into a batch size of 8. We used an initial learning rate of 0.0001 with a learning rate scheduler of the form -

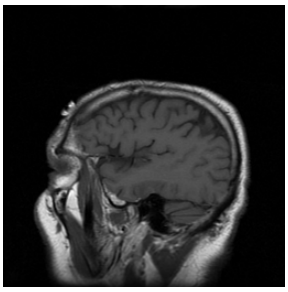
$$\eta = \frac{\eta}{1 + (10^{-4} * epoch)}$$

for all the subparts of our model. Finally, we varied the synergic hyperparameter λ from 3 to 8 giving us the best results on 3.

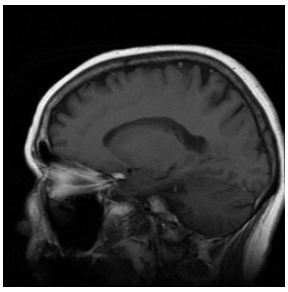
The model was written completely in Python [van Rossum and de Boer, 1991] using PyTorch. The model was trained and tested on a machine with 8GB RAM, Intel-i5-10300H CPU @ 2.5 GHz, and a dedicated NVIDIA GTX 1650ti GPU. Training of the model took 1281 seconds for 184 pairs of images, making each epoch last for an average of 5.124 seconds.



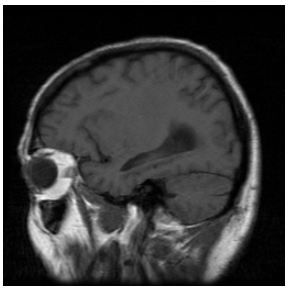
(a1) Label = 4, predictions = (4, 2).



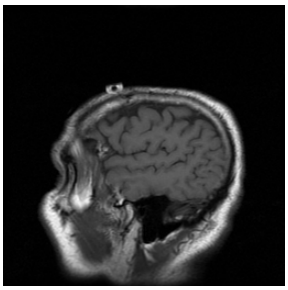
(b1) Label = 2, predictions = (2, 2).



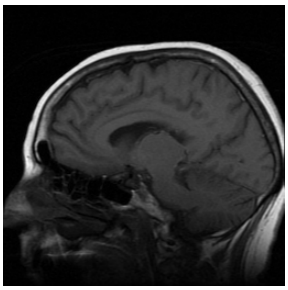
(c1) Label = 3, predictions = (3, 3).



(a2) Label = 4, predictions = (4, 4).

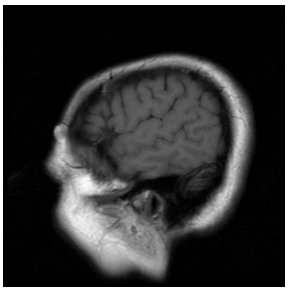


(b2) Label = 2, predictions = (2, 2).

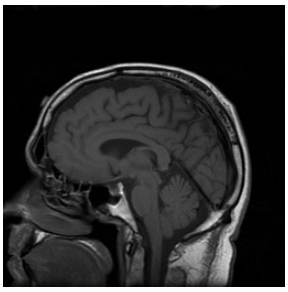


(c2) Label = 3, predictions = (3, 3).

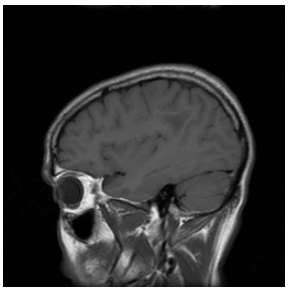
(a) Synergic label = 1 (b) Synergic label = 1 (c) Synergic label = 1



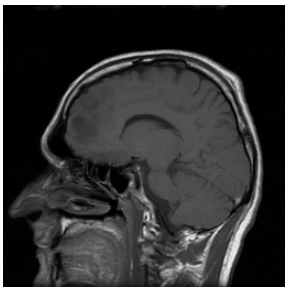
(d1) Label = 2, predictions = (2, 4).



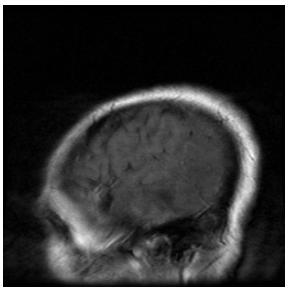
(e1) Label = 2, predictions = (2, 2).



