



# Segmentation and Classification in Digital Pathology for Glioma Research: Challenges and Deep Learning Approaches

Tahsin Kurc<sup>1\*†</sup>, Spyridon Bakas<sup>2,3,4†</sup>, Xuhua Ren<sup>5</sup>, Aditya Bagari<sup>6</sup>, Alexandre Momeni<sup>7</sup>, Yue Huang<sup>8</sup>, Lichi Zhang<sup>5</sup>, Ashish Kumar<sup>6</sup>, Marc Thibault<sup>7</sup>, Qi Qi<sup>8</sup>, Qian Wang<sup>5</sup>, Avinash Kori<sup>6</sup>, Olivier Gevaert<sup>7</sup>, Yunlong Zhang<sup>8</sup>, Dinggang Shen<sup>9,10</sup>, Mahendra Khened<sup>6</sup>, Xinghao Ding<sup>8</sup>, Ganapathy Krishnamurthi<sup>6</sup>, Jayashree Kalpathy-Cramer<sup>11†</sup>, James Davis<sup>12†</sup>, Tianhao Zhao<sup>12†</sup>, Rajarsi Gupta<sup>1,12†</sup>, Joel Saltz<sup>1†</sup> and Keyvan Farahani<sup>13†</sup>

## OPEN ACCESS

### Edited by:

John Ashburner,  
University College London,  
United Kingdom

### Reviewed by:

Houman Sotoudeh,  
University of Alabama at Birmingham,  
United States  
Javad Noorbakhsh,  
Broad Institute, United States

### \*Correspondence:

Tahsin Kurc  
tahsin.kurc@stonybrook.edu

<sup>†</sup>These authors have contributed  
equally to this work and share first  
authorship

<sup>‡</sup>People involved in the organization of  
the challenge

### Specialty section:

This article was submitted to  
Brain Imaging Methods,  
a section of the journal  
Frontiers in Neuroscience

**Received:** 29 August 2019

**Accepted:** 10 January 2020

**Published:** 21 February 2020

### Citation:

Kurc T, Bakas S, Ren X, Bagari A,  
Momeni A, Huang Y, Zhang L,  
Kumar A, Thibault M, Qi Q, Wang Q,  
Kori A, Gevaert O, Zhang Y, Shen D,  
Khened M, Ding X, Krishnamurthi G,  
Kalpathy-Cramer J, Davis J, Zhao T,  
Gupta R, Saltz J and Farahani K  
(2020) Segmentation and  
Classification in Digital Pathology for  
Glioma Research: Challenges and  
Deep Learning Approaches.  
Front. Neurosci. 14:27.  
doi: 10.3389/fnins.2020.00027

<sup>1</sup> Department of Biomedical Informatics, Stony Brook University, Stony Brook, NY, United States, <sup>2</sup> Center for Biomedical Image Computing and Analytics, University of Pennsylvania, Philadelphia, PA, United States, <sup>3</sup> Department of Radiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States, <sup>4</sup> Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States, <sup>5</sup> Institute for Medical Imaging Technology, School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai, China, <sup>6</sup> Department of Engineering Design, Indian Institute of Technology Madras, Chennai, India, <sup>7</sup> Department of Medicine and Biomedical Data Science, Stanford University, Stanford, CA, United States, <sup>8</sup> School of Informatics, Xiamen University, Xiamen, China, <sup>9</sup> Department of Radiology and BRIC, The University of North Carolina at Chapel Hill, Chapel Hill, NC, United States, <sup>10</sup> Department of Brain and Cognitive Engineering, Korea University, Seoul, South Korea, <sup>11</sup> Department of Radiology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, United States, <sup>12</sup> Department of Pathology, Stony Brook University, Stony Brook, NY, United States, <sup>13</sup> Cancer Imaging Program, National Cancer Institute, National Institutes of Health, Bethesda, MD, United States

Biomedical imaging is an important source of information in cancer research. Characterizations of cancer morphology at onset, progression, and in response to treatment provide complementary information to that gleaned from genomics and clinical data. Accurate extraction and classification of both visual and latent image features is an increasingly complex challenge due to the increased complexity and resolution of biomedical image data. In this paper, we present four deep learning-based image analysis methods from the Computational Precision Medicine (CPM) satellite event of the 21st International Medical Image Computing and Computer Assisted Intervention (MICCAI 2018) conference. One method is a segmentation method designed to segment nuclei in whole slide tissue images (WSIs) of adult diffuse glioma cases. It achieved a Dice similarity coefficient of 0.868 with the CPM challenge datasets. Three methods are classification methods developed to categorize adult diffuse glioma cases into oligodendroglioma and astrocytoma classes using radiographic and histologic image data. These methods achieved accuracy values of 0.75, 0.80, and 0.90, measured as the ratio of the number of correct classifications to the number of total cases, with the challenge datasets. The evaluations of the four methods indicate that (1) carefully constructed deep learning algorithms are able to produce high accuracy in the analysis of biomedical image data and (2) the combination of radiographic with histologic image information improves classification performance.

**Keywords:** digital pathology, radiology, segmentation, classification, image analysis, deep learning

## INTRODUCTION

Cancer is a major life-threatening health problem around the world. More than 1.7 million new cancer cases and over 600,000 cancer deaths are estimated in 2019 in the United States alone (Siegel et al., 2019). Brain cancer is one of the deadliest cancer types with low survival rates among both women and men (Siegel et al., 2016; Yuan et al., 2016). Cancer research relies on accurate and reproducible disease characterizations in order to better understand what triggers cancer and how cancer progresses so that more effective means of evaluating cancer interventions can be developed. This requires assembling observational and experimental data at multiple biological scales and fusing information from multiple data modalities.

Biomedical imaging is one of the crucial data modalities in cancer research. Features gleaned from high-resolution, detailed images play a key role in the development of correlative and predictive representations of cancer morphology. Combined with clinical and genomics data, image features can result in more effective data-driven research and healthcare delivery for cancer patients. Biomedical imaging, hence, has evolved into an indispensable tool for researchers and clinicians to extract, analyze, and interpret the complex landscape of diagnostic and prognostic information and to assess treatment strategies. Radiology and the rapidly growing field of Radiomics provide a means of quantitative study of cancer properties at the macroscopic scale. Radiomics deals with the extraction, analysis, and interpretation of large sets of visual and sub-visual image features for organ-level quantification and classification of tumors (Lambin et al., 2012; Gillies, 2013; Aerts et al., 2014; Parmar et al., 2015; Gillies et al., 2016; Zwanenburg et al., 2016). The histopathologic examination of tissue, on the other hand, reveals the effects of cancer onset and progression at the sub-cellular level (Gurcan et al., 2009; Foran et al., 2011; Kong et al., 2011; Kothari et al., 2013; Griffin and Treanor, 2017; Yonekura et al., 2018). Histopathology has been used as a primary source of information for cancer diagnosis and prognosis. Diagnosis and grading of brain tumors, for example, is traditionally done by a neuropathologist examining stained tissue sections fixed on glass slides under a light microscope. Radiology is a more prevalent imaging modality in research and clinical settings. Advancements in digital microscopes made it possible to capture high-resolution images of whole slide tissue specimens and tissue microarrays, enabling increased use of virtual slides in histopathologic analysis.

In this paper, we present the application of state-of-the-art image analysis methods for segmentation and classification tasks for radiographic and histologic image data. We describe a collection of four deep learning-based methods: one method for the segmentation of nuclei and three methods for the classification of brain tumor cases. These methods are from the challenge teams who achieved the top scores at the Computational Precision Medicine (CPM) satellite event of the 21st International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI2018) and agreed to contribute to this summary manuscript. The CPM event

was organized by a subset of the co-authors on this paper as a cluster of image analysis challenges. It is one of the series of challenges organized since 2014 to provide a platform for biomedical imaging research teams to evaluate state-of-the-art algorithms in a controlled environment.

The 2018 CPM event targeted brain diffuse glioma and consisted of two sub-challenges. The *first sub-challenge* was designed to evaluate the performance of algorithms for the detection and segmentation of nuclear material in tissue images. We describe a nucleus segmentation method from this sub-challenge. The method employs an adaptation of the Mask-RCNN algorithm to solve the problem of cell segmentation in hematoxylin and eosin (H&E) stained tissue microscopy images. The authors of this method developed pre- and post-processing steps to further improve the performance of the algorithm. The method achieved a Dice similarity coefficient score of 0.868 when evaluated against a set of manually segmented tissue images. The *second sub-challenge* asked participants to classify lower grade glioma (LGG) cases into oligodendroglioma and astrocytoma subtypes using both radiology and histopathology images. We present three classification methods from this sub-challenge. One of the methods refines lower confidence predictions from a radiology image model by combining predictions from a tissue image model. The second method implements two distinct classification models for radiographic and histologic images and combines them through a dropout-enabled ensemble learning. The third method uses multiple deep learning models: one model for classifying tissue images and two models for segmenting and classifying radiology images. A weighted average operation is then applied to the classification results from tissue and radiology images to assign a class label to each case. The methods achieved accuracy values of 0.90, 0.80, and 0.75, respectively—accuracy was measured as the number of correctly classified cases divided by the total number of cases.

In addition to presenting these algorithms, we intend to make the datasets used in the MICCAI CPM 2018 challenge publicly available to provide a valuable resource for development and refinement of future segmentation and classification algorithms.

## MATERIALS AND METHODS

In this section, we first present a brief overview of existing work on biomedical image analysis (section “Related Work”). We describe the CPM challenge and datasets in Section “[Datasets and Performance Evaluation](#).” We present the nucleus segmentation method in Section “Instance Segmentation of Nuclei in Brain Tissue Images” and the three classification methods in Section “Methods for Classification of Brain Cancer Cases.”

### Related Work

Computer-aided analysis and interpretation of image data is crucial to maximizing benefits from biomedical imaging. Common image analysis operations include segmentation of regions and objects (e.g., nodules and cells) and classification of image regions and images into categories. Image features

and quantitative measures obtained from segmentation and classification can be used in downstream analyses that integrate information from clinical and molecular data and develop predictive and correlative models. Studies have shown the value of image analysis and image features in research, and an increasing number of research projects have developed image analysis methods to efficiently, accurately, and reliably convert raw image data into rich information and new knowledge (Gurcan et al., 2009; Foran et al., 2011; Kong et al., 2011; Kothari et al., 2012, 2013; Lambin et al., 2012; Gillies, 2013; Cheng et al., 2016; Coroller et al., 2016; Gao et al., 2016; Ishikawa et al., 2016; Madabhushi and Lee, 2016; Manivannan et al., 2016; Xing and Yang, 2016; Al-Milaji et al., 2017; Bakas et al., 2017c; Lehrer et al., 2017; Chang et al., 2018a, 2019; Fabelo et al., 2018; Hu et al., 2018; Khosravi et al., 2018; Lee et al., 2018; Mobadersany et al., 2018; Peikari et al., 2018; Saltz et al., 2018; Yonekura et al., 2018; Zhou et al., 2018). Recent work on biomedical image analysis focused on the development and application of machine learning methods, in particular, deep learning models.

