CHAPTER

5

Simultaneous brain tumor segmentation and molecular profiling using deep learning and T2w magnetic resonance images

Chandan Ganesh Bangalore Yogananda[a], Bhavya R. Shah[a], Fang F. Yu[a], Sahil S. Nalawade[a], James Holcomb[a], Divya Reddy[a], Benjamin C. Wagner[a], Marco C. Pinho[a], Bruce Mickey[b], Toral R. Patel[b], Baowei Fei[c], Ananth J. Madhuranthakam[a] and Joseph A. Maldjian[a]

[a] *Advanced Neuroscience Imaging Research (ANSIR) Lab, Department of Radiology, University of Texas Southwestern Medical Center, Dallas, TX, United States of America,* [b] *Department of Neurological Surgery, University of Texas Southwestern Medical Center, Dallas, TX, United States of America.,* [c] *Department of Bioengineering, University of Texas at Dallas, Richardson, TX, United States of America.*

5.1 Introduction

In the final section of his seminal paper, "Computing Machinery and Intelligence," Alan Turing describes his "learning machine" as being capable of acting indistinguishably from a human [1]. Lady Lovelace raises the criticism that computers are inherently incapable of originality. Turing's response to this criticism is that this is a matter of training and that by the end of the century a "computing machine" will be able to outperform a human. This academic repartee is widely held to be the origins of machine learning. Approximately 70 years later, trained machines are outperforming humans in various aspects of medical imaging analysis. One such area is glioma segmentation and subsequent molecular profiling based on magnetic resonance imaging (MRI). The new World Health Organization glioma classification system relies heavily on a tumor's genetic profile. As such, molecular profiling has now become a critical step in predicting aggressiveness and response to therapy.

Gliomas are a diverse set of brain tumors that represent the most common primary brain malignancy. Gliomas are either low-grade or high-grade and the grading is dependent on histopathology, immunohistochemistry, and genetic profiling. Imaging correlates of histopathological findings are the basis for image-based segmentation of gliomas. The segmentation routines typically classify MRI findings into either enhancing/active tumor, necrotic core, or surrounding edema. The ability of these

Brain Tumor MRI Image Segmentation Using Deep Learning Techniques.
DOI: https://doi.org/10.1016/B978-0-323-91171-9.00005-3

algorithms to differentiate MRI signal of these subcomponents is what determines the accuracy of image-based glioma segmentation. Machine learning algorithms outperform human *expert* manual brain tumor segmentation by decreasing inter- and intrarater variabilities. Accurate image-based segmentation of gliomas is an important prerequisite to subsequent image-based molecular profiling. Numerous machine learning algorithms have demonstrated improved glioma segmentation with decreased variability and improvements on the time required for manual segmentation [2–4]. In this chapter, we will discuss the topic of molecular profiling in gliomas, and in particular, our work in developing convolutional neural networks (3-D Dense-Unets) for MRI -based glioma segmentation and subsequent molecular profiling. Unlike other machine learning methods, our approach facilitates clinical translation by requiring only T2w images.

5.1.1 Glioma molecular markers

In 2008 it was found that a mutation in the citric acid cycle enzyme isocitrate dehydrogenase (IDH) of some malignant brain tumors (glioblastomas) resulted in an improved prognosis compared to gliomas with the wild-type form [5]. In 2016, the World Health Organization revised their glioma classification establishing a revolution in clinical glioma evaluation [6]. At multidisciplinary tumor board meetings across the country, the knowledge of molecular status of gliomas has moved to the forefront in clinical decision-making. The first question asked is whether the tumor is IDH mutated, followed quickly by whether the tumor is 1p/19q codeleted, and if the Methylguanine-DNA Methyltransferase (MGMT) promoter is methylated. Codeletion status of 1p/19q is the most important genomic marker for oligodendrogliomas. Gliomas with 1p/19q codeletion demonstrate improved prognosis and treatment response compared to gliomas without it. Methylation of MGMT promoter silences the DNA repair MGMT enzyme epigenetically. This results in improved response to treatment and prognosis. The only way to definitively identify these markers is to obtain a tissue specimen acquired through biopsy or surgery and perform immunohistochemistry or gene sequencing. Differences in the mutation status for these markers have vital treatment implications. Thus, there is significant interest in determining these molecular profiles prior to surgical resection and therapy initiation. This is more important for gliomas that harbor significant risks to biopsy or surgery due to their location. Although the molecular profiling of gliomas is now a routine part of the work-up, it would be helpful to have this information prior to surgery. In some cases, the information would aid in planning the extent of tumor resection. For tumors in locations where resection is not possible, and the risk of a biopsy is high, accurate delineation of the molecular and genetic profile of the tumor might be used to guide empiric treatment with radiation and/or chemotherapy. While there are many new glioma molecular markers; IDH, 1p/19q, and MGMT are critical for prognosis and therapeutic planning. Highly accurate, noninvasive determination of these molecular markers would be transformational in the management of gliomas. Image-based

molecular profiling would reduce or eliminate the risks and costs associated with a neurosurgical procedure and decrease the time to definitive treatment.

5.1.2 2HG spectroscopy for IDH status

The oncometabolite 2-hydroxyglutarate (2-HG) production is catalyzed by a mutant form of the IDH enzyme [7]. Magnetic Resonance (MR) spectroscopy has been used for noninvasive determination of brain tumor 2-HG [8–11]. Clinical translation, however, has been hindered by long scan times, challenges in appropriate patient selection based on tumor morphology, adequate voxel positioning, and the need for local expertise for acquisition and processing. The time allotted for integration into a clinical protocol is frequently insufficient to acquire a good quality spectrum due to poor signal-to-noise ratio, patient motion, and additional shimming needed that cannot be performed in the routine clinical setting. Even when good quality spectra are obtained, 2-HG spectroscopy still carries a high false-positive rate of over 20% [12]. No MR spectroscopic methods are available for the other glioma molecular markers.

5.1.3 TCIA database

Work in brain tumor classification has been facilitated by The Cancer Imaging Archive (TCIA), an open-access archive of medical imaging studies for cancer research [13]. The data come from multiple clinical sites across the country. The Cancer Genome Atlas (TCGA) is a separate database with molecular characterization of a large number of tumors. Many of the cases in the TCIA database are linked by a common study ID to the TCGA database, allowing open access to imaging data, corresponding molecular marker data, and clinical data in the same subjects available through the National Cancer Institute—Genomic Data Commons (GDC) Data Portal [14]. There are two particular TCIA collections of interest to this work: The low-grade glioma (LGG) and the high-grade glioma (HGG) collections of brain tumors. Together these represent a dataset of several hundred subjects.

5.1.4 MRI and deep learning for glioma molecular markers

Recent studies have employed machine learning and deep learning with MRI to determine molecular marker status, with accuracies ranging from 80% to 93%, many using the TCIA database [15–23]. Fluid-attenuated inversion recovery (FLAIR) MR images along with genomic information have been be used to classify IDH mutation status into wild type, and IDH mutated with and without 1p/19q co-deletion [15]. A support vector machine algorithm trained using 103 LGG subjects was able to achieve an area under the curve (AUC) of 0.83 for classifying IDH status [17]. A residual deep learning network for predicting IDH status with multicontrast MRI was developed utilizing preprocessing steps including tumor segmentation, coregistration, resampling, N4biascorrection, and intensity normalization [16]. A separate group demonstrated that IDH mutation status could be determined using T2w, T2w-FLAIR and T1-weighted (T1w) pre- and postcontrast images [18]. Several 1p/19q deep

Table 5.1 T2-net and TS-net cross-validation results.