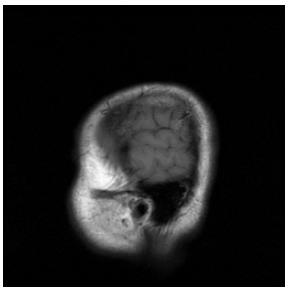
(f1) Label = 3, predictions = (3, 3).



(d2) Label = 3, predictions = (3, 3).



(e2) Label = 4, predictions = (4, 4).



(f2) Label = 4, predictions = (2, 4).

(d) Synergic label = 0 (e) Synergic label = 0 (f) Synergic label = 0

Figure 8: Input MRI slices, target labels, and predictions by the 2 DCNNs.

4 Discussion

4.1 Selecting only sagittal MRI slices

The REMBRANDT dataset includes both axial and sagittal scans, but a good number of axial slices held no information about the tumor. Additionally, a good percentage of the top and the bottom slices of every axial scan were blank and the only slices that held some information about the tumor were the ones stacked in the middle of the scan. On the other hand, almost all of the sagittal images depicted the tumor well with a very less number of outliers. To avoid feeding noisy and unnecessary MRI slices, we selected only the T1-weighted sagittal MRI slices for the SDL model.

4.2 Selecting AlexNet as the Backbone

We chose AlexNet over the ResNet models for the backbone of the SDL model to avoid overfitting. The ResNet models are comparatively bigger and have a more complicated architecture than the AlexNet model, which led to the SDL model having a high variance while training. Other smaller models like VGG19 led to the SDL model having a high bias; hence, we decided to use the AlexNet model as the SDL backbone, minimizing the bias-variance tradeoff.

4.3 Varying the synergic hyperparameter λ

We varied the value of the synergic hyperparameter λ within the stability interval described in [Zhang et al., 2019]. The testing accuracy of the SDL model went down as we increased the value of λ, matching the findings reported by [Zhang et al., 2019]. The best training accuracy was achieved at $\lambda = 7$, but $\lambda = 3$ had the best testing accuracy. Thus, we considered the model at $\lambda = 3$ as the best SDL model.

4.4 Adding and varying the size of Gaussian Filters

Inspired by the existing literature, we experimented with Gaussian filters by training the best SDL model, with and without the filter. We observed that just adding a Gaussian filter brought down the accuracy of the model. Increasing the kernel size of the filter from 3x3 to 5x5 to 7x7 brings down the accuracy further, with the reduction being proportional to the size of the kernel.

4.5 AlexNet performs poorly independently but AlexNet within SDL gives good results

In our results, one can observe that an independently fine-tuned AlexNet performs poorly on the testing dataset with an accuracy of 78.57%, average precision of 77.15%, average recall of 74.44%, and average F1 score of 75.10%. But, the same pre-trained model when fine-tuned within the SDL architecture performs exceptionally well and gave good results on both the training and testing data. This can be credited to the fact that the SDL architecture enables 2 AlexNets to mutually learn from each other, allowing them to achieve better metrics than the ones achieved by fine-tuning a single AlexNet on the same dataset. This also acts as proof that the SDL model is effective and helps in boosting the performance of a pre-trained model.

5 Conclusion

This paper presented a Synergic Deep Learning model with an AlexNet backbone for automated grading (II, III, and IV) of glioma tumor MRI scans to overcome the 2 prevalent problems present in medical image classification models - 1. Overfitting, and 2. Significant intra-class variation and inter-class similarity. The SDL model constitutes of 2 DCNNs and a synergic layer that enables the 2 DCNNs to mutually learn from each other. The work used T1-weighted sagittal MRI scans available in the REMBRANDT dataset for training the SDL model, which was augmented using random rotations and Gaussian filters. The SDL model achieved the highest training accuracy of 99.45% at $\lambda = 7$ and the highest testing accuracy of 92.86% at $\lambda = 3$. Both of the mentioned accuracies were achieved with no Gaussian filter, and adding a Gaussian filter to the SDL model brought down both, the training and the testing accuracies. Additionally, increasing the kernel size of the Gaussian filter further brought down the accuracy of the SDL model. Further, the best SDL model (at $\lambda = 3$) gave us a training accuracy of 98.36%, a testing accuracy of 92.86%, an average precision of 91.75%, average recall of 94.07%, average specificity of 96.19%, average sensitivity of 94.07%, and an average F1-score of 92.79%. The proposed SDL model with AlexNet backbone outperformed VGG19, AlexNet, ResNet50, and ResNet152 in terms of testing accuracy, recall, specificity, sensitivity, and F1 score. Finally, the AlexNet

