bioRxiv preprint doi: https://doi.org/10.1101/757385; this version posted September 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

A NOVEL FULLY AUTOMATED MRI-BASED DEEP LEARNING METHOD FOR CLASSIFICATION OF IDH MUTATION STATUS IN BRAIN GLIOMAS

Chandan Ganesh Bangalore Yogananda, Bhavya R. Shah, Maryam Vejdani-Jahromi, Sahil S. Nalawade, Gowtham K. Murugesan, Frank F. Yu, Marco C. Pinho, Benjamin C. Wagner, Bruce Mickey, Toral R. Patel, Baowei Fei, Ananth J. Madhuranthakam, Joseph A. Maldjian

Department of Radiology, University of Texas Southwestern Medical Center, Dallas, Texas (C.G.B.Y., B.R.S., M.J., S.S.N., G.K.M., F.F.Y., M.C.P., B.C.W., A.J.M., J.A.M); Department of Neurological Surgery, University of Texas Southwestern Medical Center, Dallas, Texas (B.M., T.R.P.); Department of Bioengineering, University of Texas at Dallas, Richardson, Texas (B.F.)

WITHDRAWN
see manuscript DOI for details

Running Title: IDH classification with MRI

Corresponding Author: Joseph A. Maldjian, MD, Department of Radiology, UT Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, Texas 75390-9178 (Joseph.Maldjian@utsouthwestern.edu)
Office: 214-648-6751, Fax: 214-648-3904

Funding: This work was supported by NIH/NCI U01CA207091 (AJM, JAM).

Conflict of Interest: None.

Authorship: Experimental Design (C.G.B.Y., M.V.J., S.S.N., G.K.M., B.C.W., A.J.M., J.A.M.)**;** Implementation (C.G.B.Y., M.V.J., S.S.N., G.K.M., F.F.Y., M.C.P., B.C.W., A.J.M., J.A.M.)**;** Analysis and interpretation of data (C.G.B.Y., B.R.S, M.V.J., S.S.N., G.K.M., F.F.Y., M.C.P.,B.M., T.R.P, B.F., A.J.M., J.A.M.)**;** Writing of the manuscript: (C.G.B.Y., B.R.S, M.V.J., S.S.N., G.K.M., F.F.Y., M.C.P., B.C.W, B.M., T.R.P, B.F., A.J.M., J.A.M.)

Word count*: 5101*

bioRxiv preprint doi: https://doi.org/10.1101/757385; this version posted September 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

WITHDRAWN
see manuscript DOI for details

ABSTRACT

Background: Isocitrate dehydrogenase (IDH) mutation status has emerged as an important prognostic marker in gliomas. Currently, reliable IDH mutation determination requires invasive surgical procedures. The purpose of this study was to develop a highly-accurate, MRI-based, voxel-wise deep-learning IDH-classification network using T2-weighted (T2w) MR images and compare its performance to a multi-contrast network.

Methods: Multi-parametric brain MRI data and corresponding genomic information were obtained for 214 subjects (94 IDH-mutated, 120 IDH wild-type) from The Cancer Imaging Archive (TCIA) and The Cancer Genome Atlas (TCGA). Two separate networks were developed including a T2w image only network (T2-net) and a multi-contrast (T2w, FLAIR, and T1 post-contrast), network (TS-net) to perform IDH classification and simultaneous single label tumor segmentation. The networks were trained using 3D-Dense-UNets. A three-fold cross-validation was performed to generalize the networks' performance. ROC analysis was also performed. Dice-scores were computed to determine tumor segmentation accuracy.

Results: T2-net demonstrated a mean cross-validation accuracy of 97.14% +/- 0.04 in predicting IDH mutation status, with a sensitivity of 0.97 +/- 0.03, specificity of 0.98 +/- 0.01, and an AUC of 0.98 +/- 0.01. TS-net achieved a mean cross-validation accuracy of 97.12% +/- 0.09, with a sensitivity of 0.98 +/- 0.02, specificity of 0.97 +/- 0.001, and an AUC of 0.99 +/- 0.01. The mean whole tumor segmentation Dice-scores were 0.85 +/- 0.009 for T2-net and 0.89 +/- 0.006 for TS-net.

Conclusion: We demonstrate high IDH classification accuracy using only T2-weighted MRI. This represents an important milestone towards clinical translation.

Keywords *– IDH (Isocitrate Dehydrogenase), deep learning, MRI (Magnetic Resonance Imaging), CNN (Convolutional Neural Networks), 3D-Dense-UNet, glioma, tumor segmentation, machine learning, brain tumor.*

bioRxiv preprint doi: https://doi.org/10.1101/757385; this version posted September 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Keypoints – 1. IDH status is an important prognostic marker for gliomas. 2. We developed a non-invasive, MRI based, highly accurate deep-learning method for the determination of IDH status 3. The deep-learning networks utilizes only T2 weighted MR images to predict IDH status thereby facilitating clinical translation.