Fold description Fold number	T2-net		TS-net	
	% accuracy	Dice-score	% accuracy	Dice-score
1	97.18	0.843	97.22	0.88
2	97.14	0.86	97.10	0.883
3	97.10	0.857	97.05	0.892
Average	**97.14 ± 0.04**	**0.853 ± 0.009**	**97.12 ± 0.08**	**0.885 ± 0.006**

Table 5.2 1p/19q-net cross-validation results.

Fold description Fold number	1p/19q-net		
	% accuracy	AUC	Dice-score
1	93.4	0.9571	0.8151
2	94.35	0.9688	0.8057
3	92.62	0.9351	0.8000
Average	**93.46 ± 0.86**	**0.953 ± 0.01**	**0.801 ± 0.007**

learning algorithms developed for LGGs achieved accuracies of 65.9%–87.7% [24–26]. Unfortunately, these methods have limited clinical viability, require extensive preprocessing including manual tumor presegmentation, and multicontrast MR image acquisitions that can have motion from longer scanning. Multiple radiomic approaches have also been attempted, but most have poor performance compared to the deep learning methods, and none have achieved tissue-level accuracies for clinical viability.

5.1.5 T2w versus multicontrast MRI data

Several factors can compromise multicontrast imaging data including (1) motion, (2) long acquisition times, and (3) the use of intravenous gadolinium. Considering the importance of a clinically implementable technique, the use of only T2w images is ideal, as T2w MRI are universally obtained during clinical evaluation of brain tumors. High-quality T2w images can be acquired within 2 min and are robust to patient motion [27,28]. Reliable, accurate, and noninvasive classification of molecular marker status can enhance treatment planning of brain tumors, *especially when biopsy or resection is not an option*.

5.2 Summary of the methods

In this section, we summarize our previously reported methods for classification of IDH mutation status, 1p/19q codeletion status, and MGMT methylation status using deep learning and the TCIA-TCGA database [29–31].

Table 5.3 MGMT-net cross-validation results.

Fold description Fold number	MGMT-net		
	% accuracy	Fold number	% accuracy
1	95.12	1	95.12
2	93.98	2	93.98
3	95.12	3	95.12
Average	**94.73 ± 0.66**	**Average**	**94.73 ± 0.66**

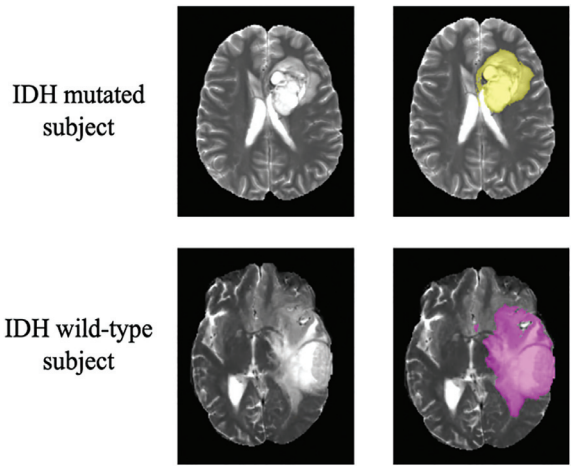

FIGURE 5.1

Yellow voxels represent IDH mutated (value of 1) and purple voxels represent IDH wild-type (value of 2). The ground truth labels have the same mutation status for all voxels in each tumor.

5.2.1 Classification of IDH mutation status

5.2.1.1 Material and methods

5.2.1.1.1 Data for the study

Multiparametric brain MRI data from TCIA were used to train several deep learning networks [13]. Corresponding genomic information was provided from the TCGA [14]. Studies were evaluated for the availability of T2w, T2w-FLAIR, and contrast enhanced T1-weighted (T1c) image series along with IDH status. Only preoperative studies were used. The final dataset included 214 subjects (94 IDH mutated, 120 IDH wild-type). Tumor masks for the datasets were available through previous expert segmentation and were used as the ground truth for the tumor segmentation in the training set [32]. Ground truth tumor masks for IDH mutated and wild-types were labeled with 1s and 2s, respectively (Fig. 5.1). Two separate networks were developed. These included a T2w image only network (T2w-net) trained only on the T2w images,

and a 3-sequence network (TS-net) trained on multicontrast MR data including T2w images, T2w-FLAIR, and T1c.

5.2.1.1.2 Preprocessing steps

The preprocessing steps included, (1) coregistering the multiparametric images to the T1C (for TS-net), (2) affine registration to the SRI24 template with resampling to 1 mm isotropic resolution, (3) N4BiasCorrection, (4) skull stripping to isolate the brain volume, and (5) intensity normalization.[32–34] Preprocessing took approximately 5 min per subject using advanced normalization tools [35].

5.2.1.1.3 Network details

Both the networks were trained and tested using a 3-D patch-based ($32 \times 32 \times 32$) approach. 3-D Dense UNets were developed and trained to perform a voxel-wise two-class whole tumor segmentation with classes 1 and 2 depicting the IDH mutated and wild type, respectively. This approach represents a significant departure from previous image-based IDH classification schemes which typically provide only the segmented tumor to the network followed by radiomic feature extraction of the tumor voxels or a CNN applied only to the segmented tumor. Our approach recasts the classification problem as a segmentation problem on the entire image. The schematics of the T2-net are shown in Fig. 5.2.

The dense connections were implemented in two steps including, (1) the input and output of a particular dense block were concatenated, and (2) feature maps generated from all the layers within a dense block were interconcatenated. The encoder portion of the network consisted of a combination of dense blocks and transition down block, whereas the decoder portion of the network was a combination of dense blocks and transition up blocks. The network also consisted of a bottleneck block that helped minimize the convolution layers. High-resolution information from the encoder to decoder was passed through skip connections. Such connections ensured that all the generated feature maps were reused while providing a direct supervision signal to every layer in the architecture [36].

5.2.1.1.4 Cross-validation

To evaluate the networks' performance and reliability, a cross-validation (threefold) was implemented on each network separately by shuffling and splitting the dataset into three equal groups ($\sim$70 subjects in each group). Each cross-validation fold demonstrates a separate training session on a distinctive combination of the three groups alternating between training, in-training validation, and held-out testing data. Seventy-five percent overlapping 3-D patches (size: $32 \times 32 \times 32$) were extracted from the training and in-training validation dataset and used to train the networks. Data augmentation included vertical flipping, horizontal flipping, random rotation, translational rotation, and down sampling. This provided $\sim$150,000 patches for both training and in-training validation separately and avoided issues with overfitting. Patches from the same subject were not mixed within the three groups. This avoided the problems of data leakage and data duplication among training and testing datasets,