The work done by Qian et al. detected and differentiated GBM from solitary brain metastases (van Griethuysen et al., 2017) using a support vector machine (SVM) model. The analysis algorithm computes a variety of radiomic features, using the PyRadiomics package (van Griethuysen et al., 2016; Lu et al., 2019), from contrast-enhanced Radiology image datasets. The experiments show that a combination of the least absolute shrinkage and selection operator (LASSO) and SVM achieves the best prognostic prediction performance and the highest stability. Lu et al. (Krizhevsky et al., 2012) proposed and evaluated an approach, which uses the AlexNet deep learning network (Abrol et al., 2018) as a feature extractor and applies transfer learning to train a model for brain disease detection in magnetic resonance imaging (MRI) data. The last three layers of AlexNet are replaced by a fully connected layer, a softmax layer, and a classification layer to implement the feature extractor function. Chang et al. (2018a) proposed and implemented a CNN model to predict isocitrate dehydrogenase (IDH) mutations in glioma patients using preoperative MRI data. Their experimental evaluation shows that incorporating the age at which a patient was diagnosed with cancer improves algorithm accuracy to 89%. Abrol et al. (Binder et al., 2018) applied feature selection and SVM-based classification methods on MRI data obtained from a group of GBM patients. Their experimental results show that three-dimensional radiomic features computed from radiology images could be used to differentiate pseudo-progression from true cancer progression in GBM patients. Binder et al. (Shukla et al., 2017) identified radiographic signatures of extracellular domain missense mutants (i.e., A289V) of the epidermal growth factor receptor (EGFR) suggestive of an invasive and proliferative phenotype, and associated with shorter patient survival. Their approach leverages the integrated analysis of advanced multiparametric MRI (Bakas et al., 2016) and biophysical tumor growth modeling (Akbari et al., 2018). Their findings were corroborated by experiments *in vitro* and *in vivo* in animal models, contributing to the discovery of a potential molecular target and presenting an opportunity for potential therapeutic

development (Shukla et al., 2017). Another study (Bakas et al., 2017a) found an imaging signature in radiology images of the most prevalent mutation of EGFR, namely, EGFRvIII, revealing a complex yet distinct macroscopic GBM radiographic phenotype. This signature showed a classification accuracy of ~90% for determining EGFRvIII GBM tumors. The study used an SVM model for multivariate integrative analysis of multiple image features to identify the signature. The features include the tumor's spatial distribution pattern leveraging a biophysical growth model (Akbari et al., 2018) and a distinct within-patient self-normalized heterogeneity index (Wang et al., 2019).

Mobadersany et al. (2018) examined the application of deep learning techniques to predict outcomes in LGG and glioblastoma multiforme (GBM) patients. Their approach combines tissue image analysis results with genomics data to achieve high accuracy. The deep learning network consists of convolutional layers, which are trained to predict image patterns associated with survival. This network is connected to fully connected layers that transform the image features for survival analysis. Survival data are modeled via a Cox proportional hazard layer. Wang et al. (Qian et al., 2019) implemented an analysis pipeline to classify glioma cases into grades II, III, and IV gliomas using whole slide tissue images (WSIs) from H&E and Ki-67 stained tissue samples. The pipeline consists of multiple steps, including region-of-interest (ROI) identification, image feature extraction, feature selection, automated grading of slides, and interpretation of the grading results. Multiple image features, such as the shapes and sizes of nuclei and image intensity distribution, are computed and pruned using a random forest method. The grading step employs machine learning models with automatic tuning of model parameters for the best classification performance. Saltz et al. (2018) employed a deep learning workflow to create maps of tumor-infiltrating lymphocytes (TILs) in more than 5,000 WSIs from 13 different cancer types in The Cancer Genome Atlas (TCGA) repository. The image analysis approach partitions each WSI into small (50  $\mu\text{m}$  by 50  $\mu\text{m}$ ) patches and classifies each patch as either TIL-positive or TIL-negative. The workflow implements an iterative learning phase in which predictions by the deep learning models are reviewed and corrected by pathologists, to refine and improve classification accuracy. The analysis method also uses a convolutional neural network (CNN) to identify and segment regions of necrosis in order to reduce false positives.

Nucleus segmentation is one of the core analysis tasks in histopathology imaging projects which study tissue morphology (Gurcan et al., 2009; Madabhushi and Lee, 2016; Xing and Yang, 2016). The nucleus segmentation task is challenging because of the relatively large variation in the intensity of captured signal and the ambiguity of boundary information when separating neighboring nuclei. Several projects proposed machine learning algorithms that use engineered image features and algorithms that perform statistical analyses of intensity and texture properties to detect and delineate nucleus boundaries (Kong et al., 2011; Gao et al., 2016; Peikari and Martel, 2016; Peikari et al., 2018). In recent years, there has been a significant shift toward the application of deep learning techniques. Yang et al. (2018) proposed a method that uses a U-Net model

to segment lesions in cervical cancer cases. The segmentation results are fed into a cascade network, which integrates the foreground and the edges of the segmented nuclei to generate instance segmentations. Wollmann et al. (2019) developed a hyperparameter optimization method that searches for the best parameters of a nucleus segmentation pipeline to improve segmentation accuracy. The authors evaluated their technique with two analysis pipelines, a clustering-based pipeline and a deep learning pipeline, using prostate cancer tissue images. Their results show that the deep learning pipeline performs better than the clustering-based pipeline. Alom et al. (2018) proposed a residual recurrent CNN built on the U-Net architecture (Ronneberger et al., 2015). While this type of network has been used for segmentation of macro-level objects such as retinal blood vessels and the lungs, the authors adopted it to segmentation of the nuclei. Xie and Li (2018) implemented a neural network method that learns object-level and pixel-level information in tissue image patches. The goal is to have the analysis pipeline carry out nucleus detection and nucleus segmentation simultaneously. Hou et al. (2019b) proposed a sparse convolutional autoencoder for the detection of nuclei and feature extraction in WSIs. The approach integrates nucleus detection and feature learning in a single network. The network encodes the nuclei into sparse feature maps, which represent the nuclei's locations and appearances and can be fine tuned for end-to-end supervised learning.

Radiology and pathology capture morphologic data at different biological scales. The non-invasive and non-ionizing property of MRI made it quite popular for oncology imaging studies such as brain tumors (Bakas et al., 2016). On the other hand, the *de facto* standard for tumor assessment and grading is whole slide tissue biopsy examined under a microscope. Combined use of image modalities from both domains can lead to improvements in image-based analyses. Lundstrom et al. (2017) argue for a tighter collaboration between radiology, pathology, and genomics teams toward enhanced integrated diagnosis of disease. The authors point to the increasing use of digital slide technologies in pathology as well as to the fact that computational approaches for radiology and pathology imaging modalities are not fundamentally different. They note that combining complementary views of the disease from multiple scales can maximize the benefits of biomedical imaging. Madabhushi and Lee (2016) note that researchers are increasingly looking at opportunities for combining radiomic data with features extracted from high-resolution pathology image for better predictive capabilities in disease prognosis. On the methodology and software front, Arnold et al. (2016) developed a web-based platform that integrates radiology and pathology data for cancer diagnosis. Saltz et al. (2017) devised methods and tools for combined computation, management, and exploration of image features from radiology and pathology image datasets. Kelahan et al. (2017) implemented a dashboard for radiologists to view pathology reports to aid with diagnosis and image-guided decision making. McGarry et al. (2018) proposed a method for combining multi-parametric MRI data with digital pathology slides to train predictive models for prostate cancer localization.

Despite a growing body of research and development on methods and tools, computerized image analysis continues to be a challenging task. Both image resolutions and data complexity continue to increase, requiring the enhancement of existing methods and the development of new techniques. For example, contemporary digital microscopy scanners are capable of imaging whole slide tissue specimens at very high resolutions (e.g., over  $80,000 \times 80,000$  pixels). These images may contain millions of cells and nuclei, and multiple types of regions (e.g., tumor, stromal, and normal tissues). There can be significant morphological heterogeneity within a specimen, as well as across specimens in both radiographic and histologic imaging, requiring novel methods that can handle heterogeneity and increasing the density of morphologic information.

## Datasets and Performance Evaluation

The approaches, which will be described in Sections “Instance Segmentation of Nuclei in Brain Tissue Images” and “Methods for Classification of Brain Cancer Cases,” were experimentally evaluated with radiographic and histologic image datasets from the MICCAI 2018 CPM challenge event. Here we provide a brief description of the challenge datasets and the methods for scoring algorithm performance. The datasets for the 2018 CPM challenge were obtained from TCGA<sup>1</sup> (Tomczak et al., 2015) and The Cancer Imaging Archive (TCIA)<sup>2</sup> (Clark et al., 2013; Prior et al., 2013) repositories, and the images had been scanned at the highest resolution. Images from these sources are publicly available and have been used in many publications (e.g., Aerts et al., 2014; Yu et al., 2016; Bakas et al., 2017c; Mobadersany et al., 2018; Saltz et al., 2018; Agarwal et al., 2019).

### Datasets for Segmentation of Nuclei in Pathology Images

A WSI may contain hundreds of thousands of nuclei; some images with large tissue coverage will have more than one million nuclei. Manually segmenting all nuclei in the entire WSIs would be infeasible. Thus, we extracted image tiles from WSIs and used the tiles in the training and test datasets in order to reduce the cost of generating high-quality ground truth data as well as the computational requirements of the training and test steps of analysis algorithms. The image tiles were selected by a pathologist and extracted from a set of GBM and LGG WSIs at the highest resolution. The training and test datasets consisted of 15 and 18 image tiles, respectively. The sizes of the tiles ranged from  $459 \times 392$  pixels to  $1032 \times 808$  pixels in the training set and from  $378 \times 322$  pixels to  $500 \times 500$  pixels in the test set. The nuclei in each image tile were segmented by two students. The segmentation results were reviewed, refined, and consolidated by the pathologist to generate the final set of segmentation data. This process generated 2905 and 2235 nuclei in the training and test sets, respectively.

In the challenge event, the performance of a segmentation algorithm was measured as the average of the standard Dice similarity coefficient and a modified version of the Dice metric.

<sup>1</sup><https://portal.gdc.cancer.gov>

<sup>2</sup><https://www.cancerimagingarchive.net>



The standard Dice score (Dice, 1945) measures the overlap between two sets of segmentation results without taking into account the individual nuclei. That is, it computes the amount of overlap between the ground truth mask and the mask generated by the segmentation algorithm without considering splitting and merging of the nuclei by the algorithm. The modified Dice metric aims to incorporate split and merge errors into the score. We refer the reader to an earlier publication (Vu et al., 2019) for a more detailed description of the modified Dice metric.

## Datasets for Combined Radiology and Pathology Classification

The datasets were matched MRI and digital pathology images obtained from the same patients and the same time point. Each case corresponded to a single patient. There was one set of MRI data (T1, T1C, FLAIR, and T2 images) and one corresponding WSI for each case. The training set contained a total of 32 cases: 16 cases that were classified as oligodendroglioma and 16 cases classified as astrocytoma. The test dataset consisted of 20 cases with 10 cases of oligodendroglioma and 10 cases of astrocytoma. We retrieved the WSI and MRI images from the TCGA and TCIA archives, respectively. These images had been obtained and classified following the protocols implemented in the TCGA project<sup>3</sup>. We obtained the ground truth classification labels of the cases from the associated clinical and metadata in the TCGA repository. These classifications were further reviewed by a pathologist and a radiologist. In the challenge event, we used the accuracy of a classification method to score its performance and rank it. We counted the number of correctly classified cases and divided that number by the total number of cases to compute the accuracy score.

In the following sections, we will present a nucleus segmentation algorithm (section “Instance Segmentation of Nuclei in Brain Tissue Images”), which achieved the second highest score in the segmentation challenge, and three classification algorithms (section “Methods for Classification of Brain Cancer Cases”), which achieved the top three scores in the classification challenge.

## Instance Segmentation of Nuclei in Brain Tissue Images

In this section, we present the nucleus segmentation algorithm developed by XR, QW, LZ, and DS. This method achieved the second highest score in the CPM challenge and its developers agreed to contribute to this manuscript.