model independently performed poorly on the testing dataset (testing accuracy = 78.57%) but performed very well (testing accuracy = 92.85%) when used as the backbone of the SDL model.

6 Future work

This paper restricts the SDL model to include exactly 2 DCNNs in its architecture but [Zhang et al., 2019] explores the SDL^n model, where n is the number of DCNNs used in a single SDL model. Such an SDL model, where n DCNNs mutually learn from each other can be applied to the REMBRANDT dataset to observe the changes in the defined metrics. Further, The MRI slices can be segmented or cropped such that only the region of interest is visible before they are fed into the model to remove noisy and unnecessary details of the scan. Hence, a joint segmentation classification model can be trained with the classification model being the SDL model. This work can also be extended to a refined dataset of axial MRI scans, allowing the model to learn the features from different angles of an MRI scan. Finally, more sophisticated methods like Reinforcement Learning can be applied to automatically select the best value of the synergic hyperparameter for classifying the MRI scans.

References

Jianpeng Zhang, Yutong Xie, Qi Wu, and Yong Xia. Skin lesion classification in dermoscopy images using synergic deep learning. In Alejandro F. Frangi, Julia A. Schnabel, Christos Davatzikos, Carlos Alberola-López, and Gabor Fichtinger, editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, pages 12–20, Cham, 2018. Springer International Publishing. ISBN 978-3-030-00934-2.

Jianpeng Zhang, Yutong Xie, Qi Wu, and Yong Xia. Medical image classification using synergic deep learning. *Medical Image Analysis*, 54:10–19, 2019. ISSN 1361-8415. doi:https://doi.org/10.1016/j.media.2019.02.010. URL `https://www.sciencedirect.com/science/article/pii/S1361841518307552`.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C.J. Burges, L. Bottou, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012. URL `https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf`.

L. Scarpace, A. E. Flanders, R. Jain, T. Mikkelsen, and D. W. Andrews. Data from rembrandt [data set]. the cancer imaging archive, 2019.

George Stoyanov and Deyan Dzhenkov. On the concepts and history of glioblastoma multiforme-morphology, genetics and epigenetics. *Folia medica*, online:ahead of print, 07 2017. doi:10.1515/folmed-2017-0069.

Fassil Mesfin and Mohammed Al-Dhahir. *Gliomas*. 08 2020.

Yiwen Jiang and Lene Uhrbom. On the origin of glioma. *Upsala journal of medical sciences*, 117:113–21, 02 2012. doi:10.3109/03009734.2012.658976.

N Upadhyay and Adam Waldman. Conventional mri evaluation of gliomas. *The British journal of radiology*, 84 Spec No 2:S107–11, 12 2011. doi:10.1259/bjr/65711810.

Xiaoli Tang. The role of artificial intelligence in medical imaging research. *BJR|Open*, 2(1):20190031, 2020. doi:10.1259/bjro.20190031. URL `https://doi.org/10.1259/bjro.20190031`.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016. doi:10.1109/CVPR.2016.90.

David Gutman, Noel C. F. Codella, Emre Celebi, Brian Helba, Michael Marchetti, Nabin Mishra, and Allan Halpern. Skin lesion analysis toward melanoma detection: A challenge at the international symposium on biomedical imaging (isbi) 2016, hosted by the international skin imaging collaboration (isic), 2016.

Alba García Seco de Herrera, Henning Müller, and Stefano Bromuri. Overview of the ImageCLEF 2015 medical classification task. In *Working Notes of CLEF 2015 (Cross Language Evaluation Forum)*, September 2015.

