



A Combined Radio-Histological Approach for Classification of Low Grade Gliomas

Aditya Bagari¹, Ashish Kumar¹, Avinash Kori¹, Mahendra Khened¹,
and Ganapathy Krishnamurthi¹✉

Indian Institute of Technology Madras, Chennai 600036, India
gankrish@iitm.ac.in

Abstract. Deep learning based techniques have shown to be beneficial for automating various medical image tasks like segmentation of lesions and automation of disease diagnosis. In this work, we demonstrate the utility of deep learning and radiomics features for classification of low grade gliomas (LGG) into astrocytoma and oligodendroglioma. In this study the objective is to use whole-slide H&E stained images and Magnetic Resonance (MR) images of the brain to make a prediction about the class of the glioma. We treat both the pathology and radiology datasets separately for in-depth analysis and then combine the predictions made by the individual models to get the final class label for a patient. The pre-processing of the whole slide images involved region of interest detection, stain normalization and patch extraction. An autoencoder was trained to extract features from each patch and these features are then used to find anomaly patches among the entire set of patches for a single Whole Slide Image. These anomaly patches from all the training slides form the dataset for training the classification model. A deep neural network based classification model was used to classify individual patches among the two classes. For the radiology dataset based analysis, each MRI scan was fed into a pre-processing pipeline which involved skull-stripping, co-registration of MR sequences to T1c, re-sampling of MR volumes to isotropic voxels and segmentation of brain lesion. The lesions in the MR volumes were automatically segmented using a fully convolutional Neural Network (CNN) trained on BraTS-2018 segmentation challenge dataset. From the segmentation maps $64 \times 64 \times 64$ cube patches centered around the tumor were extracted from the T1 MR images for extraction of high level radiomic features. These features were then used to train a logistic regression classifier. After developing the two models, we used a confidence based prediction methodology to get the final class labels for each patient. This combined approach achieved a classification accuracy of 90% on the challenge test set ($n=20$). These results showcase the emerging role of deep learning and radiomics in analyzing whole-slide images and MR scans for lesion characterization.

All authors contributed equally.

© Springer Nature Switzerland AG 2019
A. Crimi et al. (Eds.): BrainLes 2018, LNCS 11383, pp. 416–427, 2019.
https://doi.org/10.1007/978-3-030-11723-8_42

1 Introduction

1.1 Medical Image Analysis for LGG Classification

Medical imaging techniques like Magnetic Resonance Imaging (MRI) is used for detection and assessment of the abnormalities inside the body. The non-invasive and non-ionizing property of the MRI make them suitable for oncology imaging studies such as brain tumors. In addition to MR images, the gold standard for tumor assessment and grading usually employ whole-slide imaging of tissue biopsy under a microscope for assessing at cellular level. The assessment of these medical images are mostly done by visual inspection of trained radiologist or a pathologist. However, manual inspection of vast amounts of data is usually error prone, time consuming and introduces inter-rater variability. Hence, several research communities in medical image analysis are continuously working on developing methods to automate tasks such as segmentation and quantification of medical images.

Gliomas are one of the leading cause of brain cancer and usually associated with poor prognosis and lesser survival rates. The gold standard for grading of gliomas is mostly based on the pathology reports got from tissue biopsies. In, this work we propose a methodology of classification of Gliomas into Astrocytomas and Oligodendrogliomas based on MR images of the brain using radiomic features localized to segmented tumor from the T1 MR image. We also propose a methodology of refining the lower-confidence predictions from model based on MR images by combining the model's predictions with whole-slide image analysis.

2 Datasets Used

The dataset comprised of Radiology and Histopathology slides from 30 different patients. The dataset was equally distributed for the two classes of tumor-namely Astrocytoma and Oligodendroglioma. The Radiology data for each patient consisted of FLAIR, T1, T1C, T2 MR sequences. FLAIR and T2 sequences were missing in four cases provided, hence they were not included in the training set as our segmentation model required all the 4 MR sequences. The pathology data consisted of single whole-slide images for each of the 30 patients.