The method implements an application of the Mask-RCNN network (He et al., 2017) with a novel MASK non-maximum suppression (MASK-NMS) module, which can increase the robustness of the model. Mask-RCNN is a deep learning network extended from the Faster-RCNN model (Ren et al., 2015) and is used to carry out semantic and object instance segmentation (see Figure 1). In our implementation, we used ResNet-101 to build a Mask-RCNN pyramid network backbone for the segmentation of nuclei in WSIs. This adaptation is based on

an existing implementation by Matterport<sup>4</sup>. We have extended this implementation in several ways to improve segmentation performance. First, we have reduced the region proposal network (RPN) anchor sizes and increased the number of anchors to be used because the nuclei are small objects and can be found anywhere in a tissue image. Second, we have increased the maximum number of predicted objects, since even a small image tile from a tissue slide can contain 1000 or more nuclei. Moreover, rather than training the network end-to-end from the start, we initialized the model using weights from the pre-training on the MSCOCO dataset (Lin et al., 2014). We train the layers in multiple stages. We first train the network heads after they are randomly initialized. We later train the upper layers of the network. After this, we reduce the learning rate by a factor of 10 and train the entire network end to end. In our experiments, the training took 300 epochs using stochastic gradient descent with momentum set to 0.9. During training and testing, input tissue images were cropped to 600 × 600.

In addition to the above extensions, we implemented a set of pre-processing steps to further improve the algorithm performance. Holes in the masks are filled by an image morphology operation. Fused nucleic masks are split by applying morphological erosion and dilation. To help avoid overfitting, data augmentation, which could increase the amount of training data, is applied in the form of random crops, random rotations, Gaussian blurring, and random horizontal and vertical flips.

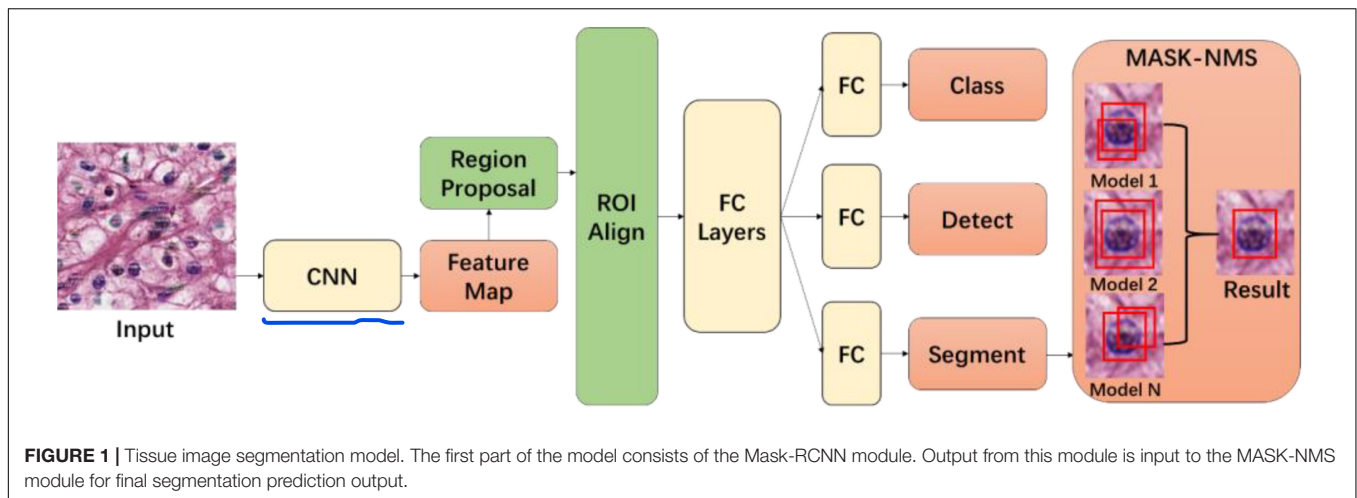
Our implementation combines predictions from fivefold cross training models in a post-processing step (see Figure 1). We have implemented this step in a novel module called MASK-NMS, which is one of our contributions in the segmentation method. MASK-NMS takes unions of masks with maximum overlap and removes false-positive masks with a small overlap. It starts with a set of segmentation results. This set is called  $I$ . Each result in set  $I$  is assigned a score  $S$ , which is the value of the classification probability from the Mask-RCNN module and corresponds to the confidence level of the segmentation result. After selecting the segmentation with the maximum score  $M$  (the maximum score among scores  $S$ ), MASK-NMS removes it from the set  $I$  and appends it to the final segmentation set  $D$ .  $D$  is initialized to an empty set. It also removes any segmentations with an overlap greater than a threshold  $N$  in the set  $I$ , where the intersection over union (IOU) is used as the overlap metric. IOU is also known as the Jaccard similarity index (Jaccard, 1901), which measures the similarity between finite sample sets. It is defined as the size of the intersection between two sets divided by the size of the union of the sets. The selection process repeats until set  $I$  becomes empty. Finally, we obtain the segmentation results in set  $D$ . The MASK-NMS module assembles multiple results together and reduces false positives and false negatives.

## Methods for Classification of Brain Cancer Cases

In this section, we present three classification algorithms, which achieved the top three scores in the classification challenge and the developers of which agreed to contribute to this manuscript.

<sup>3</sup><https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga/using-tcga/types>

<sup>4</sup>[https://github.com/matterport/Mask\\_RCNN](https://github.com/matterport/Mask_RCNN)



### An Approach for Classification of Low-Grade Gliomas Using Combined Radiology and Pathology Image Data

The top-performing method (developed by AB, AsK, AvK, MK, and GK) (Bagari et al., 2018) in the classification challenge uses an MRI classification model and a WSI classification model and combines the predictions from the two models to assign a class to a given case. The overall analysis pipeline is depicted in **Figures 2–4** and described below.

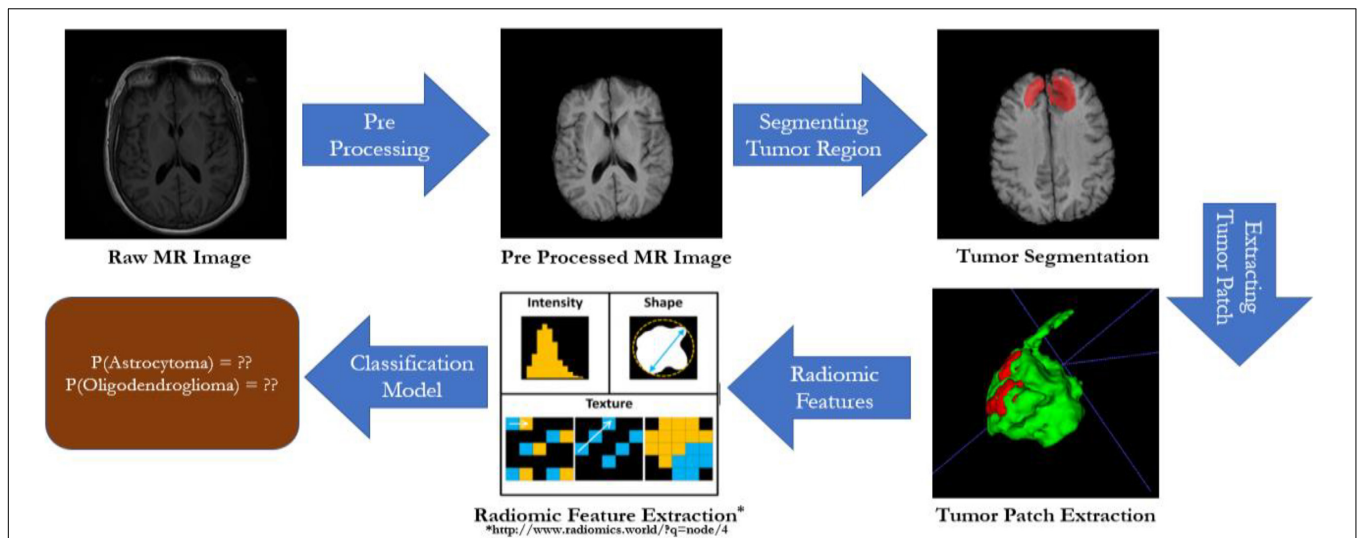
#### Radiology image analysis pipeline

Different pulse sequences in MRI, including native T1- and T2-weighted, T2-Flair, and T1-weighted post-contrast imaging, can be used to enhance different parts of a tumor. In this part of the pipeline (**Figure 2**), we execute a segmentation pipeline consisting of the following steps on these images before features are computed from the images and used in the classification model: (1) Skull stripping: It is necessary to remove the skull from MRI as its presence can be wrongly interpreted as a tumor, and most segmentation networks are trained using skull-stripped images. (2) Co-registration and re-sampling to isotropic voxel spacing: Following 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 acquired on a different day or using a different scanner. Registered images are spatially correlated across channels and can be used for tumor segmentation. We register sequences T1, FLAIR, and T2 with respect to T1c scan. The MRI volumes are re-sampled to an isotropic voxel resolution of  $1 \text{ mm}^3$  after the co-registration step. (3) Segmentation of tumor regions using a CNN: Tumor regions are segmented by a fully CNN trained on the BraTS-2018 dataset (Menze et al., 2014; Bakas et al., 2017b,c, 2018; Crimi et al., 2018). After the segmentation step, a set of 105 radiomic features are computed on segmented regions using the pyradiomic library (Lu et al., 2019). These features include shape features, first-order statistics, features from gray level co-occurrence matrix, features from gray level run length matrix and gray level size zero matrix, and neighboring gray tone difference matrix. The 105-dimensional radiomic feature

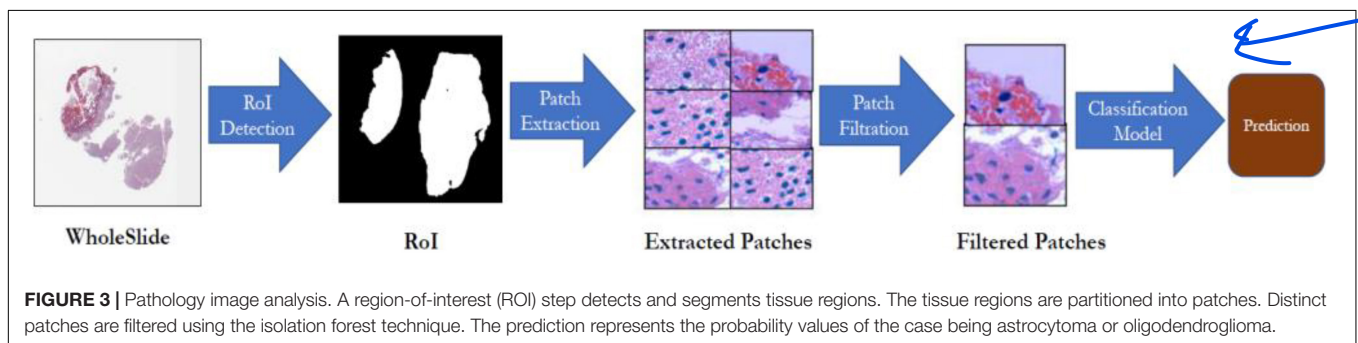
vectors are reduced to a 16-dimensional feature vector using the principal component analysis. A classification model is trained with 16-dimensional feature vectors as input. If the training dataset has  $N$  cases, the model is trained with an  $(N,16)$  input using logistic regression with the liblinear optimization algorithm (Fan et al., 2008) and a fivefold cross-validation process. This process fits a logistic regression model on the entire training data. Classification predictions from the MRI data are obtained using this model.