Alba García Seco de Herrera, Roger Schaer, Stefano Bromuri, and Henning Müller. Overview of the ImageCLEF 2016 medical task. In *Working Notes of CLEF 2016 (Cross Language Evaluation Forum)*, September 2016.

Noel C. F. Codella, David Gutman, M. Emre Celebi, Brian Helba, Michael A. Marchetti, Stephen W. Dusza, Aadi Kalloo, Konstantinos Liopyris, Nabin Mishra, Harald Kittler, and Allan Halpern. Skin lesion analysis

toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic), 2018.

Shankar Kathiresan, Abdul Rahaman Wahab Sait, Deepak Gupta, Lakshmanaprabu S.K, Ashish Khanna, and Hari Mohan Pandey. Automated detection and classification of fundus diabetic retinopathy images using synergic deep learning model. *Pattern Recognition Letters (Elsevier)*, 133:210–216, 5 2020. doi:10.1259/bjr/65711810.

Etienne Decencière, Xiwei Zhang, Guy Cazuguel, Bruno Lay, Béatrice Cochener, Caroline Trone, Philippe Gain, Richard Ordonez, Pascale Massin, Ali Erginay, Béatrice Charton, and Jean-Claude Klein. Feedback on a publicly distributed image database: The messidor database. *Image Analysis Stereology*, 33(3):231–234, 2014. ISSN 1854-5165. doi:10.5566/ias.1155. URL `https://www.ias-iss.org/ojs/IAS/article/view/1155`.

Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *International Conference on Learning Representations*, 2015.

S. Gutta, J. Acharya, M.S. Shiroishi, D. Hwang, and K.S. Nayak. Improved glioma grading using deep convolutional neural networks. *American Journal of Neuroradiology*, 42(2):233–239, 2021. ISSN 0195-6108. doi:10.3174/ajnr.A6882. URL `https://www.ajnr.org/content/42/2/233`.

Taohui Xiao, Wenqing Hua, Cheng Li, and Shanshan Wang. Glioma grading prediction by exploring radiomics and deep learning features. In *Proceedings of the Third International Symposium on Image Computing and Digital Medicine*, ISICDM 2019, page 208–213, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450372626. doi:10.1145/3364836.3364877. URL `https://doi.org/10.1145/3364836.3364877`.

Bjoern H. Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, Roland Wiest, Levente Lanczi, Elizabeth Gerstner, Marc André Weber, Tal Arbel, Brian B. Avants, Nicholas Ayache, Patricia Buendia, D. Louis Collins, Nicolas Cordier, Jason J. Corso, Antonio Criminisi, Tilak Das, Hervé Delingette, Çağatay Demiralp, Christopher R. Durst, Michel Dojat, Senan Doyle, Joana Festa, Florence Forbes, Ezequiel Geremia, Ben Glocker, Polina Golland, Xiaotao Guo, Andac Hamamci, Khan M. Iftekharuddin, Raj Jena, Nigel M. John, Ender Konukoglu, Danial Lashkari, José António Mariz, Raphael Meier, Sérgio Pereira, Doina Precup, Stephen J. Price, Tammy Riklin Raviv, Syed M.S. Reza, Michael Ryan, Duygu Sarikaya, Lawrence Schwartz, Hoo Chang Shin, Jamie Shotton, Carlos A. Silva, Nuno Sousa, Nagesh K. Subbanna, Gabor Szekely, Thomas J. Taylor, Owen M. Thomas, Nicholas J. Tustison, Gozde Unal, Flor Vasseur, Max Wintermark, Dong Hye Ye, Liang Zhao, Binsheng Zhao, Darko Zikic, Marcel Prastawa, Mauricio Reyes, and Koen Van Leemput. The multimodal brain tumor image segmentation benchmark (brats). *IEEE Transactions on Medical Imaging*, 34(10):1993–2024, 10 2015. ISSN 0278-0062. doi:10.1109/TMI.2014.2377694.

Spyridon Bakas, Hamed Akbari, Aristeidis Sotiras, Michel Bilello, Martin Rozycki, Justin Kirby, John Freymann, Keyvan Farahani, and Christos Davatzikos. Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. *Scientific Data*, 4, 09 2017a. doi:10.1038/sdata.2017.117.