3 Histopathology Approach

3.1 Preprocessing

The pathology dataset contained 30 wholeslide images each for a single patient. Each pathology slide for a given patient was a large scale image typically spanning across 10–50k pixels across each of the 2 dimensions. A single pathology slide was acquired in multiple scales of resolution. The whole slide image (WSI) contained large areas of white space irrelevant for the training process. The first

step in our pathology analysis was finding the Regions of Interest (RoI) in a WSI. The WSI was converted from RGB colorspace to HSV colorspace for better contrast enhancement. The lower and upper thresholds were applied on the pixel values to get binary masks from the WSI (Fig. 1).



Fig. 1. Finding the region of interest (RoI) from the Whole Slide Image

On the generated binary mask, morphological closing and opening operations were applied to fill small holes and remove scattered foreground pixels (of size less than 3×3) from the foreground pixels. From the processed binary masks, bounding boxes around all the discrete contours (each contour encompasses a connected region of tissue in the RoI) were obtained. The generated bounding boxes served as a blueprint for the patch extraction process.

From each of the WSIs, patches of size 224×224 were extracted from the entire RoI with a stride of $(224, 224)$ in both dimensions. To limit the number of patches from a slide, the maximum number of patches from a bounding box were limited to 2k.

We observed that the color intensity variation of WSI across different cases in the dataset was huge and hence stain normalization technique proposed by Reinhard et al. (2001) was employed to obtain uniform patches across multiple whole-slides. This method works by transferring an image from RGB space to a $l\alpha\beta$ color space where the correlation between the different color axes is minimal, hence transformation in color channels can be applied independently, without having any undesirable cross channel artifacts. The normalization technique ensured that the extracted patches from different WSIs had minimal variation in intensity (Figs. 2 and 3).

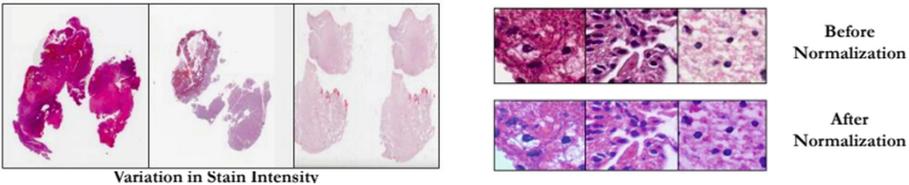


Fig. 2. Stain normalization for getting uniform colour intensity patterns across different slides

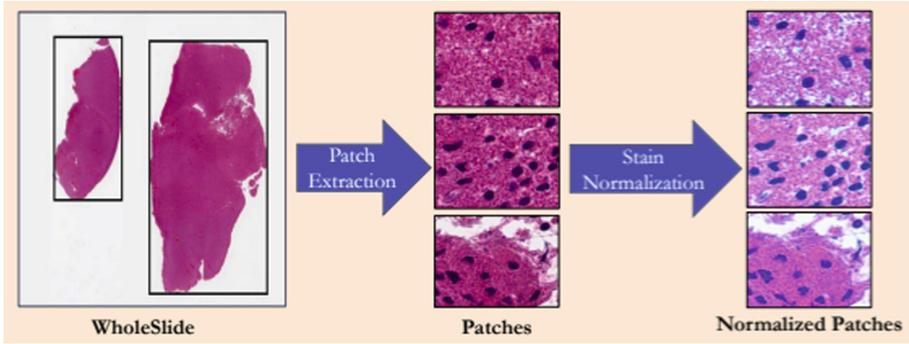


Fig. 3. Patch extraction from the Whole Slide Image followed by stain normalization

3.2 Feature Extraction and Anomaly Detection

Due to the presence of similar normal regions in the WSIs of the two classes, it was important to remove such patches which potentially came from regions that were non-tumour. To find out potential tumor patches, it was essential to obtain a good feature representation for each patch. Autoencoders are employed for extracting features from unlabelled data. We use an autoencoder based approach to extract features from each patch. We train an autoencoder using all the patches extracted from each of the slides. The autoencoder has five convolutional layers to downsample a patch followed by a single fully connected layer in the middle and again five convolutional layers to upsample the patch back to its original size. The central fully connected layer has 128 nodes giving a feature vector of size 128×1 for each patch. We use pixel wise reconstruction loss to train this autoencoder (Table 1).