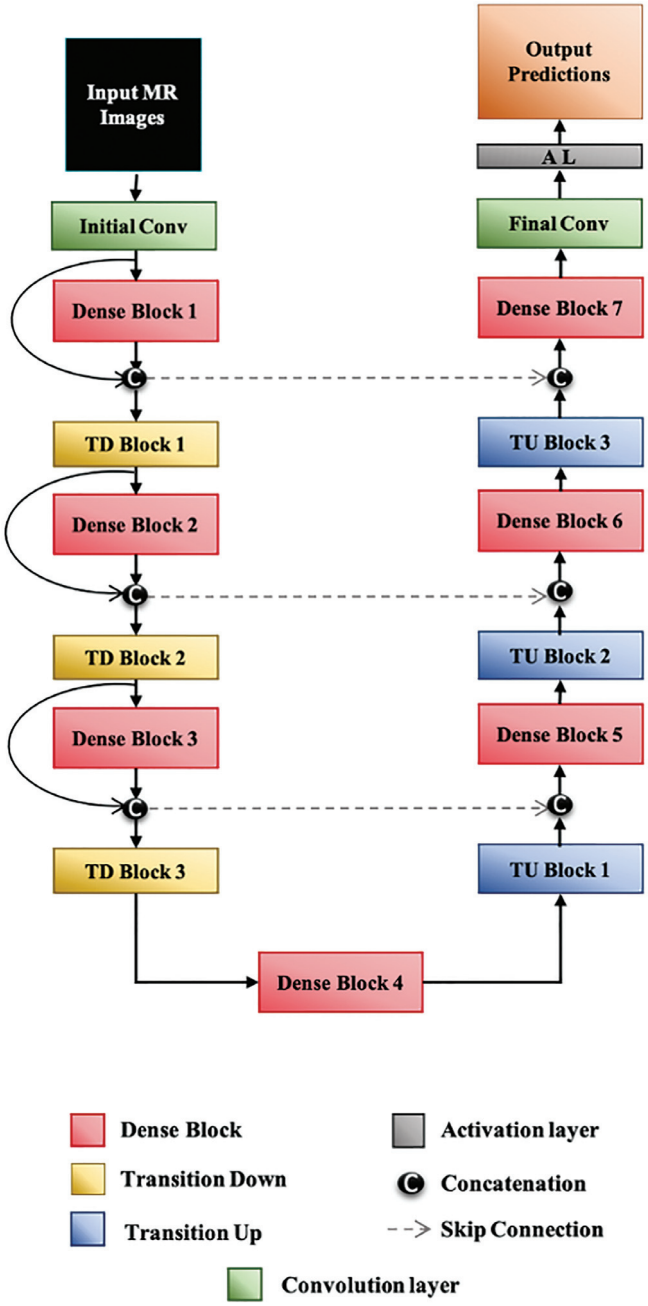

FIGURE 5.2

Network architecture for both T2-net and TS-net IDH networks.

a problem that is commonly encountered in 2-D networks [37,38]. The networks were developed using keras python packages, Tensorflow backend engines, an Adam optimizer, and pycharm Integrated Development Environment (IDE) [39–41]. The networks were trained on a Tesla P100 NVIDIA-GPU with a batch size of 4, an initial learning rate of 10^{-5}, and maximal iterations of 100. Majority-voting over IDH-mutated and IDH-wild-type voxel-wise classes were used to determine each subject's IDH mutation status. Additionally, an ROC curve (receiver operating characteristic curve) was plotted using the subject-wise IDH classification. A dual-volume fusion (DVF) (Fig. 5.3) method was designed in MATLAB using the largest connected component method (3-D connected component) that provided a single label whole tumor segmentation volume (Fig. 5.4). The majority-voting (50% threshold) was used to obtain the accuracy, specificity, sensitivity, negative predictive value (NPV), and positive predictive value (PPV).

5.2.1.2 Results

T2w-net achieved an average cross-fold validation accuracy of 97.1 ± 0.04%, sensitivity of 97 ± 0.03%, specificity of 98 ± 0.01%, PPV of 98 ± 0.01%, NPV of 97 ± 0.01%, and AUC of 0.98 ±0.01. The multicontrast TS-net demonstrated similar accuracy of 97 ±0.09%, a sensitivity of 98 ±0.02%, specificity of 97 ±0.001%, PPV of 97 ± 0.002%, NPV of 97 ±0.001%, and AUC of 0.99 ±0.01, (see Table 5.1). Fig. 5.5 demonstrates the ROC curves for the two networks with very similar performance. The performance of these networks exceeds what has been reported in the literature [15,16,18,42].

5.2.2 Classification of 1p/19q codeletion status

Codeletion status of 1p/19q is a key genomic marker for oligodendrogliomas. Gliomas with 1p/19q codeletion demonstrate improved prognosis and treatment response [43,44]. Using the previously trained 3-D-IDH network (T2w-net), a transfer learning approach was implemented for classifying the codeletion status of 1p/19q. We demonstrated a remarkable 93% accuracy in determination of codeletion status, outperforming any prior reports on noninvasive 1p/19q prediction, approaching histological/molecular accuracies and rivaling Fluorescence in situ hybridization (FISH). [30].

5.2.2.1 Material and methods

5.2.2.1.1 Data and preprocessing

A total of 368 T2w brain MR images and corresponding genomic information of glioma patients were obtained from TCIA) and TCGA databases, respectively [13,14]. The dataset included 130 1p/19q codeleted and 238 1p/19q non-codeleted cases. Expert segmented tumor masks for these datasets were used as ground truth in training the tumor segmentation step [32]. Ground truth tumor masks for the codeleted and

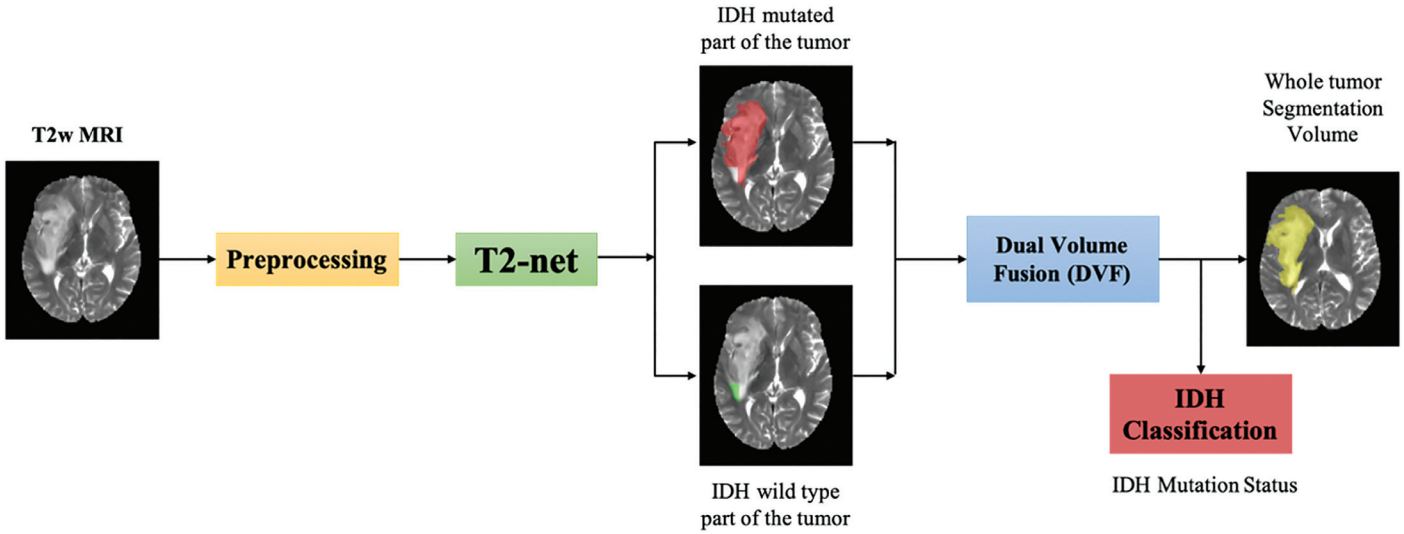

FIGURE 5.3

T2-net overview. Input T2 image is preprocessed then provided to T2-net. Voxel-wise classification is performed for IDH-mutated and IDH-wild type. A dual-volume fusion step is performed to combine the volumes and eliminate outliers from the main tumor cluster. Majority voting is then performed over the two classes to determine the final IDH mutation status.