Spyridon Bakas, Mauricio Reyes, Andras Jakab, Stefan Bauer, Markus Rempfler, Alessandro Crimi, Russell Takeshi Shinohara, Christoph Berger, Sung Min Ha, Martin Rozycki, Marcel Prastawa, Esther Alberts, Jana Lipkova, John Freymann, Justin Kirby, Michel Bilello, Hassan Fathallah-Shaykh, Roland Wiest, Jan Kirschke, Benedikt Wiestler, Rivka Colen, Aikaterini Kotrotsou, Pamela Lamontagne, Daniel Marcus, Mikhail Milchenko, Arash Nazeri, Marc-Andre Weber, Abhishek Mahajan, Ujjwal Baid, Elizabeth Gerstner, Dongjin Kwon, Gagan Acharya, Manu Agarwal, Mahbubul Alam, Alberto Albiol, Antonio Albiol, Francisco J. Albiol, Varghese Alex, Nigel Allinson, Pedro H. A. Amorim, Abhijit Amrutkar, Ganesh Anand, Simon Andermatt, Tal Arbel, Pablo Arbelaez, Aaron Avery, Muneeza Azmat, Pranjal B., W Bai, Subhashis Banerjee, Bill Barth, Thomas Batchelder, Kayhan Batmanghelich, Enzo Battistella, Andrew Beers, Mikhail Belyaev, Martin Bendszus, Eze Benson, Jose Bernal, Halandur Nagaraja Bharath, George Biros, Sotirios Bisdas, James Brown, Mariano Cabezas, Shilei Cao, Jorge M. Cardoso, Eric N Carver, Adrià Casamitjana, Laura Silvana Castillo, Marcel Catà, Philippe Cattin, Albert Cerigues, Vinicius S. Chagas, Siddhartha Chandra, Yi-Ju Chang, Shiyu Chang, Ken Chang, Joseph Chazalon, Shengcong Chen, Wei Chen, Jefferson W Chen, Zhaolin Chen, Kun Cheng, Ahana Roy Choudhury, Roger Chylla, Albert Clérigues, Steven Colleman, Ramiro German Rodriguez Colmeiro, Marc Combalia, Anthony Costa, Xiaomeng Cui, Zhenzhen Dai, Lutao Dai, Laura Alexandra Daza, Eric Deutsch, Changxing Ding, Chao Dong, Shidu Dong, Wojciech Dudzik, Zach Eaton-Rosen, Gary Egan, Guilherme Escudero, Théo Estienne, Richard Everson, Jonathan Fabrizio, Yong Fan, Longwei Fang, Xue Feng, Enzo Ferrante, Lucas Fidon, Martin Fischer, Andrew P. French, Naomi Fridman, Huan Fu, David Fuentes, Yaozong Gao, Evan Gates, David Gering,