After obtaining the feature vectors for each patch in a given Whole Slide Image, it was required to find a subset of these patches which can contain potential tumor regions. We treated the task of finding this subset as the problem of anomaly detection, where the tumor (anomaly) patches were required to be filtered out from the entire set of patches. We used the Isolation Forest Liu et al. (2008) technique to perform the task of anomaly detection. Isolation forest uses the feature vector representation of all patches from a WSI image and isolates anomaly patches based on these features in a two stage process. The first stage builds isolation trees (random decision trees) and in the second stage, test instances are passed through the trees to obtain an anomaly score for each instance based on the path length required to isolate an observation.

After extracting patches exhaustively from a WSI, we selected anomaly patches and used only the selected patches for further training. All these filtered patches from a WSI were assigned the same label as of the WSI. A total of 60k patches combined for both the classes were obtained after selecting these anomaly patches. We further used these patches for training a classification model for the two classes, Astrocytoma and Oligodendroglioma (Fig. 4).

Table 1. The table below describes the architecture used for the auto-encoder to extract features from each patch. Note: Op - Operation, MP - Max-pooling, US - Up-sampling.

| Layer | Op | Kernel | Stride | # Kernels | NonLin |
|-------|-----------|--------------|--------|-----------|--------|
| 1 | Conv + MP | 3×3 | (1, 1) | 8 | ReLU |
| 2 | Conv + MP | 3×3 | (1, 1) | 16 | ReLU |
| 3 | Conv + MP | 3×3 | (1, 1) | 32 | ReLU |
| 4 | Conv + MP | 3×3 | (1, 1) | 64 | ReLU |
| 5 | Conv + MP | 3×3 | (1, 1) | 128 | ReLU |
| 6 | FC | - | - | 128 | ReLU |
| 7 | US + Conv | 3×3 | (1, 1) | 64 | ReLU |
| 8 | US + Conv | 3×3 | (1, 1) | 32 | ReLU |
| 9 | US + Conv | 3×3 | (1, 1) | 16 | ReLU |
| 10 | US + Conv | 3×3 | (1, 1) | 8 | ReLU |
| 11 | US + Conv | 3×3 | (1, 1) | 3 | ReLU |

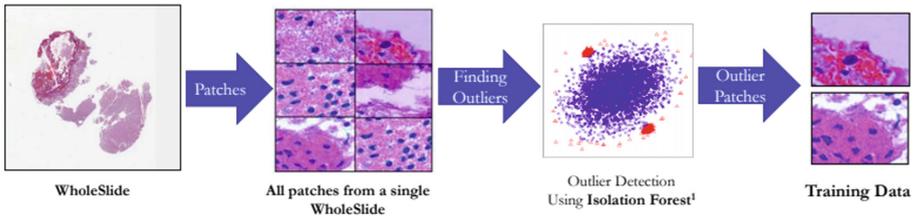


Fig. 4. Features were extracted using an autoencoder based approach and Isolation Forest was used to find anomaly patches from the entire set of patches for a single Whole Slide Image

3.3 Two-Class Classification

DenseNet-161 (Huang et al. (2017)) network was chosen to distinguish patches from Astrocytoma and Oligodendroglioma. The network was trained using binary cross entropy loss function (Fig. 5).

Testing: During testing, a stride based patch extraction was used to obtain all the patches in the region of interest, using a stride of (224, 224). From these patches, anomaly patches were found using Isolation Forest, in a way similar to the training phase. All the filtered patches from a particular whole-slide were fed to the densenet. After obtaining the prediction scores for all the patches from a WSI, a majority class voting was performed among all the predictions to obtain the class label for the slide. The class which had a higher frequency in the predictions was assigned as the class label for the slide.