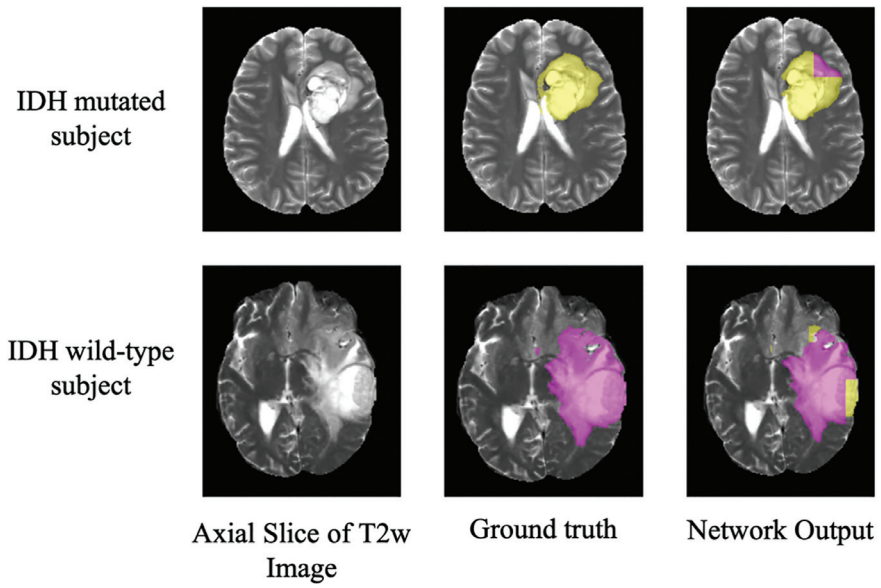

FIGURE 5.4

Native T2 image (A). Ground truth segmentation (B). Network output without DVF (C). Yellow voxels correspond to IDH-mutated class and purple voxels correspond to IDH wild-type.

non-codeleted types were labeled with 1s and 2s, respectively, similar to the IDH network. Data preprocessing steps were the same as for the IDH network.

5.2.2.1.2 Network details and cross-validation

Transfer learning was implemented on the trained 3-D IDH network (T2w-net), by fine-tuning only the network's decoder portion for a voxel-wise two-class segmentation of the whole tumor. The network hyperparameters and training optimizer used during the transfer learning were kept constant from the IDH study. The transfer learning was implemented such that classes 1 and 2 represent the codeleted and the non-codeleted type, respectively. This was used for the final classification of 1p/19q codeletion status of an individual subject. The network architecture is shown in Fig. 5.6.

A threefold cross-validation procedure identical to the IDH study was implemented with each group consisting of ~122 subjects. The patch extraction and primary data augmentation procedures were similar to the IDH study. However, additional data augmentation steps included projective transformation, and addition of salt and pepper noise and Gaussian noise. DVF and statistical analysis for evaluating network (majority voting, accuracy, specificity, sensitivity, NPV, PPV, and ROC curve) were also kept the same as the IDH study.

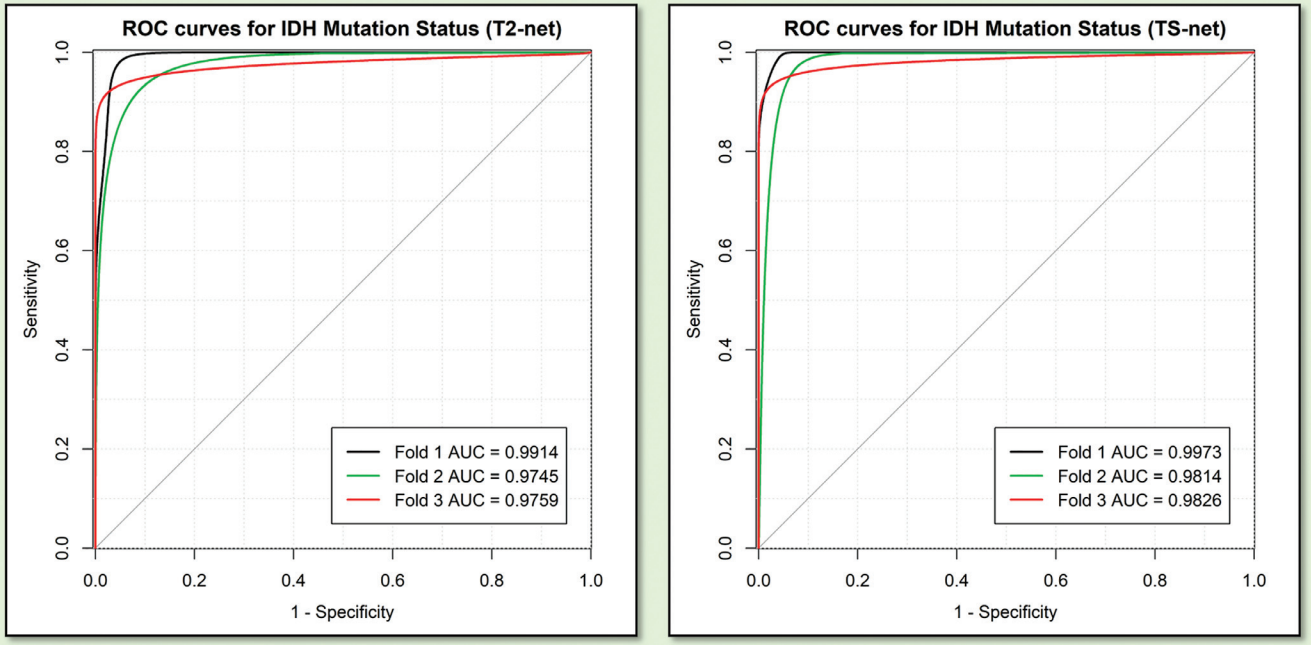

FIGURE 5.5

ROC curves for T2-net and TS-net.

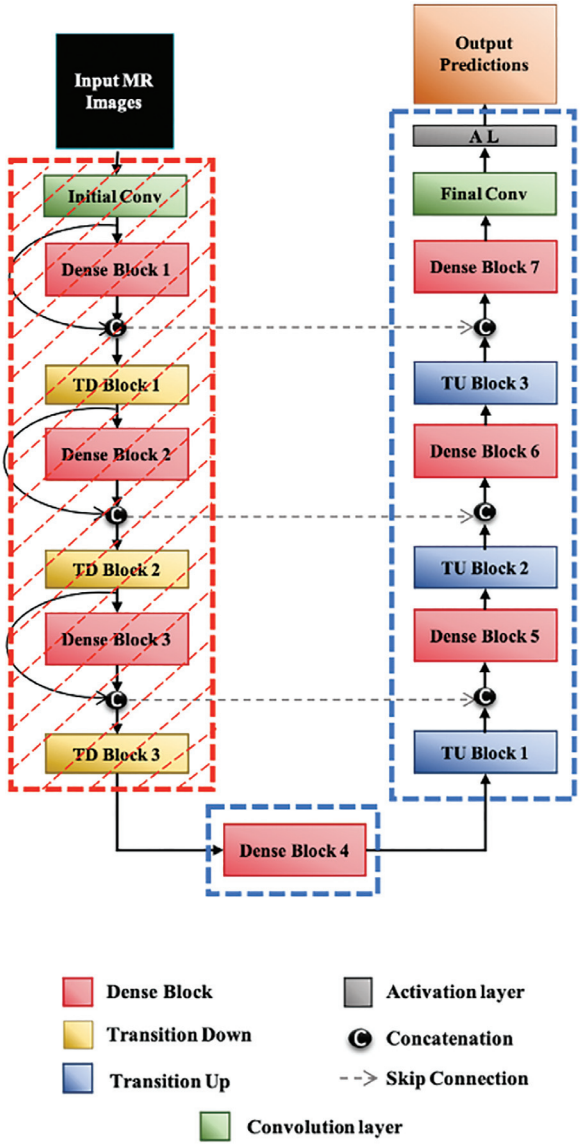

FIGURE 5.6

Network architecture for the 1p/19q-net. The previously trained 3-D IDH network was used. The left arm of the Dense U-net (striped red box) is the encoder part of the network, the right arm of the network (blue box) is the decoder part. The encoder part of the network was frozen to retain the pretrained weights from the 3-D IDH network. The decoder part of the network was fine-tuned for a dual class segmentation with class 1 representing 1p/19q codeleted type and class 2 representing 1p/19q non codeleted type.