#### Analysis pipeline for whole slide tissue images

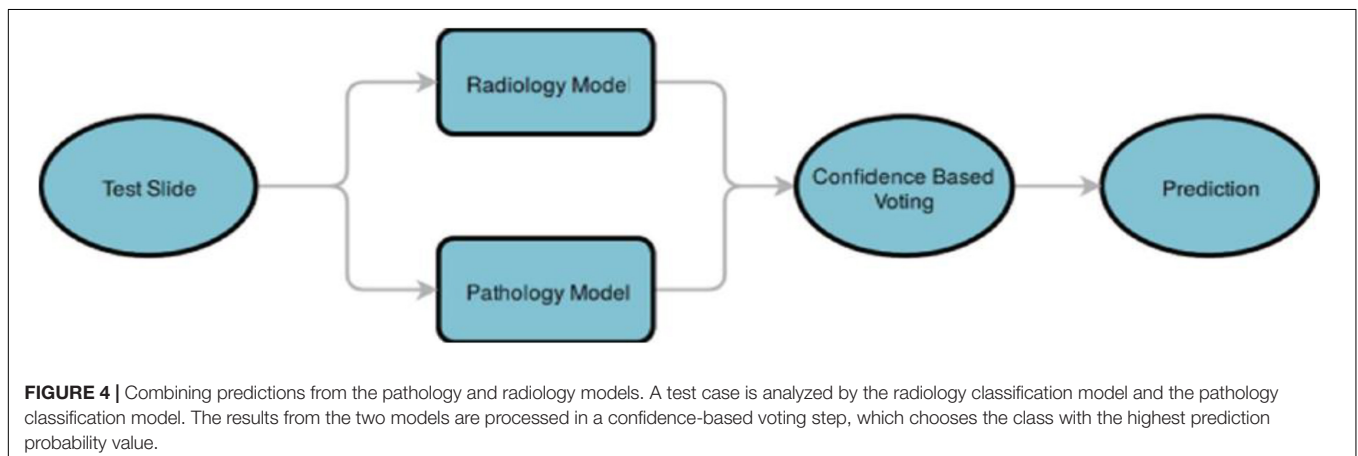
Tissue slides may contain large areas of glass background that are irrelevant to image analysis and should be removed. In this part of the pipeline (**Figure 3**), in order to detect and segment tissue regions and remove regions corresponding to glass background, a tissue image is first converted from the RGB color space to the HSV color space. Then, lower and upper thresholds are applied on color intensities to get a binary mask. The binary mask is processed to fill in small holes and remove clustered clumps from foreground pixels. After this step, bounding boxes around all the discrete contours are obtained. The bounding boxes serve as blueprints for the patch extraction process. The patch extraction process partitions the segmented tissue region into  $224 \times 224$ -pixel patches. The  $224 \times 224$ -pixel patches are color-normalized (Reinhard et al., 2001) and assigned the same label as the label of the WSI. A subset of distinct patches is filtered out using an outlier detection technique called the Isolation Forest (Liu et al., 2008). The filtering step is executed as follows. We train an autoencoder with a pixel-wise reconstruction loss to generate feature vector representations of patches from the input image. The isolation forest method is then executed with these feature vectors to find outlier patches. The remaining patches after the outlier detection step are used to refine a DenseNet-161 network, which has been pre-trained on ImageNet. Binary cross entropy is used as the loss function. During the prediction phase, test patches extracted from a WSI are classified using the trained model, and a probability score is assigned to the image based on a voting of classes predicted for individual patches.



**FIGURE 2 |** Radiology image analysis. Images are pre-processed (i.e., skull stripping and co-registration) before they are analyzed through the remaining steps of the analysis pipeline. After the pre-processing step, tumor regions in the images are segmented via a CNN model. This step is followed by computation of a set of 105 radiomic features in segmented regions. The high-dimensional feature vector is reduced to a 16-dimensional feature vector using the principle component analysis method. A classification network is trained with these feature vectors.



**FIGURE 3 |** Pathology image analysis. A region-of-interest (ROI) step detects and segments tissue regions. The tissue regions are partitioned into patches. Distinct patches are filtered using the isolation forest technique. The prediction represents the probability values of the case being astrocytoma or oligodendroglioma.



**FIGURE 4 |** Combining predictions from the pathology and radiology models. A test case is analyzed by the radiology classification model and the pathology classification model. The results from the two models are processed in a confidence-based voting step, which chooses the class with the highest prediction probability value.

### Combining predictions

As is shown in **Figure 4**, finally, predictions from both the radiology and pathology models are compared, and the class label of a case is determined based on the model, which gives a prediction with a higher probability score.

### Dropout-Enabled Ensemble Learning for Multi-Scale Biomedical Image Classification

This method is the second best performing (developed by AM, MT, and OG) (Momeni et al., 2018) and proposes two distinct classification models for radiographic and histopathologic

images and their integration through dropout-enabled ensemble learning.

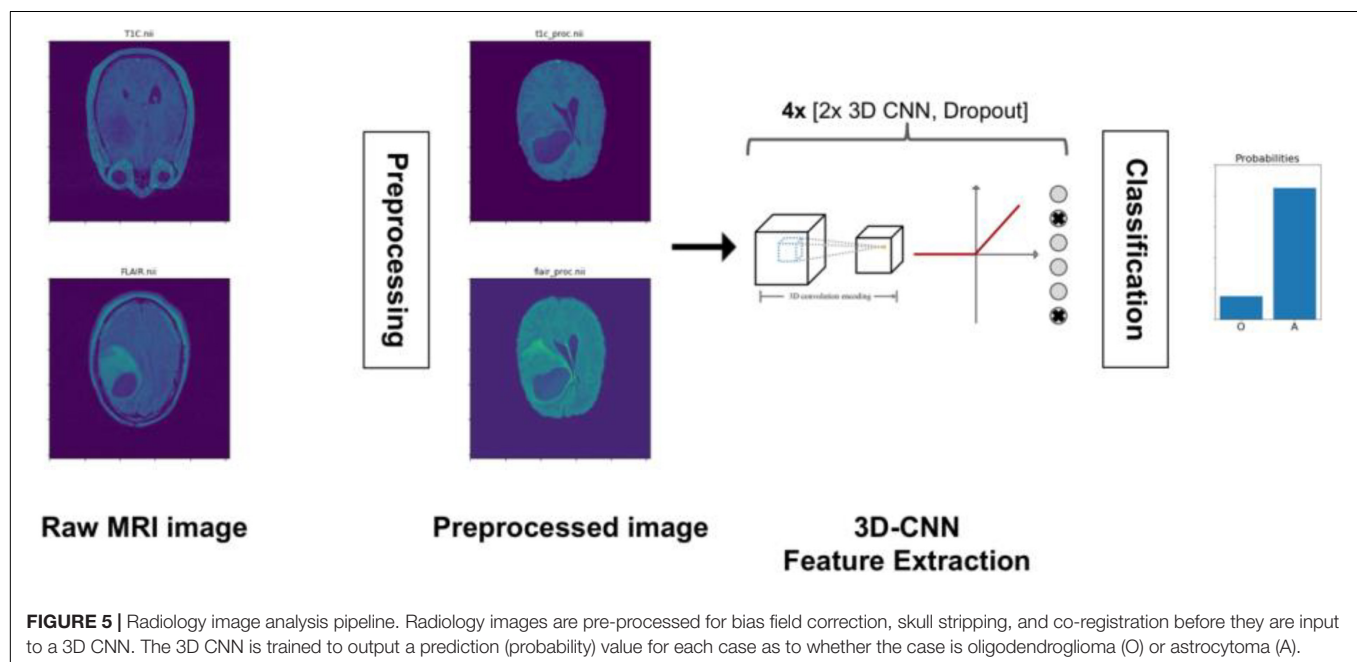
### Radiology classification model

As is shown in **Figure 5**, radiology images are pre-processed through a pipeline of bias field correction, skull-stripping, and co-registration steps before they are input to a 3D CNN network. The 3D CNN consists of eight layers to extract deep features from MRI and three 3D max pooling to reduce the sample size. The input of the 3D CNN is a 3D voxel image in the form of three spatial dimensions and two modalities per voxel. In this work, the 3D CNN is trained with the T1c and T2-FLAIR modalities only because these are the most informative for LGG segmentation. After the last convolutional layer, the extracted features are averaged over all of the 3D space to yield a unique 100-dimensional feature vector per case. This vector is connected to a 1D output for classification with cross-entropy loss. The whole network is then trained. To avoid overfitting, we use classical data augmentation techniques (rotation, cropping, etc.)

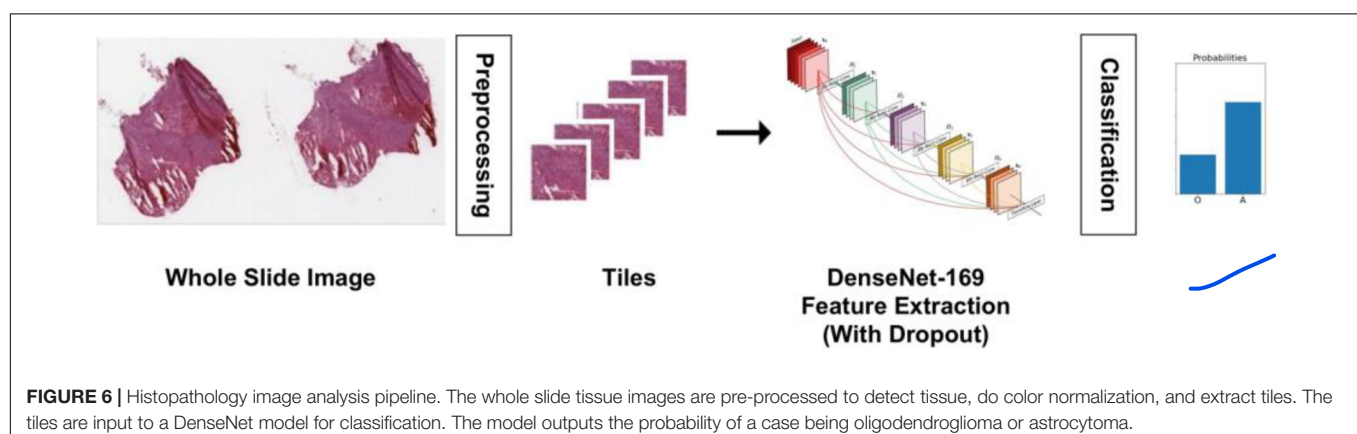
as well as dropout. Eight dropout layers are placed throughout the network to avoid overfitting and for the ensemble learning step.