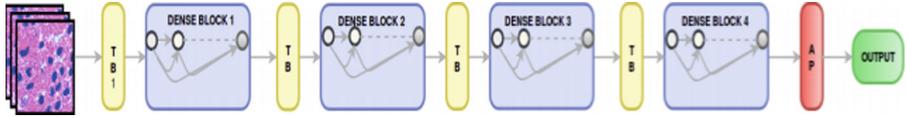


Fig. 5. A Densenet based network was used for the two-class classification task

4 Radiology

4.1 Understanding MR Images

Different pulse sequences of MR imaging system are used to enhance different parts of the tumor. For assessment of brain tumor the following pulse sequences are generally used (Fig. 6):

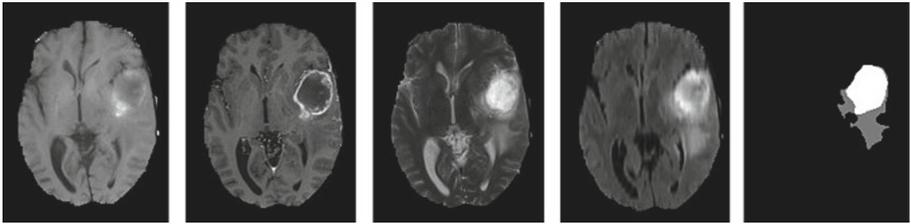


Fig. 6. From left to right: T1, T1c, T2, FLAIR, segmented tumor

1. **T1 weighted:** T1 image of tissues affected by brain tumors are of low signal intensity.
2. **T2 weighted:** T2 image of tissues affected by brain tumors are typically of high signal intensity. Calcifications due to tumor are mostly dark on T2.
3. **Fluid Attenuated Inversion Recovery (Flair):** Uses attenuation of intensity in Cerebro-spinal fluids (CSF) to differentiate between CSF and abnormalities. Generally gives the whole tumor region.
4. **T1-weighted post contrast imaging (T1c):** Gadolinium is used to enhance images and is useful in identifying vascular structures and breakdown in the blood-brain barrier typically found in the Necrotic region of the brain.

4.2 Pre Processing of MR Images

Magnetic resonance images were pre-processed to remove structures that could interfere with image segmentation (Fig. 7).

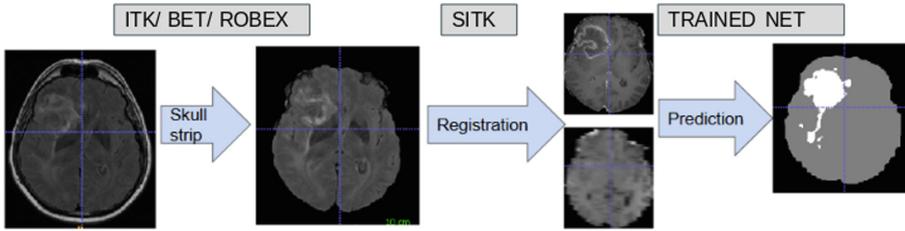


Fig. 7. MR image sequence pre-processing pipeline

1. Skull Stripping

It is necessary to remove the skull from the MRI as the enhancement arising due to its presence can be wrongly interpreted as tumor. Most segmentation networks are trained using skull stripped images as input and hence it is important to maintain this while providing new data. This was done using the ITK library Ibanez et al. (2005).

2. Co-registration and re-sampling to isotropic voxel spacing

Followed by skull stripping is the step of co-registering the MRI sequences to a reference sequence. Generally, there can be movement between scans if the patient does not remain still or if the scan is being done on a different day or while using a different machine. As a standard, we registered sequences T1, FLAIR and T2 with respect to T1c scan for all patients. After co-registration the MR volumes were re-sampled to 1 mm isotropic voxels across all dimensions.

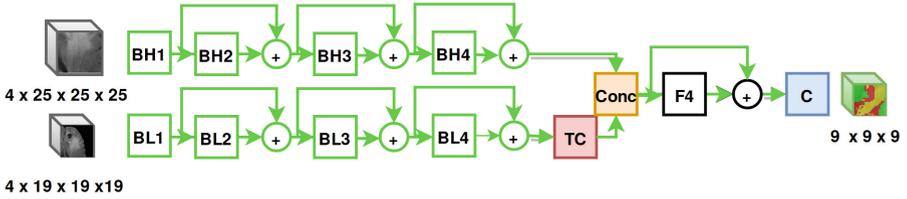
4.3 Segmentation of MR Images

Segmentation of the tumor in Magnetic resonance imaging (MRI) is the first step towards diagnosis. Features like the size, location of the tumor can dictate the stage of the tumor and the appropriate treatment. We trained a segmentation network with the following properties:

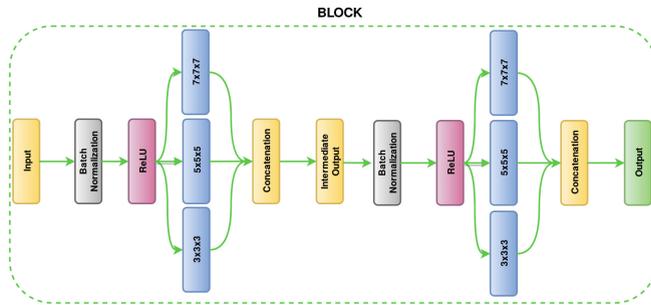
1. **Network architecture:** 3-D CNN as shown in the Fig. 8 was used for the task of semantic segmentation.
2. **Data:** Our network architecture was inspired from Kamnitsas et al. (2017) and was trained using the data provided by BraTS Menze et al. (2015) 2018 challenge. 25^3 and 51^3 (re-sized to 19^3) sized patches were extracted from all the four sequences (T1, T2, Flair and T1c) and the network was trained to predict center 9^3 .
3. **Training:** The weights of the network were initialized using Xavier initializer Glorot and Bengio (2010), and training was done using the weighted combination of weighted cross entropy and dice loss. Adam optimizer was used with initial learning rate of 0.001 and with the decay factor of 0.1.
4. **CNN based brain tumor segmentation:** The segmentation of the whole tumor region was done using a in-house fully convolution neural network trained on BraTS-2018 dataset (Menze et al. (2015)).

4.4 Radiomic Feature Extraction

The segmentations generated via the above-mentioned model were post processed to ensure to remove noise or any false positive region segmented. By applying connected component analysis, only the biggest tumor prediction was kept. As low grade gliomas are small in size, a patch of size $64 * 64 * 64$ of the segmentation and the corresponding T1 sequence image per case, centered around the predicted tumor region were extracted (Fig. 9).



(a)



(b)

Fig. 8. Semantic segmentation network for segmentation of gliomas from MR volumes. (a) The top portion of the network accepts high-resolution patches (25^3) while the bottom pathway accepts low-resolution input (51^3 patches resized to 19^3) as input. Both the high and low-resolution pathway is composed of inception modules so as to learn multi-resolution features. *TC* in the network stands for transposed convolution and is used to match the features of the spatial dimension of the low-resolution pathway with those learned in the high-resolution path. The BL and BH refer to building blocks for low and high resolution pathway. (b) The building block of the network. In the block, the dimension of the feature map in an inception module is maintained by setting the padding to 0, 1, 2 for 3×3 , 5×5 and 7×7 respectively.

As 3D images contain a lot of spatial and physical information, the 3D T1 image patches and the segmentation patches were used to extract features such as shape, texture, first order, second order and other higher order features using radiomics. The approach to extract high-through quantitative features relies on the radiomics platform provided by pyradiomics Griethuysen et al. (2017).