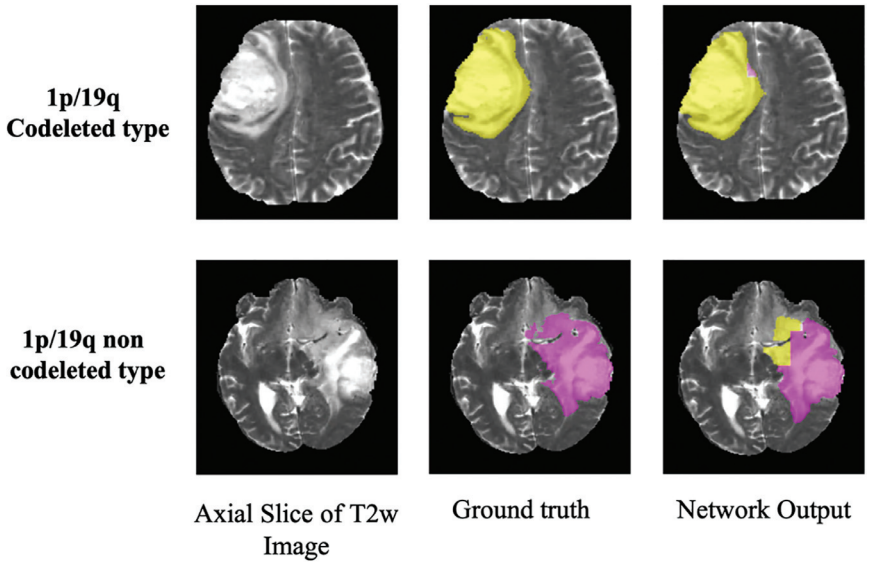

FIGURE 5.7

Voxel-wise segmentations of 1p/19q-net. Native T2 image (A). Ground truth segmentation (B). Network output without DVF (C). Yellow voxels correspond to 1p/19q codeleted class and purple voxels correspond to 1p/19q non-codeleted class.

5.2.2.2 Results

The 1p/19q-net demonstrated mean subject-wise cross-validation accuracy of 93.4%, a sensitivity of 0.90 ±0.003, specificity of 0.95 ±0.01, PPV of 0.91 ±0.02, NPV of 0.95 ±0.0003 and AUC of 0.95 ±0.01, (see Table 5.2). Fig. 5.8 demonstrates the ROC curves for the 1p/19q network. Fig. 5.7 depicts the Voxel-wise segmentations of 1p/19q-net.

5.2.3 MGMT methylation promoter

Methylation of the O^6- MGMT promoter silences the DNA repair MGMT enzyme epigenetically resulting in improved prognosis and response to treatment [45]. Again, using our previously trained 3-D-IDH network (T2w-net), we implemented a transfer learning approach for determining the MGMT promoter status using only T2w MRI from the TCIA database [31]. We demonstrated 95% mean cross-validation accuracy with sensitivity and specificity of 96% and 92%, respectively, *rivaling* the accuracy, sensitivity, and specificity of invasive histological methods such as immunofluorescence, methylation-specific Polymerase Chain Reaction (PCR), and pyrosequencing.

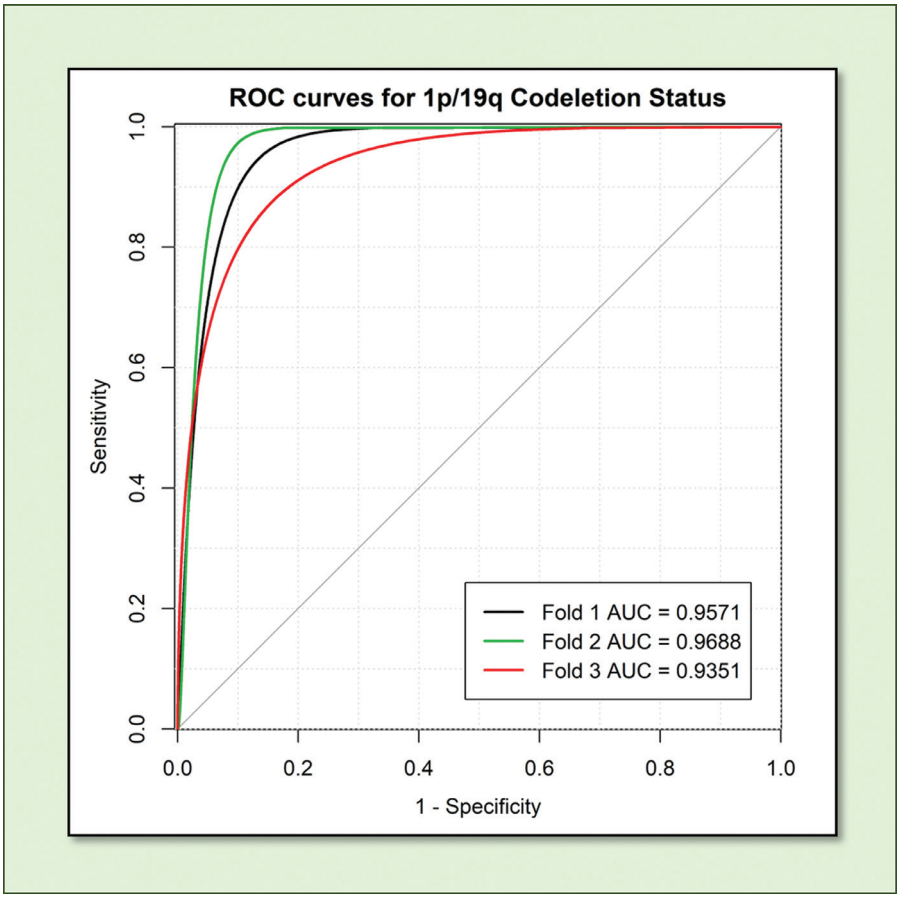

FIGURE 5.8

ROC curves for 1p/19q cross-validation.

5.2.3.1 Material and methods

5.2.3.1.1 Data and preprocessing

T2w brain MR images and corresponding genomic information of 247 glioma patients were retrieved from the TCIA and TCGA databases, respectively [13,14]. The dataset consisted of 163 MGMT methylated and 84 MGMT unmethylated subjects. Expert segmented tumor masks for these datasets were used as ground truth in training the tumor segmentation step [32]. Ground truth tumor masks for MGMT methylated and unmethylated types were labeled with 1s and 2s, respectively similar to the IDH network. Data preprocessing steps were also the same.

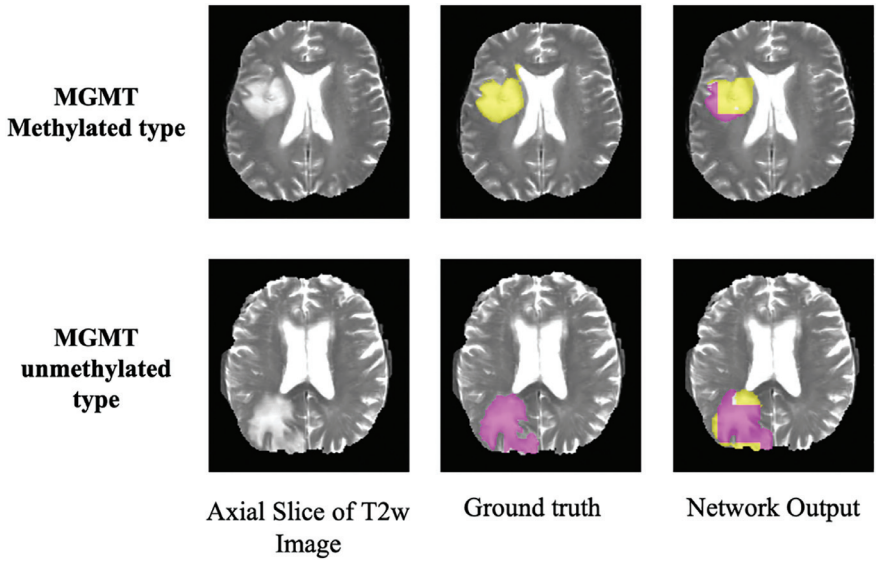

FIGURE 5.9

Example of voxel-wise segmentations of MGMT-net. Native T2 image (A). Ground truth segmentation (B). Network output without DVF (C). Yellow voxels correspond to MGMT methylated class and purple voxels correspond to MGMT unmethylated class.