### Histopathology classification model

A multiple instance learning approach, as shown in **Figure 6**, is implemented for the histopathology images. The learning step is carried out after a pre-processing phase. The pre-processing steps here consist of tissue detection, color normalization, and tiling. Tissue detection is done with Otsu thresholding to detect and segment tissue regions only, eliminating regions that are glass background. A simple histogram equalization algorithm is used for color normalization prior to tiling. The tiling step extracts  $20\,448 \times 448$ -pixel patches from a WSI by uniform random sampling. Once the image patches have been extracted, a DenseNet network pretrained on ImageNet is fine tuned, after removing its last fully connected layer. The remainder of the network is used as a fixed feature extractor for tissue images, and two fully connected layers with dropout are used for classification. As with the radiology model, we used

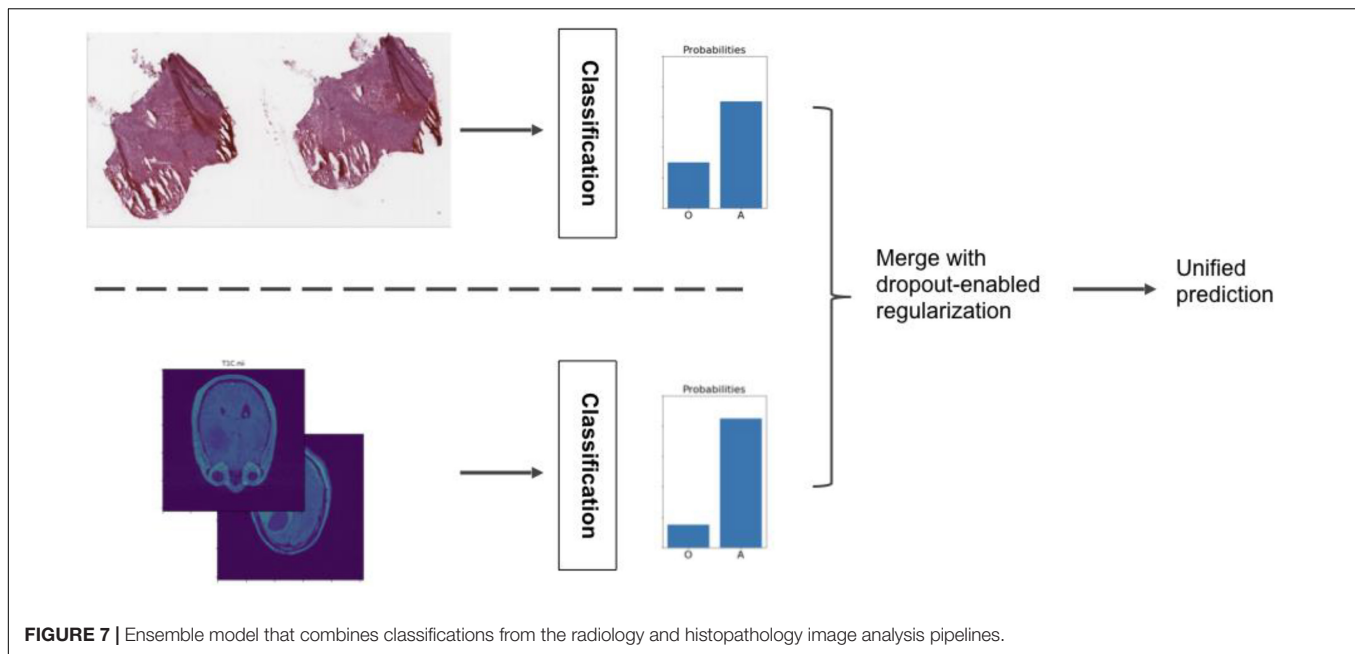


**FIGURE 5 |** Radiology image analysis pipeline. Radiology images are pre-processed for bias field correction, skull stripping, and co-registration before they are input to a 3D CNN. The 3D CNN is trained to output a prediction (probability) value for each case as to whether the case is oligodendroglioma (O) or astrocytoma (A).



**FIGURE 6 |** Histopathology image analysis pipeline. The whole slide tissue images are pre-processed to detect tissue, do color normalization, and extract tiles. The tiles are input to a DenseNet model for classification. The model outputs the probability of a case being oligodendroglioma or astrocytoma.





classical data augmentation techniques along with dropout to eliminate overfitting.

### Ensemble learning model

The main contribution of our approach is a meta-algorithm that combines the histopathology and radiology classification models, as is shown in **Figure 7**. In this ensemble learning methodology, each model is trained separately. Their predictions are combined into a single, more robust output. The basic idea is to extract the one-to-last feature layer from each individual classification model and form a single feature vector for each case/patient by concatenating the two feature vectors. An SVM model is then trained with the combined feature vectors to classify the cases. However, if the training dataset is small (which is the case with the CPM challenge dataset; we have 50-dimensional feature vectors from both the classification models and only 32 cases in the training dataset), the classification problem can become under-determined and result in overfitting of the models. To address this problem, we use regularization through dropout in the ensemble learning step. The idea is to enable the dropout values of the models in the test phase, so that individual models produce multiple (typically thousands) feature vectors for each subject. These many feature vectors can then be concatenated to form the combined feature vectors, creating a training dataset big enough for the SVM model (Momeni et al., 2018). Dropout at test time results in sampled feature vectors that are both distinct and informative and provides sufficient variance in the training dataset. Hence, the ensemble learning method can learn a more accurate and robust model from the newly produced dataset.

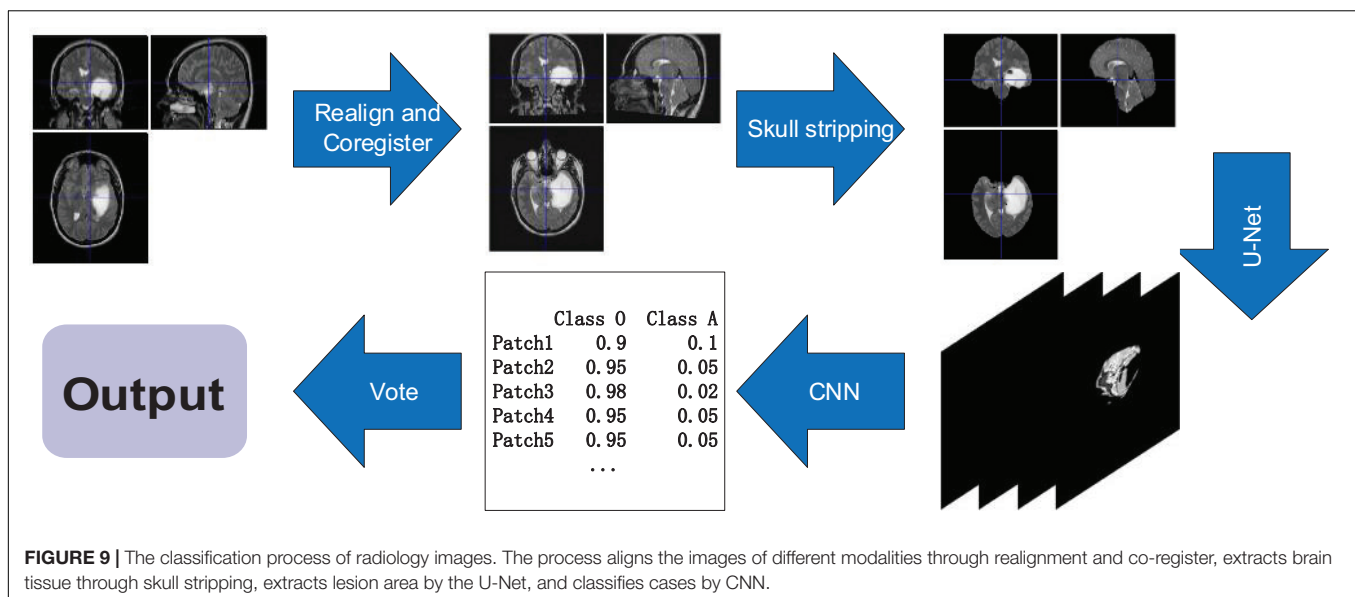
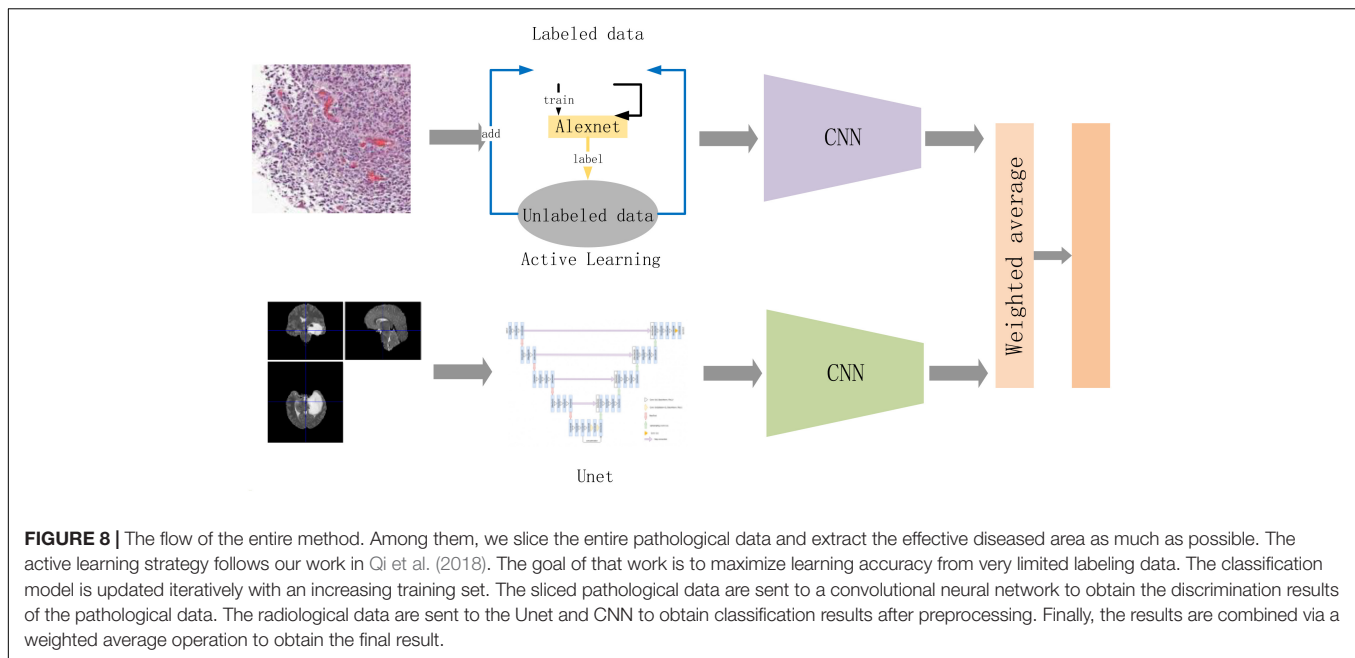
### A Weighted Average-Based Classification Method

The third best performing method (developed by QQ, YZ, YH, and XD) is illustrated in **Figure 8**. It analyzes each imaging modality (radiology images and pathology images) separately and

combines the prediction results via a weighted average operation. We describe the individual classification models and weighted average operation below.

Classification of pathology images is carried out by identifying tissue characteristics that differentiate oligodendroglioma from astrocytoma. Astrocytoma is noted to have more grades, as well as necrosis, increased cell density, calcification, and nuclear atypia. On the other hand, fried egg-like cells, and the tissue characteristics of chicken-cage-like blood vessels are unique to oligodendroglioma. In the proposed method, each histologic image is partitioned into  $512 \times 512$  patches. A sample set is created to identify typical samples of both subtypes of brain diffuse gliomas to assess imbalance in the data. In order to prevent the classification error caused by data imbalance, our method expands the sample set by rotating the original image in symmetrical and asymmetrical directions. The balanced samples are then sent to a CNN classifier network, which is trained to fully recognize the tissue and cell characteristics of oligodendroglioma and astrocytoma (see **Figure 8**). The method uses the VGG16 CNN network (Simonyan and Zisserman, 2014). We use data augmentation and add dropout layers or batch normalization layers to the classification model to reduce the risk of overfitting the model.

The classification model for radiology images is shown in **Figure 9**. Radiology images are pre-processed using methods from the SPM12 software (Penny et al., 2011). The methods include *Realign*, *Estimate*, and *Re-slice* to register data of the same modality in different cases; *Co-register* and *Estimate* and *Re-slice* to register different modal data of the same case; and *Segment* and *ImCalc* to extract the intracranial cavity. The pre-processed images are then segmented using the U-Net (Ronneberger et al., 2015) segmentation network. Patches with tumor, which are predicted by the segmentation network, are used as training data for a 2D Densenet (Huang et al., 2017) network. We classify each



patch, set the threshold value of 0.99, and select effective patches. The ensemble of multiple patches can effectively improve the robustness of the classifier.