Amir Gholami, Willi Gierke, Ben Glocker, Mingming Gong, Sandra González-Villá, T. Grosges, Yuanfang Guan, Sheng Guo, Sudeep Gupta, Woo-Sup Han, Il Song Han, Konstantin Harmuth, Huiguang He, Aura Hernández-Sabaté, Evelyn Herrmann, Naveen Himthani, Winston Hsu, Cheyu Hsu, Xiaojun Hu, Xiaobin Hu, Yan Hu, Yifan Hu, Rui Hua, Teng-Yi Huang, Weilin Huang, Sabine Van Huffel, Quan Huo, Vivek HV, Khan M. Iftekharuddin, Fabian Isensee, Mobarakol Islam, Aaron S. Jackson, Sachin R. Jambawalikar, Andrew Jesson, Weijian Jian, Peter Jin, V Jeya Maria Jose, Alain Jungo, B Kainz, Konstantinos Kamnitsas, Po-Yu Kao, Ayush Karnawat, Thomas Kellermeier, Adel Kermi, Kurt Keutzer, Mohamed Tarek Khadir, Mahendra Khened, Philipp Kickingereder, Geena Kim, Nik King, Haley Knapp, Urspeter Knecht, Lisa Kohli, Deren Kong, Xiangmao Kong, Simon Koppers, Avinash Kori, Ganapathy Krishnamurthi, Egor Krivov, Piyush Kumar, Kaisar Kushibar, Dmitrii Lachinov, Tryphon Lambrou, Joon Lee, Chengen Lee, Yuehchou Lee, M Lee, Szidonia Lefkovits, Laszlo Lefkovits, James Levitt, Tengfei Li, Hongwei Li, Wenqi Li, Hongyang Li, Xiaochuan Li, Yuexiang Li, Heng Li, Zhenye Li, Xiaoyu Li, Zeju Li, XiaoGang Li, Wenqi Li, Zheng-Shen Lin, Fengming Lin, Pietro Lio, Chang Liu, Boqiang Liu, Xiang Liu, Mingyuan Liu, Ju Liu, Luyan Liu, Xavier Llado, Marc Moreno Lopez, Pablo Ribalta Lorenzo, Zhentai Lu, Lin Luo, Zhigang Luo, Jun Ma, Kai Ma, Thomas Mackie, Anant Madabushi, Issam Mahmoudi, Klaus H. Maier-Hein, Pradipta Maji, CP Mammen, Andreas Mang, B. S. Manjunath, Michal Marcinkiewicz, S McDonagh, Stephen McKenna, Richard McKinley, Miriam Mehl, Sachin Mehta, Raghav Mehta, Raphael Meier, Christoph Meinel, Dorit Merhof, Craig Meyer, Robert Miller, Sushmita Mitra, Aliasgar Moiyadi, David Molina-Garcia, Miguel A. B. Monteiro, Grzegorz Mrukwa, Andriy Myronenko, Jakub Nalepa, Thuyen Ngo, Dong Nie, Holly Ning, Chen Niu, Nicholas K Nuechterlein, Eric Oermann, Arlindo Oliveira, Diego D. C. Oliveira, Arnau Oliver, Alexander F. I. Osman, Yu-Nian Ou, Sebastien Ourselin, Nikos Paragios, Moo Sung Park, Brad Paschke, J. Gregory Pauloski, Kamlesh Pawar, Nick Pawlowski, Linmin Pei, Suting Peng, Silvio M. Pereira, Julian Perez-Beteta, Victor M. Perez-Garcia, Simon Pezold, Bao Pham, Ashish Phophalia, Gemma Piella, G. N. Pillai, Marie Piraud, Maxim Pisov, Anmol Popli, Michael P. Pound, Reza Pourreza, Prateek Prasanna, Vesna Prkovska, Tony P. Pridmore, Santi Puch, Élodie Puybareau, Buyue Qian, Xu Qiao, Martin Rajchl, Swapnil Rane, Michael Rebsamen, Hongliang Ren, Xuhua Ren, Karthik Revanuru, Mina Rezaei, Oliver Rippel, Luis Carlos Rivera, Charlotte Robert, Bruce Rosen, Daniel Rueckert, Mohammed Safwan, Mostafa Salem, Joaquim Salvi, Irina Sanchez, Irina Sánchez, Heitor M. Santos, Emmett Sartor, Dawid Schellingerhout, Klaudius Scheufele, Matthew R. Scott, Artur A. Scussel, Sara Sedlar, Juan Pablo Serrano-Rubio, N. Jon Shah, Nameetha Shah, Mazhar Shaikh, B. Uma Shankar, Zeina Shboul, Haipeng Shen, Dinggang Shen, Linlin Shen, Haocheng Shen, Varun Shenoy, Feng Shi, Hyung Eun Shin, Hai Shu, Diana Sima, M Sinclair, Orjan Smedby, James M. Snyder, Mohammadreza Soltaninejad, Guidong Song, Mehul Soni, Jean Stawiaski, Shashank Subramanian, Li Sun, Roger Sun, Jiawei Sun, Kay Sun, Yu Sun, Guoxia Sun, Shuang Sun, Yannick R Suter, Laszlo Szilagyi, Sanjay Talbar, Dacheng Tao, Dacheng Tao, Zhongzhao Teng, Siddhesh Thakur, Meenakshi H Thakur, Sameer Tharakan, Pallavi Tiwari, Guillaume Tochon, Tuan Tran, Yuhsiang M. Tsai, Kuan-Lun Tseng, Tran Anh Tuan, Vadim Turlapov, Nicholas Tustison, Maria Vakalopoulou, Sergi Valverde, Rami Vanguri, Evgeny Vasiliev, Jonathan Ventura, Luis Vera, Tom Vercauteren, C. A. Verrastro, Lasitha Vidyaratne, Veronica Vilaplana, Ajeet Vivekanandan, Guotai Wang, Qian Wang, Chiatse J. Wang, Weichung Wang, Duo Wang, Ruixuan Wang, Yuanyuan Wang, Chunliang Wang, Guotai Wang, Ning Wen, Xin Wen, Leon Weninger, Wolfgang Wick, Shaocheng Wu, Qiang Wu, Yihong Wu, Yong Xia, Yanwu Xu, Xiaowen Xu, Peiyuan Xu, Tsai-Ling Yang, Xiaoping Yang, Hao-Yu Yang, Junlin Yang, Haojin Yang, Guang Yang, Hongdou Yao, Xujiong Ye, Changchang Yin, Brett Young-Moxon, Jinhua Yu, Xiangyu Yue, Songtao Zhang, Angela Zhang, Kun Zhang, Xuejie Zhang, Lichi Zhang, Xiaoyue Zhang, Yazhuo Zhang, Lei Zhang, Jianguo Zhang, Xiang Zhang, Tianhao Zhang, Sicheng Zhao, Yu Zhao, Xiaomei Zhao, Liang Zhao, Yefeng Zheng, Liming Zhong, Chenhong Zhou, Xiaobing Zhou, Fan Zhou, Hongtu Zhu, Jin Zhu, Ying Zhuge, Weiwei Zong, Jayashree Kalpathy-Cramer, Keyvan Farahani, Christos Davatzikos, Koen van Leemput, and Bjoern Menze. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the brats challenge, 2019.