5.2.3.1.2 Network details and cross-validation

A transfer learning approach identical to the 1p/19q study was used based on the 3-D IDH network (T2-net) by fine-tuning only the decoder part of the network for a voxel-wise two-class whole tumor segmentation. The hyperparameters and training optimizer were the same as the IDH and 1p/19q studies. Classes 1 and 2 of the transfer learning represent the methylated and the unmethylated types, respectively. This was used for the final classification of MGMT promoter status for each individual subject. The network architecture was similar to the 1p/19q-net.

A threefold cross-validation procedure identical to the 1p/19q study was implemented with each group consisting of ~82 cases. The procedures of patch extraction, data augmentation, DVF, and statistical analysis were also the same.

5.2.3.2 Results

The MGMT-net achieved a mean cross-validation accuracy of 94.7% using majority voting with a sensitivity of 0.96 ±0.004%, specificity of 0.91 ±0.02%, PPV of 0.95 ±0.09%, NPV of 0.92 ±0.01% and AUC of 0.93 ±0.03, (see Table 5.3). Fig. 5.10 demonstrates the ROC curves for the MGMT network. Fig. 5.9 depicts the voxel-wise segmentation of the MGMT network.

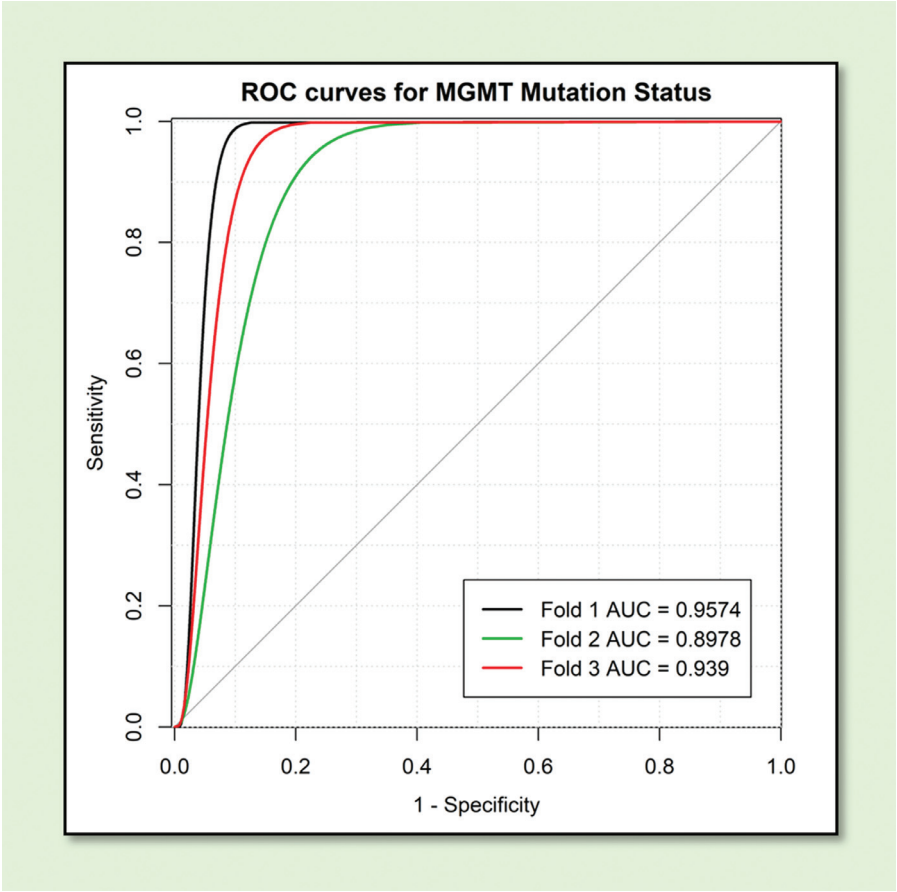

FIGURE 5.10

ROC curves for MGMT-net cross-validation.

5.3 Discussion

Using T2w MR images to noninvasively determine the genetic underpinnings of gliomas allows our "computing machine" to outperform humans. Diagnostic, therapeutic, and prognostic considerations of gliomas are now predominantly centered around the presence or absence of three specific mutations: IDH mutations, 1p/19q co-deletions, and methylation of the MGMT promoter.

In this chapter we showed that our 3-D Dense-Unet CNNs have mean cross-validation accuracies of 97.14 $\pm$0.04%, 93.4 $\pm$0.80%, 94.73 $\pm$0.66% for determining IDH mutation, 1p/19q co-deletion, and methylation of the MGMT promoter, respectively. Additionally, our 3-D Dense-Unet CNNs have a mean sensitivity and

specificity of 0.97 ±0.03% and 0.98 ±0.01%, 0.90 ±0.003% and 0.95 ±0.01%, 0.96 ±0.04% and 0.91 ±0.2% for determining IDH mutation status, 1p/19q co-deletions, and methylation of the MGMT promoter respectively. Finally, we demonstrate that our networks are able to achieve mean AUCs of 0.98 ±0.01, 0.95 ±0.01, and 0.93 ±0.01, for these three markers, respectively. Altogether, our networks achieved accuracies approaching current histopathological methodologies in determining glioma molecular status. There are many transformative implications of this work.

The most important implication is the ability to obviate the need for an invasive procedure to determine definitive treatment. After the WHO reclassification of gliomas, performing a biopsy to obtain tissue for molecular characterization of gliomas is the gold standard. The first step is determining if the tissue sample is adequate. This refers not only to the volume of tissue acquired but also whether all populations of cells are represented in the tissue sample. For example, it is not uncommon for gliomas to degenerate into more malignant cell lines. If these cell lines are not represented in the tissue sample, sampling bias results in inaccurate molecular characterization of the glioma. This becomes particularly important when molecular status determinations are made based on arbitrary "cut-off" values or percentages of cells that demonstrate a particular mutation [46–48]. What makes this determination even more challenging is that there are no standardized guidelines for "cut-off" values [46–48]. Therefore, it is not uncommon for sites to classify a glioma differently. Although advanced imaging techniques such as MR perfusion offer insight into the relative distribution of glioma metabolic activity, it is unclear whether more metabolically active regions always correlate to more aggressive cell lines [49,50]. Our noninvasive, image-based, deep learning networks overcome this fundamental bias by interrogating the entire tumor volume and then using majority voting across voxels to make a final determination.

Once the tissue sample has been deemed adequate in volume, the tissue will undergo several steps to process it and prepare it for analysis. Improper processing can lead to delays or the need for more tissue. Additionally, several next-generation molecular imaging techniques require tissue to be sent out for analysis. These factors add time and cost. Our networks are able to make molecular determinations based on T2w MR images alone in minutes. This substantially reduces the time to diagnosis and facilitates definitive treatment.