WITHDRAWN
see manuscript DOI for details

IMPORTANCE OF THE STUDY

One of the most important recent discoveries in brain glioma biology has been the identification of the isocitrate dehydrogenase (IDH) mutation status as a marker for therapy and prognosis. The mutated form of the gene confers a better prognosis and treatment response than gliomas with the non-mutated or wild-type form. Currently, the only reliable way to determine IDH mutation status is to obtain glioma tissue either via an invasive brain biopsy or following open surgical resection. The ability to non-invasively determine IDH mutation status has significant implications in determining therapy and predicting prognosis. We developed a highly accurate, deep learning network that utilizes only T2-weighted MR images and outperforms previously published methods. The high IDH classification accuracy of our T2w image only network (T2-net) marks an important milestone towards clinical translation. Imminent clinical translation is feasible because T2-weighted MR imaging is widely available and routinely performed in the assessment of gliomas.

bioRxiv preprint doi: https://doi.org/10.1101/757385; this version posted September 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

INTRODUCTION

Isocitrate dehydrogenase (IDH) mutation status has emerged as one of the most important markers for glioma diagnosis and therapy. Gliomas with this mutant enzyme have a better prognosis than tumors of the same grade with wild-type IDH. This observation led the World Health Organization (WHO) to revise their classification of gliomas in 2016 [1]. IDH mutated tumors also have different management and therapeutic approaches than tumors with wild-type mutation status. At the present time, the only way to definitively identify an IDH mutated glioma is to perform immunohistochemistry or gene sequencing on a tissue specimen, acquired through biopsy or surgical resection. Because the differences between IDH mutated and IDH wild-type gliomas may have critical treatment implications, there is great interest in attempting to distinguish between these two tumor types prior to surgery. This becomes even more important for brain tumors that are inaccessible for biopsy or resection due to a high risk of severe post-operative complications and impairment.

WITHDRAWN
see manuscript DOI for details

MR spectroscopy can potentially be used to determine IDH mutation status. The mutant IDH enzyme catalyzes the production of the oncometabolite 2-hydroxyglutarate (2-HG) [2]. MR spectroscopic methods have been developed for identification of 2-HG [3-6] noninvasively in brain tumors. While these methods appear to work well in a research setting, in the busy clinical environment, the spectroscopic imaging data are frequently uninterpretable due to artifact, patient motion, poor shimming, small voxel sizes, non-ideal tumor location, or presence of hemorrhage or calcification affecting measurements. Even in the setting of good quality spectra, reliable clinical implementation using 2-HG spectroscopy is further compounded by the recently described high false positive rate of over 20% using this technique in the best hands [7].

Early determination of IDH mutation status directly impacts treatment decisions. Tumors that appear to be low-grade gliomas, but are IDH wild-type, are typically treated with early intervention rather

bioRxiv preprint doi: https://doi.org/10.1101/757385; this version posted September 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

than observation. Specific chemotherapeutic interventions are more effective in IDH-mutated gliomas (e.g., temozolomide) [8-12]. Additionally, surgical resection of non-enhancing tumor volume (beyond gross total resection of enhancing tumor components) in Grade III-IV IDH-mutated tumors has been demonstrated to have a survival benefit [13]. However, the determination of IDH mutation status continues to be performed using direct tissue sampling. Development of a robust non-invasive approach would be transformational in the care of these patients.

Advances in deep-learning methods are outperforming traditional machine-learning methods in predicting the genetic and molecular biology of tumors based on MRI. For example, Zhang et al. used a radiomics approach integrating a support vector machine (SVM)-based model and multimodal MRI features with an accuracy of 80% for IDH detection [14]. Recent studies by Chang et al., have used deep learning techniques to noninvasively determine IDH mutation status based on MRI, with accuracies of 94% using the TCIA database [15]. Unfortunately, none of these methods are clinically viable, requiring either manual pre-segmentation of the tumor, extensive pre-processing, or multi-contrast acquisitions that are frequently affected by patient motion due to the long scan times. Additionally, these existing methodologies use a 2D (slice-wise) classification approach. A known limitation in designing and developing a slice-wise classification model is the *data-leakage problem* [16,17]. 2D slice-wise models working with cross-sectional imaging data are particularly prone to data-leakage because they perform slice randomization across all subjects to generate the training, validation, and testing slices. As a result, adjacent slices from the same subject may be found in the training, validation, or testing data subsets. Because adjacent slices often share considerable information, this methodology may artificially boost accuracies by introducing bias in the testing phase. The previously reported studies do not appear to adhere to this caveat, potentially resulting in artificially boosted accuracies.