Classification results from the radiology image dataset and the pathology image dataset are combined via a weighted average operation (see **Figure 8**):

$$\hat{y} = \alpha * f(X_p) + (1 - \alpha) * g(X_r)$$

where the classifiers for pathology data and radiology data,  $X_p$  and  $X_r$  are VGG16 and DenseNet, respectively.  $f(\bullet)$  and  $g(\bullet)$  represent the probabilities acquired from softmax function in  $X_p$  and  $X_r$ . The weight  $\alpha$  is empirically estimated in predicting the final classification label  $\hat{y}$ .

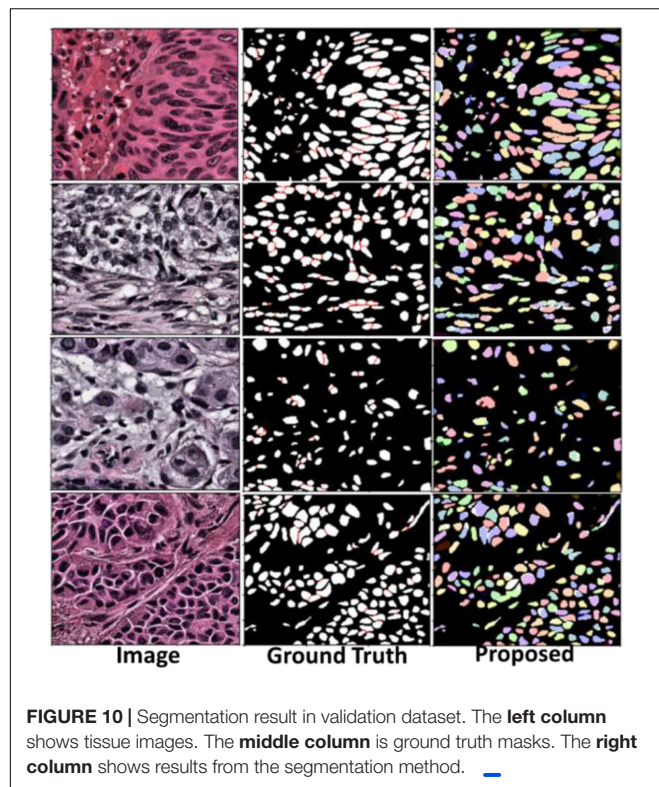
## EXPERIMENTAL RESULTS

### Segmentation of Nuclei

The Mask-RCNN model with ResNet-101 backbone obtained the 45.02% mean IOU (mIOU) on fivefold validation dataset. mIOU is the average precision score for each IOU with different thresholds (from 0.05 to 0.95 in the challenge). Tissue images with nuclei detections and segmentations are illustrated in **Figure 10**. A Dice score of 0.868 was achieved with the test dataset.

### Classification of Cancer Cases

On testing the algorithms on a dataset containing 20 radiology and pathology images, the three methods in Section “Methods



for Classification of Brain Cancer Cases” achieved the accuracy scores (i.e., the number of correctly classified cases divided by the total number of cases) as shown in **Table 1**.

## DISCUSSION AND CONCLUSION

Biomedical imaging has made great strides in image resolution and image capture speeds over the past decade. Radiology has enjoyed a widespread adoption for many years in both research and clinical settings. New imaging technologies are now allowing researchers to capture larger volumes of more detailed radiology data. Digital microscopy scanners were emerging technologies about 20 years ago. They required constant attention to capture sharp images of tissue and took many hours to scan a tissue specimen at moderate magnification levels. Nowadays, hundreds of slide tissues can be automatically imaged in several minutes. New scanning technologies and tissue staining methods are enabling researchers to capture richer morphological information at unprecedented resolutions. We anticipate that the FDA’s approval in 2017 of WSIs as a primary diagnostic tool will fuel a rapid increase in adoption of virtual slide technologies by researchers and clinicians. Combined with cheaper storage space, more powerful computing capabilities (via multi-core CPUs and accelerators such as graphics processing units), and Cloud computing infrastructures, biomedical imaging is rapidly becoming an essential tool in cancer research.

On the image analysis front, deep learning methods have seen a tremendous intake from the imaging community. These methods have demonstrated excellent results in the analysis

**TABLE 1** | Accuracy scores of the classification methods presented in Section “Methods for Classification of Brain Cancer Cases.”

Method	Score
Section “An Approach for Classification Of Low-Grade Gliomas Using Combined Radiology and Pathology Image Data”	0.90
Section “Dropout-Enabled Ensemble Learning for Multi-Scale Biomedical Image Classification”	0.80
Section “A Weighted Average-Based Classification Method”	0.75

of natural images. A rapidly growing collection of efforts are adapting these methods and extending them in innovative ways for application in biomedical image analysis. The segmentation method presented in this work shows the use of Mask-RCNN along with a non-maximum suppression (NMS) module for robust segmentation of nuclei in WSIs. The image classification methods employ a variety of deep learning methods and combine information from both radiology and pathology images to improve classification accuracy. All the methods described in this paper were evaluated with image ground truth data generated in the MICCAI CPM 2018 challenge (organized by a subset of the co-authors as denoted in the author list). The experimental results for nucleus segmentation show that high performance (i.e., high Dice scores) can be achieved by integrated use of Mask-RCNN and NMS for nucleus segmentation. The results for the classification methods show that a carefully assembled set of pipelines for each imaging modality and combination of prediction results from individual models can produce high classification accuracy.

While our work and works by other research teams have shown significant progress with more accurate, efficient, and robust image analysis algorithms, there remain challenges. One of the major challenges in machine learning analysis of biomedical imaging data is the lack of large curated and annotated training datasets, primarily because of time effort and domain expertise required for manual segmentations and classifications of tissue regions and micro-anatomic structures, such as nuclei and cells, as well as because of privacy and ownership concerns of source datasets. Some initial studies in the field of distributed learning in medicine attempted to address the data privacy and ownership challenge (Chang et al., 2018b; Sheller et al., 2018). These approaches need more investigation and adoption to facilitate collaboration across multiple medical institutions. Some projects have looked at the use of synthetic training datasets. Mahmood et al. (2018), for example, devised a method based on a conditional generative adversarial network (GAN) to improve deep learning-based segmentation of nuclei. Their method trains segmentation models using synthetic and real data. The authors employed a cycle GAN method to generate pairs of synthetic image patches and segmentation masks with varying amounts of touching and clumped nuclei. Such nuclei are difficult to segment by automated algorithms. In another work, Hou et al. (2019a) proposed a GAN architecture for the generation of synthetic tissue images and segmentation masks. The GAN architecture consists of multiple CNNs; a set of CNNs generates and refines synthetic images and masks to reference styles, and

another CNN is trained online with these images and masks to generate a segmentation model. Another GAN approach was proposed by Senaras et al. (2018b) for tumor grading. The GAN network generates synthetic image datasets with known amounts of positive and negative nuclei in immunohistochemistry-stained tissue specimens (Senaras et al., 2018b).

Another major challenge in automated biomedical image analysis is the quality assessment of input datasets and analysis results. This also is a time-consuming and labor-intensive task, as automated algorithms can process large numbers of images and generate large volumes of analysis output to be reviewed and validated, thanks to advances in computing systems. There is a need to automate the quality assessment and validation processes. Some projects are looking at this problem. A recent work by Senaras et al. (2018a) used deep learning methods to detect out-of-focus regions in WSIs so that image analysis pipelines can avoid such regions. An approach proposed by Wen et al. utilized multiple machine learning methods, namely, SVM, random forest, and CNN, to assess the quality of nuclear segmentation results. The proposed approach made use of texture and intensity features extracted from image patches in a WSI to train the quality control models (Wen et al., 2017, 2018).

As our capability to capture complex radiology and pathology image data more rapidly and at higher resolutions evolves, manual training data generation and quality evaluation will become increasingly infeasible. We expect that (semi-)automated approaches, for training data generation, for assessing the quality of data and analysis results, and for iterative refinement of deep learning models, will become important tools in a researcher's and clinician's imaging toolset. We also believe image analysis challenges, such as the MICCAI 2018 CPM challenge, are important in efforts to develop more robust methods for image analysis and method assessment and validation. One of the issues that face machine/deep learning algorithm developers is the limited amount of ground truth datasets in biomedical imaging—the small dataset size is a limitation in our work as well. Thus, in addition to providing a platform for researchers to evaluate their methods in a controlled environment, image analysis challenge events contribute to a growing set of curated datasets that are

valuable resources for development and refinement of future segmentation and classification algorithms. As part of our work, we make the datasets used in this challenge available to other researchers upon request.

## DATA AVAILABILITY STATEMENT

The datasets analyzed for this study can be found in The Cancer Genome Atlas (<https://portal.gdc.cancer.gov>) and The Cancer Imaging Archive repositories (<https://www.cancerimagingarchive.net>). Please contact the corresponding author for details about the datasets and how to obtain them.

## AUTHOR CONTRIBUTIONS

TK, SB, JK-C, JS, and KF were the main organizers of the MICCAI 2018 CPM satellite event presented in Section “Segmentation of Nuclei.” JD, TZ, RG, SB, and TK generated the datasets used in the challenge, supervised the collection of ground truth data, and developed the methods for evaluation of the challenge submissions. XR, LZ, QW, and DS proposed and implemented the nucleus segmentation algorithm presented in Section “Related Work.” YH, QQ, YZ, and XD proposed and implemented the classification algorithm described in Section “Datasets for Segmentation of Nuclei in Pathology Images.” AB, AsK, AvK, MK, and GK proposed and implemented the classification algorithm described in Section “Datasets for Combined Radiology and Pathology Classification.” All of the authors contributed text to the paper and edited it.

## FUNDING

This work was supported in part by the National Institutes of Health under award numbers NCI:U24CA180924, NCI:U24CA215109, NCI:UG3CA225021, NCI:U24CA189523, NINDS:R01NS042645, and R01LM011119 and R01LM009239 from the U.S. National Library of Medicine. The content of this publication is solely the responsibility of the authors and does not represent the official views of the NIH.