When compared to other deep-learning or machine learning algorithms, the use of T2w MR images alone is also a significant advantage. T2w images are routinely and quickly acquired as part of a conventional brain MRI protocol. High-quality T2w images can often be acquired even with patient motion degradation. There are several other factors that improve the performance of our networks when compared to other image-based algorithms. The architecture of our 3-D networks utilizes dense connections that offer several advantages: (1) they reduce overfitting, (2) they incorporate information from all previous layers into the successive layers, and (3) they are easier to train. The 3-D nature of our networks naturally represent interslice information more accurately than alternative 2-D networks. Our networks also use a DVF step that eliminates spurious voxels not connected to the tumor. Finally, because

the networks are voxel-wise classifiers they provide a classification for each voxel in the imaging volume. The other advantage of deep-learning networks is that as new molecular markers are discovered, this methodology can be applied to create new networks for those determinations.

Almost 75 years ago, speaking at a talk at the London Mathematical Society, Alan Turing described a trainable machine that will eventually outperform its initial instruction. He describes it by saying, "It would be like a pupil who had learnt much from his teacher but had added much more by his own work. When this happens, I feel that one is obliged to regard the machine as showing intelligence." Today, deep learning is revolutionizing how we noninvasively determine the genetic underpinnings of gliomas. It is likely that the next step in evolution will be a more efficient network that will automatically determine definitive treatment and accurately predict survival. To bridge the gap between the present and the future, several limitations will need to be overcome.

As our approach employ voxel-wise classification, the networks perform a *simultaneous single-label tumor segmentation on the entire image volume*. The networks demonstrated excellent tumor segmentation with Dice scores of 89% and 84% for TS-net and the T2w-net, respectively. Similarly, as the networks perform voxel-wise classification, certain tumor portions are classified as IDH wild-type, and others as IDH mutant. This is an interesting topic as heterogeneous variations in tumor biology of gliomas can occur locally and over time from cell-wise variation in genetics and/or gene expression [18,51]. Monoallelic gene expression can explain the heterogeneity of IDH and reports of false negativity in some gliomas [49,50]. There have been some reports of glioma heterogeneity in IDH mutation status. However, in our application it is more likely that the imaging features of the mutation status are heterogeneous. This results in a mixed mutated-wild type tumor segmentation output. Regardless, our voxel-wise approach provides classification accuracies that well outperform other methods.

Typically, deep-learning studies require a substantial amount of data to achieve good performance. Although several data augmentation strategies exist, larger training and validation datasets are needed to improve network performance. The number of subjects with ground truth molecular marker status and MRI from the TCIA database is rather small compared to those used for typical deep-learning networks. The TCIA consists of MR images from multiple institutions representing real-world clinical experience. Additionally, the imaging vendor platforms and acquisition parameters are diverse across the imaging centers contributing data to TCIA, which enhances the generalizability of algorithms trained using the dataset.

Although the results are very encouraging for clinical translation, performance of these algorithms needs further evaluation on larger clinical datasets. Before applying this approach in a clinical environment, it will be essential to train and validate the algorithms using additional independent datasets. It will also be important to evaluate the performance in the setting of patient motion and other imaging artifacts. This can also provide the opportunity to implement deep learning–based approaches to mitigate these artifacts.

5.4 Conclusion

Using convolutional neural networks, we developed algorithms for automatically segmenting and predicting IDH mutation, 1p/19q co-deletion, and MGMT methylation status of gliomas. These molecular markers now play a critical role in prognostic and therapeutic considerations as highlighted by the recent WHO reclassification. Our network performance rivals that of the gold- standard, invasive tissue–based approaches. In addition to being entirely non-invasive, our approach requires only routinely obtained T2w MRI images. This greatly enhances the clinical applicability by increasing robustness to real-world challenges such as patient motion, lengthy exam times, and the use of intravenous contrast. Further validation and fine-tuning is needed to facilitate clinical implementation. MR image–based deep-learning approaches to the molecular profiling of gliomas are transformative, potentially decreasing the need for an invasive neurosurgical procedure, decreasing the time and cost to diagnosis, and facilitating patient care by accelerating the time to definitive treatment.

Acknowledgments

Support for this research was provided by NCI U01CA207091 (AJM, JAM) and R01CA260705 (JAM).

References

[1] A.M. Turing, I.—Computing machinery and intelligence., Mind, LIX (236) (1950) 433–460.

[2] L. Pei, et al., Improved brain tumor segmentation by utilizing tumor growth model in longitudinal brain MRI, Proc. SPIE Int. Soc. Opt. Eng. (2017) 10134.

[3] K. Kamnitsas, E.Ferrante W.B., S. McDonagh, M. Sinclair, N. Pawlowski, M. Rajchl, M. Lee, B. Kainz, D. Rueckert, B. Glocker, Ensembles of multiple models and architectures for robust brain tumour segmentation, Lecture Notes in Computer Science 2017 (2018) 450–462.

[4] G.L. Wang, Wenqi & Ourselin, Sebastien & Vercauteren, Tom., Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks, In International MICCAI Brainlesion Workshop, Springer, Cham (2017) pp. 178–190.

[5] D.W. Parsons, et al., An integrated genomic analysis of human glioblastoma multiforme, Science 321 (5897) (2008) 1807–1812.

[6] D.N. Louis, et al., The 2016 World Health Organization classification of tumors of the central nervous system: a summary, Acta Neuropathol. 131 (6) (2016) 803–820.

[7] H. Yan, et al., IDH1 and IDH2 mutations in gliomas, New Engl. J. Med. 360 (8) (2009) 765–773.

[8] W.B. Pope, et al., Non-invasive detection of 2-hydroxyglutarate and other metabolites in IDH1 mutant glioma patients using magnetic resonance spectroscopy, J. Neuro-oncol. 107 (1) (2012) 197–205.

[9] C. Choi, et al., 2-hydroxyglutarate detection by magnetic resonance spectroscopy in IDH-mutated patients with gliomas, Nat. Med. 18 (4) (2012) 624.
[10] M.I. de la Fuente, et al., Integration of 2-hydroxyglutarate-proton magnetic resonance spectroscopy into clinical practice for disease monitoring in isocitrate dehydrogenase-mutant glioma, Neuro-Oncology 18 (2) (2015) 283–290.
[11] A. Tietze, et al., Noninvasive assessment of isocitrate dehydrogenase mutation status in cerebral gliomas by magnetic resonance spectroscopy in a clinical setting, J. Neurosurg. 128 (2) (2017) 391–398.
[12] C.H. Suh, et al., False-positive measurement at 2-hydroxyglutarate MR spectroscopy in isocitrate dehydrogenase wild-type glioblastoma: a multifactorial analysis, Radiology 291 (3) (2019) 752–762.
[13] K. Clark, et al., The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository, J. Digit. Imaging 26 (6) (2013) 1045–1057.
[14] M. Ceccarelli, et al., Molecular profiling reveals biologically discrete subsets and pathways of progression in diffuse glioma, Cell 164 (3) (2016) 550–563.
[15] R.L. Delfanti, et al., Imaging correlates for the 2016 update on WHO classification of grade II/III gliomas: implications for IDH, 1p/19q and ATRX status, J. Neurooncol. 135 (3) (2017) 601–609.
[16] K. Chang, et al., Residual convolutional neural network for the determination of IDH status in low- and high-grade gliomas from MR imaging, Clin. Cancer Res. 24 (5) (2018) 1073–1081.
[17] X. Zhang, et al., Radiomics strategy for molecular subtype stratification of lower-grade glioma: detecting IDH and TP53 mutations based on multimodal MRI, J. Magn. Reson. Imaging 48 (4) (2018) 916–926.
[18] P. Chang, et al., Deep-learning convolutional neural networks accurately classify genetic mutations in gliomas, AJNR Am. J. Neuroradiol. 39 (7) (2018) 1201–1207.
[19] S. Liang, et al., Multimodal 3D DenseNet for IDH genotype prediction in gliomas, Genes (Basel) 9 (8) (2018) 382.
[20] K.S. Choi, S.H. Choi, B. Jeong, Prediction of IDH genotype in gliomas with dynamic susceptibility contrast perfusion MR imaging using an explainable recurrent neural network, Neuro-Oncology (2019).
[21] H. Zhou, et al., Machine learning reveals multimodal MRI patterns predictive of isocitrate dehydrogenase and 1p/19q status in diffuse low- and high-grade gliomas, J. Neurooncol, 142 (2) (2019) 299–307.
[22] Z. Li, et al., Deep learning based radiomics (DLR) and its usage in noninvasive IDH1 prediction for low grade glioma, Sci. Rep. 7 (1) (2017) 5467.
[23] J. Yu, et al., Noninvasive IDH1 mutation estimation based on a quantitative radiomics approach for grade II glioma, Eur. Radiol. 27 (8) (2017) 3509–3522.
[24] Y. Matsui, et al., Prediction of lower-grade glioma molecular subtypes using deep learning, J. Neurooncol. 146 (2) (2020) 321–327.
[25] Z. Akkus, et al., Predicting deletion of chromosomal arms 1p/19q in low-grade gliomas from MR images using machine intelligence, J. Digit. Imaging 30 (4) (2017) 469–476.
[26] S.R. van der Voort, et al., Predicting the 1p/19q codeletion status of presumed low-grade glioma with an externally validated machine learning algorithm, Clin. Cancer Res. 25 (24) (2019) 7455–7462.
[27] B.M. Ellingson, et al., Consensus recommendations for a standardized brain tumor imaging protocol in clinical trials, Neuro-Oncology 17 (9) (2015) 1188–1198.