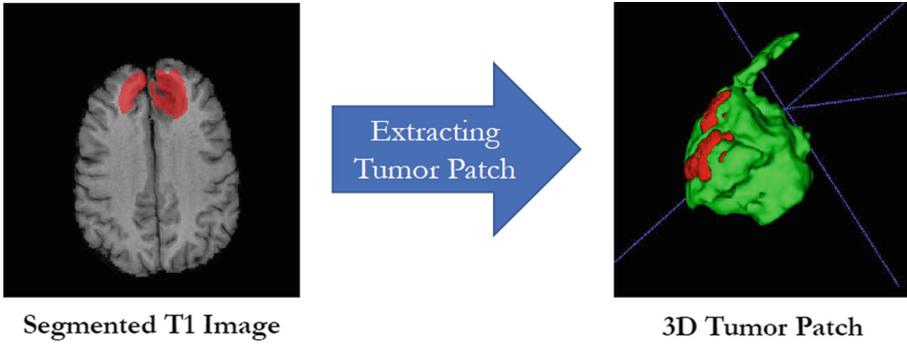


Fig. 9. 3D patch extraction from generated segmentation

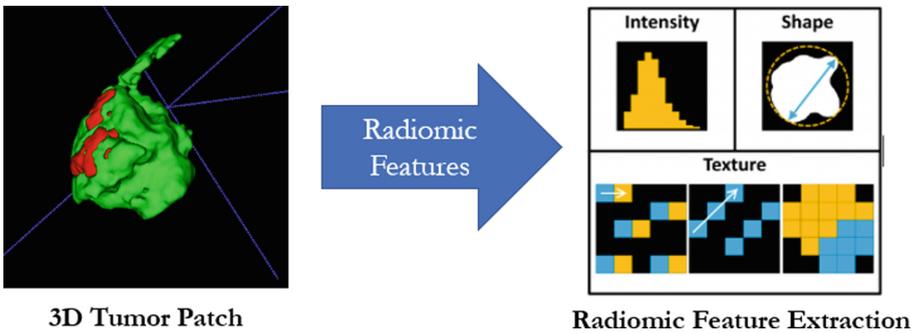


Fig. 10. Extracting radiomic features from the 3D T1 image patches www.radiomics.world (2018)

In total 105 radiomic features were extracted from the images which included 3 major kinds of radiomic features namely, Shape Features (13), First Order Features (18) and Texture Features (74). The complete list of extracted radiomic features are detailed below (Fig. 10).

1. **Shape Features (13):**

Elongation, Flatness, Least Axis, Major Axis, Maximum 2D Diameter Column, Maximum 2D Diameter Row, Maximum 2D Diameter Slice, Maximum 3D Diameter, Minor Axis, Sphericity, Surface Area, Surface Volume Ratio, Volume

2. **First order statistics (18):**

10 Percentile, 90 Percentile, Energy, Entropy, Interquartile Range, Kurtosis, Maximum, Mean, Mean Absolute Deviation, Median, Minimum, Range, Robust Mean Absolute Deviation, Root Mean Squared, Skewness, Standard Deviation, Total Energy, Uniformity, Variance

3. **GLCM (Gray Level Co-occurrence Matrix) (23):**

Auto-correlation, Cluster Prominence, Cluster Shade, Cluster Tendency,

Contrast, Correlation, Difference Average, Difference Entropy, Difference Variance, ID, IDM, IDMN, IDN, IMC1, IMC2, Inverse Variance, Joint Average, Joint Energy, Joint Entropy, Maximum Probability, Sum Average, Sum Entropy, Sum Squares

4. **GLRLM (Gray Level Run Length Matrix) (30):**

Dependence Entropy, Dependence Non-Uniformity, Dependence Non-Uniformity Normalized, Dependence Variance, Gray Level Non-Uniformity, Gray Level Variance, High Gray Level Emphasis, Large Dependence Emphasis, Large Dependence High Gray Level Emphasis, Large Dependence Low Gray Level Emphasis, Low Gray Level Emphasis, Small Dependence Emphasis, Small Dependence High Gray Level Emphasis, Small Dependence Low Gray Level Emphasis, Gray Level Non-Uniformity, Gray Level Non-Uniformity Normalized, Gray Level Variance, High Gray Level Run Emphasis, Long Run Emphasis, Long Run High Gray Level Emphasis, Long Run Low Gray Level Emphasis, Low Gray Level Run Emphasis, Run Entropy, Run Length Non-Uniformity, Run Length Non-Uniformity Normalized, Run Percentage, Run Variance, Short Run Emphasis, Short Run High Gray Level Emphasis, Short Run Low Gray Level Emphasis

5. **GLSZM (Gray Level Size Zero Matrix) (16):**

Gray Level Non-Uniformity, Gray Level Non-Uniformity Normalized, Gray Level Variance, High Gray Level Zone Emphasis, Large Area Emphasis, Large Area High Gray Level Emphasis, Large Area Low Gray Level Emphasis, Low Gray Level Zone Emphasis, Size Zone Non-Uniformity, Size Zone Non-Uniformity Normalized, Small Area Emphasis, Small Area High Gray Level Emphasis, Small Area Low Gray Level Emphasis, Zone Entropy, Zone Percentage, Zone Variance

6. **NGTDM (Neighbouring Gray Tone Difference Matrix) (5):**

Busyness, Coarseness, Complexity, Contrast, Strength

4.5 Training Methodology

MR Image based training methodology:

1. For a given MRI sequence, the 105-length deep feature vectors extracted from radiomics were reduced using Principal Component Analysis to a 16-length deep feature vector. This 16-length feature vector is then trained against the classification status present in the training data.
2. We took 27 training samples and generated (27, 16) shaped feature vectors and trained a logistic regression classifier with LIBLINEAR as the optimization algorithm on a 5-fold cross validation basis.
3. Prior to testing on the 20 test samples, the logistic regression model was fitted on the entire training data of 27 train samples. The features for the test set were extracted from the radiomic feature extraction system, normalized using the mean and standard deviation obtained during training and probabilistic classification predictions were obtained from the fitted logistic regression model.

5 Combining the Pathology and Radiology Predictions

For combining the predictions obtained by the pathology and radiology model, we use a simple higher confidence based voting criteria. Given the predictions from both the pathology based model and the radiology based model, we compare the probability values given by each of the model for a particular class and assign the final prediction based on the model which outputs the class label with a higher probability (Fig. 11).

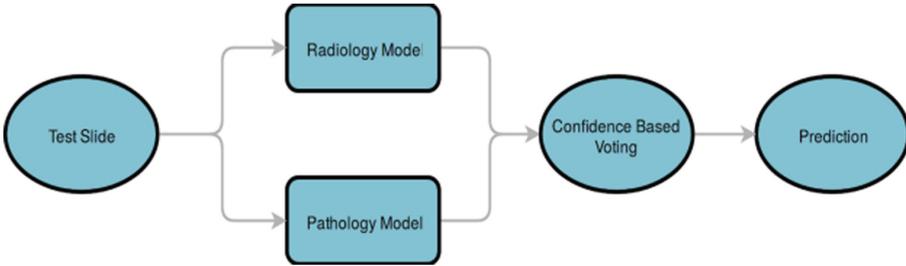


Fig. 11. Pipeline for combining the predictions from both the pathology and radiology model.

6 Results

Performance on Challenge Test Dataset

On testing the algorithms on a dataset containing 20 radiology and pathology images, an accuracy of 80% individually was observed both by the radiology and pathology models. The combined radiology and pathology model boosted the accuracy by 10% resulting in an accuracy of 90% on the entire dataset.

7 Conclusions

The results of our study show the feasibility of deep-learning based models for analyzing MR and whole-slide images. Our algorithm treats the histopathology and MR dataset separately and combines the individual predictions from both the models to output a single classification label. We demonstrated that anomaly detection based patch extraction can improve the classification results in Whole Slide Image based analysis. We also showed that radiomics based features aid in accurate classification of low-grade gliomas. One limitation of our work is that the MR based model requires all the 4 MR sequences of the patient for detecting the tumor region and even if a single sequence is missing the MR model fails to correctly narrow down on the tumor region. Another limitation could be the

high processing time required to generate predictions during inference stage due to the heavy preprocessing and postprocessing required for both the models.

Additional work needs to be done to explore potential ways to combine the two models on a feature level earlier in the classification pipeline. Another area of future work involves further investigation of anomaly detection based approaches for extracting relevant patches where the pixel wise annotations are not available. The model also needs to be tested on a large cohort for establishing generalization.

References

- Glorot, X., Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks. In: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, pp. 249–256 (2010)
- Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: CVPR, vol. 1, p. 3 (2017)
- Ibanez, L., Schroeder, W., Ng, L., Cates, J.: The ITK software guide (2005)
- Kamnitsas, K., et al.: Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med. Image Anal.* **36**, 61–78 (2017)
- Liu, F.T., Ting, K.M., Zhou, Z.-H.: Isolation forest. In: 2008 Eighth IEEE International Conference on Data Mining, pp. 413–422. IEEE (2008)
- Menze, B.H., et al.: The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Trans. Med. Imaging* **34**(10), 1993–2024 (2015)
- Reinhard, E., Adhikhmin, M., Gooch, B., Shirley, P.: Color transfer between images. *IEEE Comput. Graph. Appl.* **21**(5), 34–41 (2001)
- van Griethuysen, J.J., et al.: Computational radiomics system to decode the radiographic phenotype. *Cancer Res.* **77**(21), e104–e107 (2017)
- www.radiomics.world (2018)