## REFERENCES

- Abrol, S., Kotrotsou, A., Hassan, A., Elshafeey, N., Idris, T., Manohar, N., et al. (2018). Abstract 3040: radiomics discriminates pseudo-progression from true progression in glioblastoma patients: a large-scale multi-institutional study. *Cancer Res.* 78:3040.
- Aerts, H. J., Velazquez, E. R., Leijenaar, R. T., Parmar, C., Grossmann, P., Carvalho, S., et al. (2014). Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat. Commun.* 5:4006. doi: 10.1038/ncomms5006
- Agarwal, R., Diaz, O., Lladó, X., Yap, M. H., and Martí, R. (2019). Automatic mass detection in mammograms using deep convolutional neural networks. *J. Med. Imaging* 6:031409.
- Akbari, H., Bakas, S., Pisapia, J. M., Nasrallah, M. P., Rozycki, M., Martinez-Lage, M., et al. (2018). In vivo evaluation of EGFRvIII mutation in primary glioblastoma patients via complex multiparametric MRI signature. *Neuro Oncol.* 20, 1068–1079. doi: 10.1093/neuonc/noy033
- Al-Milaji, Z., Ersoy, I., Hafiane, A., Palaniappan, K., and Bunyak, F. (2017). Integrating segmentation with deep learning for enhanced classification of epithelial and stromal tissues in H&E images. *Pattern Recognit. Lett.* 119, 214–221. doi: 10.1016/j.patrec.2017.09.015
- Alom, M. Z., Yakopcic, C., Taha, T. M., and Asari, V. K. (2018). “Nuclei Segmentation with recurrent residual convolutional neural networks based U-Net (R2U-Net),” in *Proceedings of the NAECON 2018 - IEEE National Aerospace and Electronics Conference*, Dayton, OH, 228–233.
- Arnold, C. W., Wallace, W. D., Chen, S., Oh, A., Abtin, F., Genshaft, S., et al. (2016). RadPath: a web-based system for integrating and correlating radiology and pathology findings during cancer diagnosis. *Acad. Radiol.* 23, 90–100. doi: 10.1016/j.acra.2015.09.009



- Bagari, A., Kumar, A., Kori, A., Khened, M., and Krishnamurthi, G. (2018). "A combined radio-histological approach for classification of low grade gliomas," in *Proceedings of the International MICCAI Brainlesion Workshop*, (Berlin: Springer), 416–427. doi: 10.1007/978-3-030-11723-8\_42
- Bakas, S., Akbari, H., Pisapia, J., Martinez-Lage, M., Rozycki, M., Rathore, S., et al. (2017a). In Vivo detection of EGFRvIII in glioblastoma via perfusion magnetic resonance imaging signature consistent with deep Peritumoral infiltration: the phi-Index. *Clin. Cancer Res.* 23, 4724–4734. doi: 10.1158/1078-0432.CCR-16-1871
- Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J. S., et al. (2017b). Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. *Sci. Data* 4:170117. doi: 10.1038/sdata.2017.117
- Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J. S., et al. (2017c). Segmentation labels and radiomic features for the pre-operative scans of the TCGA-LGG collection. *Cancer Imaging Arch.* 2017:286. doi: 10.1038/sdata.2017.117
- Bakas, S., Reyes, M., Jakab, A., Bauer, S., Rempfler, M., Crimi, A., et al. (2018). Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge. *arXiv [preprint]*
- Bakas, S., Zeng, K., Sotiras, A., Rathore, S., Akbari, H., Gaonkar, B., et al. (2016). GLISTRboost: combining multimodal MRI segmentation, registration, and biophysical tumor growth modeling with gradient boosting machines for glioma segmentation. *Brainlesion* 9556, 144–155. doi: 10.1007/978-3-319-30858-6\_1
- Binder, Z. A., Thorne, A. H., Bakas, S., Wileyto, E. P., Bilello, M., Akbari, H., et al. (2018). Epidermal growth factor receptor extracellular domain mutations in glioblastomas present opportunities for clinical imaging and therapeutic development. *Cancer Cell* 34:163–177.e7. doi: 10.1016/j.ccell.2018.06.006
- Chang, H. Y., Jung, C. K., Woo, J. I., Lee, S., Cho, J., Kim, S. W., et al. (2019). Artificial Intelligence in Pathology. *J. Pathol. Transl. Med.* 53, 1–12.
- Chang, K., Bai, H. X., Zhou, H., Su, C., Bi, W. L., Agboda, E., et al. (2018a). Residual convolutional neural network for the determination of idh status in low- and high-grade gliomas from MR imaging. *Clin. Cancer Res.* 24, 1073–1081. doi: 10.1158/1078-0432.CCR-17-2236
- Chang, K., Balachandar, N., Lam, C., Yi, D., Brown, J., Beers, A., et al. (2018b). Distributed deep learning networks among institutions for medical imaging. *J. Am. Med. Inform. Assoc.* 25, 945–954. doi: 10.1093/jamia/ocy017
- Cheng, C. L., Azhar, R., Sng, S. H. A., Chua, Y. Q., Hwang, J. S. G., Chin, J. P. F., et al. (2016). Enabling digital pathology in the diagnostic setting: navigating through the implementation journey in an academic medical centre. *J. Clin. Pathol.* 69, 784–792. doi: 10.1136/jclinpath-2015-203600
- Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., et al. (2013). The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. *J. Digit. Imaging* 26, 1045–1057. doi: 10.1007/s10278-013-9622-7
- Coroller, T. P., Grossmann, P., Hou, Y., Rios, V., Leijenaar, R. T., Hermann, G., et al. (2016). CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma. *Radiother. Oncol.* 114, 345–350. doi: 10.1016/j.radonc.2015.02.015
- Crimi, A., Bakas, S., Kuijff, H., Menze, B., and Reyes, M. (2018). "Brainlesion: glioma, multiple sclerosis, stroke and traumatic brain injuries," in *Proceedings of the Third International Workshop, BrainLes 2017, Held in Conjunction with MICCAI 2017*, Quebec City, QC.
- Dice, L. R. (1945). Measures of the amount of ecologic association between species. *Ecology* 26, 297–302. doi: 10.2307/1932409
- Fabelo, H., Ortega, S., Ravi, D., Kiran, B. R., Sosa, C., Bulters, D., et al. (2018). Spatio-spectral classification of hyperspectral images for brain cancer detection during surgical operations. *PLoS One* 13:e0193721. doi: 10.1371/journal.pone.0193721
- Fan, R.-E., Chang, W. K., Hsieh, C.-J., Wang, X.-R., and Lin, C.-J. (2008). LIBLINEAR: a library for large linear classification. *J. Machine Learn. Res.* 9, 1871–1874. doi: 10.1021/ci100073w
- Foran, D. J., Yang, L., Chen, W., Hu, J., Goodell, L. A., Reiss, M., et al. (2011). ImageMiner: a software system for comparative analysis of tissue microarrays using content-based image retrieval, high-performance computing, and grid technology. *J. Am. Med. Inform. Assoc.* 18, 403–415. doi: 10.1136/amiajnl-2011-000170
- Gao, Y., Ratner, V., Zhu, L., Diprima, T., Kurc, T., Tannenbaum, A., et al. (2016). "Hierarchical nucleus segmentation in digital pathology images," in *Proceedings of the SPIE International Society Optical Engineering*, Washington, DC.
- Gillies, R. (2013). Radiomics: informing cancer heterogeneity. *J. Nucl. Med.* 31, 271–279.
- Gillies, R. J., Kinahan, P. E., and Hricak, H. (2016). Radiomics: images are more than pictures, they are data. *Radiology* 278, 563–577. doi: 10.1148/radiol.2015151169
- Griffin, J., and Treanor, D. (2017). Digital pathology in clinical use: where are we now and what is holding us back? *Histopathology* 70, 134–145. doi: 10.1111/his.12993
- Gurcan, M. N., Boucheron, L., Can, A., Madabhushi, A., Rajpoot, N., and Yener, B. (2009). Histopathological image analysis: a review. *IEEE Rev. Biomed. Eng.* 2:147. doi: 10.1109/RBME.2009.2034865
- He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017). "Mask r-cnn," in *Proceedings of the IEEE International Conference On Computer Vision*, Seattle, 2961–2969.
- Hou, L., Agarwal, A., Samaras, D., Kurc, T. M., Gupta, R. R., and Saltz, J. H. (2019a). "Robust histopathology image analysis: to label or to synthesize?," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Seattle, 8533–8542.
- Hou, L., Nguyen, V., Kanevsky, A. B., Samaras, D., Kurc, T. M., Zhao, T., et al. (2019b). Sparse autoencoder for unsupervised nucleus detection and representation in histopathology images. *Pattern Recogn.* 86, 188–200. doi: 10.1016/j.patcog.2018.09.007
- Hu, Z. L., Tang, J. S., Wang, Z. M., Zhang, K., Zhang, L., and Sun, Q. L. (2018). Deep learning for image-based cancer detection and diagnosis - A survey. *Pattern Recogn.* 83, 134–149. doi: 10.1016/j.patcog.2018.05.014
- Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. (2017). "Densely connected convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Seattle, 4700–4708.
- Ishikawa, Y., Washiya, K., Aoki, K., and Nagahashi, H. (2016). "Brain tumor classification of microscopy images using deep residual learning," in *SPIE Biophotonics Australasia*, eds M. R. Hutchinson, and E. M. Goldys, (Washington, DC: SPIE), 10013.
- Jaccard, P. (1901). Distribution de la flore alpine dans le bassin des Dranses et dans quelques régions voisines. *Bull. Soc. Vaudoise Sci. Nat.* 37, 241–272.
- Kelahan, L. C., Kalaria, A. D., and Filice, R. W. (2017). PathBot: a radiology-pathology correlation dashboard. *J. Digit. Imaging* 30, 681–686. doi: 10.1007/s10278-017-9969-2
- Khosravi, P., Kazemi, E., Imielinski, M., Elemento, O., and Hajirasouliha, I. (2018). Deep convolutional neural networks enable discrimination of heterogeneous digital pathology images. *EBiomedicine* 27, 317–328. doi: 10.1016/j.ebiom.2017.12.026
- Kong, J., Cooper, L. A. D., Wang, F., Gutman, D. A., Gao, J., Chisolm, C., et al. (2011). Integrative, multimodal analysis of glioblastoma using TCGA molecular data, pathology images, and clinical outcomes. *IEEE Trans. Biomed. Eng.* 58, 3469–3474. doi: 10.1109/TBME.2011.2169256
- Kothari, S., Phan, J. H., Osunkoya, A. O., and Wang, M. D. (2012). Biological interpretation of morphological patterns in histopathological whole-slide images. *ACM BCB* 2012, 218–225. doi: 10.1145/2382936.2382964
- Kothari, S., Phan, J. H., Stokes, T. H., and Wang, M. D. (2013). Pathology imaging informatics for quantitative analysis of whole-slide images. *J. Am. Med. Inform. Assoc.* 20, 1099–1108. doi: 10.1136/amiajnl-2012-001540
- Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, eds M. I. Jordan, Y. LeCun, and S. A. Solla, (Cambridge, MA: MIT Press), 1097–1105.
- Lambin, P., Rios-Velazquez, E., Leijenaar, R., Carvalho, S., van Stiphout, R. G., Granton, P., et al. (2012). Radiomics: extracting more information from medical images using advanced feature analysis. *Eur. J. Cancer* 48, 441–446. doi: 10.1016/j.ejca.2011.11.036
- Lee, J. J., Jedrych, J., Pantanowitz, L., and Ho, J. (2018). Validation of digital pathology for primary histopathological diagnosis of routine, inflammatory dermatopathology cases. *Am. J. Dermatopathol.* 40, 17–23. doi: 10.1097/dad.0000000000000888