[28] E. Nyberg, et al., Comparison of brain MR images at 1.5T using BLADE and rectilinear techniques for patients who move during data acquisition, AJNR Am. J. Neuroradiol. 33 (1) (2012) 77–82.
[29] C.G. Bangalore Yogananda, et al., A novel fully automated MRI-based deep-learning method for classification of IDH mutation status in brain gliomas, Neuro-Oncology 22 (3) (2020) 402–411.
[30] C.G.B. Yogananda, et al., A novel fully automated MRI-based deep-learning method for classification of 1p/19q co-deletion status in brain gliomas, Neuro-Oncol. Adv. 2 (Supplement_4) (2021) iv42–iv48.
[31] C.G.B. Yogananda, et al., MRI-based deep-learning method for determining glioma MGMT promoter methylation status, Am. J. Neuroradiol. 42 (5) (2021) 845–852. https://doi.org/10.3174/ajnr.A7029.
[32] S. Bakas, et al., Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features, Sci. Data 4 (2017) 170117.
[33] T. Rohlfing, et al., The SRI24 multi-channel brain atlas: construction and applications, Proc. SPIE Int. Soc. Opt. Eng. 6914 (2008) 691409.
[34] N.J. Tustison, et al., Large-scale evaluation of ANTs and FreeSurfer cortical thickness measurements, Neuroimage 99 (2014) 166–179.
[35] B.B. Avants, et al., A reproducible evaluation of ANTs similarity metric performance in brain image registration, Neuroimage 54 (3) (2011) 2033–2044.
[36] S. Jegou, Michal. Drozdzal, David. Vazquez, Adriana. Romero, Yoshua Bengio, The One Hundred Layers Tiramisu: Fully Convolutional DenseNets for Semantic Segmentation, In Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 2017, pp. 11–19.
[37] V. Wegmayr, J.A.S. Buhmann, N.K. Mori, Classification of brain MRI with big data and deep 3D convolutional neural networks. Proceedings, Medical Imaging 2018: Computer-Aided Diagnosis (2018) 1057501.
[38] J.Y. Xinyang Feng, Zachary C Lipton, Scott A Small, Frank A Provenzano, Deep learning on MRI affirms the prominence of the hippocampal formation in Alzheimer's disease classification, bioRxiv 2018 (2018) 456277.
[39] Pedregosa, F., et al., Scikit-learn: Machine learning in Python. 2011. 12(Oct), 2825–2830.
[40] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Dais, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D.G. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, X. Zheng, G. Brain, Tensorflow: a system for large-scale machine learning, in: Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI '16), 2016, pp. 265–284.
[41] D.P. Kingma, J.L. Ba, Adam: A method for stochastic optimization, {ICLR} 2015, San Diego, CA, USA, May 7–9, 2015, Conference Track Proceedings. In Conference Track Proceedings (2015).
[42] M. Zhang, et al., Identification of new biomarkers associated with IDH mutation and prognosis in astrocytic tumors using nanostring N counter analysis system, Appl. Immunohistochem. Mol. Morphol. 26 (2) (2018) 101–107.
[43] J. Jr.Polivka, et al., Co-deletion of 1p/19q as prognostic and predictive biomarker for patients in West Bohemia with anaplastic oligodendroglioma, Anticancer Res. 36 (1) (2016) 471–476.

[44] Y.-X. Li, et al., Not all 1p/19q non-codeleted oligodendroglial tumors are astrocytic, Oncotarget 7 (40) (2016) 64615–64630.
[45] M.E. Hegi, et al., Correlation of O6-methylguanine methyltransferase (MGMT) promoter methylation with clinical outcomes in glioblastoma and clinical strategies to modulate MGMT activity, J. Clin. Oncol. 26 (25) (2008) 4189–4199.
[46] D. Lee, et al., IDH1 mutations in oligodendroglial tumors: comparative analysis of direct sequencing, pyrosequencing, immunohistochemistry, nested PCR and PNA-mediated clamping PCR, Brain Pathol. 23 (3) (2013) 285–293.
[47] R. Senetta, et al., A "weighted" fluorescence in situ hybridization strengthens the favorable prognostic value of 1p/19q codeletion in pure and mixed oligodendroglial tumors, J. Neuropathol. Exp. Neurol. 72 (5) (2013) 432–441.
[48] A. Woehrer, J.A. Hainfellner, Molecular diagnostics: techniques and recommendations for 1p/19q assessment, CNS Oncol. 4 (5) (2015) 295–306.
[49] M. Preusser, et al., Value and limitations of immunohistochemistry and gene sequencing for detection of the IDH1-R132H mutation in diffuse glioma biopsy specimens, J. Neuropathol. Exp. Neurol. 70 (8) (2011) 715–723.
[50] C. Horbinski, What do we know about IDH1/2 mutations so far, and how do we use it? Acta neuropathologica 125 (5) (2013) 621–636.
[51] S. Pusch, et al., Glioma IDH1 mutation patterns off the beaten track, Neuropathol. Appl. Neurobiol. 37 (4) (2011) 428–430.

Non-Print Items

Abstract

Gliomas demonstrate diverse imaging features, variable response to therapy, and differences in prognosis. This is largely a function of genetic heterogeneity. Several key mutations serve as therapeutic and prognostic markers such as isocitrate dehydrogenase (IDH) mutation status, O6-methyl guanine-DNA methyltransferase (MGMT) promoter status, and 1p/19q co-deletion status. Currently, the gold standard for molecular marker determination requires tissue from either an invasive brain biopsy or surgical resection. Here we describe our work in developing highly accurate simultaneous deep learning segmentation and classification approaches for noninvasive profiling of molecular markers using T2-weighted magnetic resonance images only.

Keywords

Isocitrate dehydrogenase; 1p/19q; Methyl guanine-DNA methyltransferase; Deep learning; Convolutional Neural Networks (CNN); Magnetic resonance imaging; Glioma; Dense-U-net