S Bakas, H Akbari, A Sotiras, M Bilello, M Rozycki, J Kirby, J Freymann, K Farahani, and C Davatzikos. Segmentation labels and radiomic features for the pre-operative scans of the tcga-gbm collection. *The Cancer Imaging Archive*, 2017b. doi:10.7937/K9/TCIA.2017.KLXWJJ1Q.

S Bakas, H Akbari, A Sotiras, M Bilello, M Rozycki, J Kirby, J Freymann, K Farahani, and C Davatzikos. Segmentation labels and radiomic features for the pre-operative scans of the tcga-gbm collection. *The Cancer Imaging Archive*, 2017c. doi:10.7937/K9/TCIA.2017.GJQ7R0EF.

Bukke Babu and Varadarajan Sourirajan. Detection of brain tumour in mri scan images using tetrolet transform and svm classifier. *indian journal of science and technology*, 10:1–10, 05 2017. doi:10.17485/ijst/2017/v10i19/113721.

Jens Krommweh. Tetrolet transform: A new adaptive haar wavelet algorithm for sparse image representation. *Journal of Visual Communication and Image Representation*, 21:364–374, 05 2010. doi:10.1016/j.jvcir.2010.02.011.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019. URL `http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf`.

Mu Li, Tong Zhang, Yuqiang Chen, and Alexander J. Smola. Efficient mini-batch training for stochastic optimization. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '14, page 661–670, New York, NY, USA, 2014. Association for Computing Machinery. ISBN 9781450329569. doi:10.1145/2623330.2623612. URL `https://doi.org/10.1145/2623330.2623612`.

K Clark, B Vendt, K Smith, J Freymann, J Kirby, P Koppel, S Moore, S Phillips, D Maffitt, M Pringle, L Tarbox, and F Prior. The cancer imaging archive (tcia): Maintaining and operating a public information repository. *Journal of Digital Imaging*, 26:1045–1057, 2013. doi:https://doi.org/10.1007/s10278-013-9622-7.

G van Rossum and J de Boer. Interactively testing remote servers using the python programming language. *CWI Quarterly*, 4:283–304, 1991.