- Lehrer, M., Powell, R. T., Barua, S., Kim, D., Narang, S., and Rao, A. (2017). "Radiogenomics and histomics in glioblastoma: the promise of linking image-derived phenotype with genomic information," in *Advances in Biology and Treatment of Glioblastoma*, ed. K. Somasundaram, (Berlin: Springer), 143–159. doi: 10.1007/978-3-319-56820-1\_6
- Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., et al. (2014). "Microsoft coco: common objects in context," in *Proceedings of the European Conference on Computer Vision*, (Berlin: Springer), 740–755. doi: 10.1007/978-3-319-10602-1\_48
- Liu, F. T., Ting, K. M., and Zhou, Z.-H. (2008). "Isolation forest," in *Proceedings of the 2008 Eighth IEEE International Conference on Data Mining*, (Amsterdam: IEEE), 413–422.
- Lu, S. Y., Lu, Z. H., and Zhang, Y. D. (2019). Pathological brain detection based on AlexNet and transfer learning. *J. Comput. Sci.* 30, 41–47. doi: 10.1016/j.jocs.2018.11.008
- Lundstrom, C. F., Gilmore, H. L., and Ros, P. R. (2017). Integrated diagnostics: the computational revolution catalyzing cross-disciplinary practices in radiology. *Pathol. Genomics Radiol.* 285, 12–15. doi: 10.1148/radiol.2017170062
- Madabhushi, A., and Lee, G. (2016). Image analysis and machine learning in digital pathology: challenges and opportunities. *Med. Image Anal.* 33, 170–175. doi: 10.1016/j.media.2016.06.037
- Mahmood, F., Borders, D., Chen, R., McKay, G. N., Salimian, K. J., Baras, A., et al. (2018). Deep adversarial training for multi-organ nuclei segmentation in histopathology images. *IEEE Trans. Med. Imaging* 36, 1–10. doi: 10.1109/TMI.2019.2927182
- Manivannan, S., Li, W., Akbar, S., Wang, R., Zhang, J., and McKenna, S. J. (2016). An automated pattern recognition system for classifying indirect immunofluorescence images of HEP-2 cells and specimens. *Pattern Recogn.* 51, 12–26. doi: 10.1016/j.patcog.2015.09.015
- McGarry, S. D., Hurrell, S. L., Iczkowski, K. A., Hall, W., Kaczmarowski, A. L., Banerjee, A., et al. (2018). Radio-pathomic maps of epithelium and lumen density predict the location of high-grade prostate cancer. *Int. J. Radiat. Oncol. Biol. Phys.* 101, 1179–1187. doi: 10.1016/j.ijrobp.2018.04.044
- Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., et al. (2014). The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Trans. Med. Imaging* 34, 1993–2024.
- Mobadersany, P., Yousefi, S., Amgad, M., Gutman, D. A., Barnholtz-Sloan, J. S., Velázquez Vega, J. E., et al. (2018). Predicting cancer outcomes from histology and genomics using convolutional networks. *Proc. Natl. Acad. Sci. U.S.A.* 115, E2970–E2979. doi: 10.1073/pnas.1717139115
- Momeni, A., Thibault, M., and Gevaert, O. (2018). "Dropout-enabled ensemble learning for multi-scale biomedical data," in *Proceedings of the International MICCAI Brainlesion Workshop*, (Berlin: Springer), 407–415. doi: 10.1007/978-3-030-11723-8\_41
- Parmar, C., Grossmann, P., Bussink, J., Lambin, P., and Aerts, H. J. (2015). Machine learning methods for quantitative radiomic biomarkers. *Sci. Rep.* 5:13087. doi: 10.1038/srep13087
- Peikari, M., and Martel, A. L. (2016). "Automatic cell detection and segmentation from H and E stained pathology slides using colorspace decorrelation stretching," in *Proceedings of the Medical Imaging 2016: Digital Pathology*, International Society for Optics and Photonics, Vol. 9791, Houston, TX, 979114.
- Peikari, M., Salama, S., Nofech-Mozes, S., and Martel, A. L. (2018). A cluster-then-label semi-supervised learning approach for pathology image classification. *Sci. Rep.* 8:7193. doi: 10.1038/s41598-018-24876-0
- Penny, W. D., Friston, K. J., Ashburner, J. T., Kiebel, S. J., and Nichols, T. E. (2011). *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Amsterdam: Elsevier.
- Prior, F. W., Clark, K., Commey, P., Freymann, J., Jaffe, C., Kirby, J., et al. (2013). "TCIA: an information resource to enable open science," in *Proceedings of the 2013 35th Annual International Conference of the IEEE Engineering in Medicine (EMBC)*, (Piscataway, NJ: IEEE), 1282–1285.
- Qi, Q., Li, Y., Wang, J., Zheng, H., Huang, Y., Ding, X., et al. (2018). Label-efficient breast cancer histopathological image classification. *IEEE J. Biomed. Health Inform.* 23, 2108–2116. doi: 10.1109/JBHI.2018.2885134
- Qian, Z., Li, Y., Wang, Y., Li, L., Li, R., Wang, K., et al. (2019). Differentiation of glioblastoma from solitary brain metastases using radiomic machine-learning classifiers. *Cancer Lett.* 451, 128–135. doi: 10.1016/j.canlet.2019.02.054
- Reinhard, E., Adhikhmin, M., Gooch, B., and Shirley, P. (2001). Color transfer between images. *IEEE Comput. Graph. Appl.* 21, 34–41.
- Ren, S., He, K., Girshick, R., and Sun, J. (2015). "Faster r-cnn: towards real-time object detection with region proposal networks," in *Advances in neural information processing systems*, eds M. I. Jordan, Y. LeCun, and S. A. Solla, (Cambridge, MA: MIT Press), 91–99.
- Ronneberger, O., Fischer, P., and Brox, T. (2015). "U-net: convolutional networks for biomedical image segmentation," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, (Berlin: Springer), 234–241. doi: 10.1007/978-3-319-24574-4\_28
- Saltz, J., Almeida, J., Gao, Y., Sharma, A., Bremer, E., DiPrima, T., et al. (2017). Towards generation, management, and exploration of combined radiomics and pathomics datasets for cancer research. *AMIA Jt Summits Transl. Sci. Proc.* 2017, 85–94.
- Saltz, J., Gupta, R., Hou, L., Kurc, T., Singh, P., Nguyen, V., et al. (2018). Spatial organization and molecular correlation of tumor-infiltrating lymphocytes using deep learning on pathology images. *Cell Rep.* 23:181–193.e7. doi: 10.1016/j.celrep.2018.03.086
- Senaras, C., Niazi, M. K. K., Lozanski, G., and Gurcan, M. N. (2018a). DeepFocus: detection of out-of-focus regions in whole slide digital images using deep learning. *PLoS One* 13:e0205387. doi: 10.1371/journal.pone.0205387
- Senaras, C., Niazi, M. K. K., Sahiner, B., Pennell, M. P., Tozbikian, G., Lozanski, G., et al. (2018b). Optimized generation of high-resolution phantom images using cGAN: application to quantification of Ki67 breast cancer images. *PLoS One* 13:e0196846. doi: 10.1371/journal.pone.0196846
- Sheller, M. J., Reina, G. A., Edwards, B., Martin, J., and Bakas, S. (2018). "Multi-institutional deep learning modeling without sharing patient data: a feasibility study on brain tumor segmentation," in *Proceedings of the International MICCAI Brainlesion Workshop*, (Berlin: Springer), 92–104. doi: 10.1007/978-3-030-11723-8\_9
- Shukla, G., Alexander, G. S., Bakas, S., Nikam, R., Talekar, K., Palmer, J. D., et al. (2017). Advanced magnetic resonance imaging in glioblastoma: a review. *Chin. Clin. Oncol.* 6:40. doi: 10.21037/cco.2017.06.28
- Siegel, R. L., Miller, K. D., and Jemal, A. (2016). Cancer statistics, 2016. *CA* 66, 7–30. doi: 10.3322/caac.21332
- Siegel, R. L., Miller, K. D., and Jemal, A. (2019). Cancer statistics, 2019. *CA* 69, 7–34.
- Simonyan, K., and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv [preprint]*
- Tomczak, K., Czerwińska, P., and Wiznerowicz, M. (2015). The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemp. Oncol.* 19:A68. doi: 10.5114/wo.2014.47136
- van Griethuysen, J., Fedorov, A., Aucoin, N., Fillion-Robin, J.-C., Hosny, A., Pieper, S., et al. (2016). *Welcome to Pyradiomics Documentation*. Available at: <https://pyradiomics.readthedocs.io/en/latest/> (accessed February 7, 2020).
- van Griethuysen, J. J., Fedorov, A., Parmar, C., Hosny, A., Aucoin, N., Narayan, V., et al. (2017). Computational radiomics system to decode the radiographic phenotype. *Cancer Res.* 77, e104–e107. doi: 10.1158/0008-5472.CAN-17-0339
- Vu, Q. D., Graham, S., Kurc, T., To, M. N. N., Shaban, M., Qaiser, T., et al. (2019). Methods for segmentation and classification of digital microscopy tissue images. *Front. Bioeng. Biotechnol.* 7:53. doi: 10.3389/fbioe.2019.00053
- Wang, X. Y., Wang, D. Q., Yao, Z. G., Xin, B. W., Wang, B., Lan, C. J., et al. (2019). Machine learning models for multiparametric glioma grading with quantitative result interpretations. *Front. Neurosci.* 12:1046. doi: 10.3389/fnins.2018.01046
- Wen, S., Kurc, T. M., Gao, Y., Zhao, T., Saltz, J. H., and Zhu, W. (2017). A methodology for texture feature-based quality assessment in nucleus segmentation of histopathology image. *J. Pathol. Inform.* 8:38. doi: 10.4103/jpi.jpi\_43\_17
- Wen, S., Kurc, T. M., Hou, L., Saltz, J. H., Gupta, R. R., Batiste, R., et al. (2018). Comparison of different classifiers with active learning to support quality control in nucleus segmentation in pathology images. *AMIA Jt Summits Transl. Sci. Proc.* 2017, 227–236.
- Wollmann, T., Bernhard, P., Gunkel, M., Braun, D. M., Meiners, J., Simon, R., et al. (2019). "Black-box hyperparameter optimization for nuclei segmentation in prostate tissue images," in *Bildverarbeitung für die Medizin 2019*, eds H.

- Handels, T. M., Deserno, A., Maier, K. H., Maier-Hein, and C. Palm, (Berlin: Springer), 345–350. doi: 10.1007/978-3-658-25326-4\_75
- Xie, L., and Li, C. (2018). “Simultaneous detection and segmentation of cell nuclei based on convolutional neural network,” in *Proceedings of the 2nd International Symposium on Image Computing and Digital Medicine*, (Chengdu: ACM), 129–132.
- Xing, F., and Yang, L. (2016). Robust nucleus/cell detection and segmentation in digital pathology and microscopy images: a comprehensive review. *IEEE Rev. Biomed. Eng.* 9, 234–263. doi: 10.1109/RBME.2016.2515127
- Yang, Q., Wu, K., Cheng, H., Gu, C., Liu, Y., Casey, S. P., et al. (2018). “Cervical nuclei segmentation in whole slide histopathology images using convolution neural network,” in *Proceedings of the International Conference on Soft Computing in Data Science*, (Berlin: Springer), 99–109. doi: 10.1007/978-981-13-3441-2\_8
- Yonekura, A., Kawanaka, H., Prasath, V. B. S., Aronow, B. J., and Tsuruoka, S. (2018). “Glioma subtypes clustering method using histopathological image analysis,” in *Proceedings of the 7th International Conference on Informatics, Electronics & Vision (Iciev) and 2018 2nd International Conference on Imaging, Vision & Pattern Recognition (Icivpr)*, 2018, Fukuoka, 442–446.
- Yu, K.-H., Zhang, C., Berry, G. J., Altman, R. B., Ré, C., Rubin, D. L., et al. (2016). Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. *Nat. Commun.* 7:12474. doi: 10.1038/ncomms12474
- Yuan, Y., Shi, Q., Li, M., Nagamuthu, C., Andres, E., and Davis, F. G. (2016). Canadian brain cancer survival rates by tumour type and region: 1992–2008. *Can. J. Public Health* 107, e37–e42. doi: 10.17269/cjph.107.5209
- Zhou, M., Scott, J., Chaudhury, B., Hall, L., Goldgof, D., Yeom, K. W., et al. (2018). Radiomics in brain tumor: image assessment, quantitative feature descriptors, and machine-learning approaches. *Am. J. Neuroradiol.* 39, 208–216. doi: 10.3174/ajnr.A5391
- Zwanenburg, A., Leger, S., Vallières, M., and Löck, S. (2016). Image biomarker standardisation initiative. *arXiv [preprint]*
- Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Kurc, Bakas, Ren, Bagari, Momeni, Huang, Zhang, Kumar, Thibault, Qi, Wang, Kori, Gevaert, Zhang, Shen, Khened, Ding, Krishnamurthi, Kalpathy-Cramer, Davis, Zhao, Gupta, Saltz and Farahani. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.