The purpose of this study is to develop a highly accurate fully automated deep learning IDH-classification 3D network using T2-weighted images only and compare its performance to a multi-contrast

WITHDRAWN
see manuscript DOI for details

bioRxiv preprint doi: https://doi.org/10.1101/757385; this version posted September 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

3D network. The use of T2 images only provides strong clinical translation capability. T2 images are routinely acquired as part of any MRI brain tumor evaluation. These images are robust to motion and can be obtained within 2 minutes. On modern MRI scanners available in most clinical settings, high quality T2w images can be obtained even in the presence of active patient motion using commonly available motion resistant acquisition techniques [18].

WITHDRAWN
see manuscript DOI for details

MATERIAL & METHODS

Data and Pre-processing

Multi-parametric brain MRI data of glioma patients were obtained from the Cancer Imaging Archive (TCIA) [19] database. Genomic information was provided from TCGA (the cancer genome atlas) database [20]. Only pre-operative studies were used. Studies were screened for the availability of IDH status and T2w, T2w-FLAIR, and contrast enhanced T1-weighted (T1c) image series. The final dataset included 214 subjects (94 IDH-mutated, 120 IDH wild-type). Tumor masks for 87 data sets were available through previous expert segmentation [21] and were used as the ground truth for the tumor segmentation in the training set. Ground truth whole tumor masks for IDH mutated type were labelled with 1s and the ground truth tumor masks for IDH wild-type were labelled with 2s (Figure 1). Tumor masks for the remaining 127 subjects were manually drawn and validated by in-house neuro-radiologists. Data preprocessing was minimal, including (a) N4BiasCorrection to remove RF inhomogeneity, (b) co-registration of the multi-contrast data to the T1c (for TS-net only), and (c) intensity normalization to zero-mean and unit variance[22]. The pre-processing was developed using the Advanced Normalization Tools (ANTS) software routines[23].

Network Details

Two separate networks were developed. These included a T2w image only network (T2-net) trained only on the T2w images (Figure 2A), and a 3-sequence network (TS-net) trained on multi-contrast MR data including T2w images, T2w-FLAIR, and T1c. A 3D 32×32×32 patch-based training and testing

bioRxiv preprint doi: https://doi.org/10.1101/757385; this version posted September 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

approach was implemented for both networks. Dense-UNets were designed and trained for a voxel-wise dual-class segmentation of the whole tumor with Classes 1 and 2 representing IDH-mutated and IDH-wild-type, respectively. The schematics for the network architecture is shown in Figure 2B. Each network consisted of seven dense blocks: three transition down blocks, three transition up blocks, an initial convolution layer, and a final convolution layer followed by an activation layer at the end. Each dense block was made up of five layers. Each layer was connected to every other layer in that particular dense block. This dense connection was implemented by concatenating the feature maps from one layer with feature maps from every other layer of that dense block. The input to a dense block was also concatenated with the output of that dense block. Every dense block on the encoder part of the network was followed by a transition down block, while every dense block on the decoder part of the network was preceded by a transition up block. The bottleneck block was used to keep the convolution layers to a smaller number in order to avoid having large convolution layers. With these connecting patterns, all feature maps were reused such that every layer in the architecture received a direct supervision signal [24]. A detailed description of the network is given in the supplemental material section.

WITHDRAWN
see manuscript DOI for details

Training Details

70 subjects were used for training (40 IDH mutated, 30 IDH wild-type), 17 subjects for in-training validation (8 IDH mutated, 9 IDH wild-type), and 127 held-out subjects were used for testing (46 IDH mutated, 81 IDH wild-type). To avoid the data leakage problem, no patch from the same subject was mixed with the training, validation or testing datasets. [16,17]. The data augmentation steps included horizontal flipping, vertical flipping, random and translational rotation. Data augmentation provided a total 0f 150,000 patches for training and 37,000 patches for validation. Networks were implemented using Keras and Tensorflow [25] with an Adaptive Moment Estimation optimizer (Adam) [26]. The initial learning rate was set to 10^{-5} with a batch size of 4 and maximal iterations of 100. Training was implemented on Tesla P100, P40 and K80 NVIDIA-GPUs.

bioRxiv preprint doi: https://doi.org/10.1101/757385; this version posted September 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Testing Details

Each network yields two segmentation volumes. Volume 1 provides the voxel-wise prediction of IDH mutated tumor and Volume 2 identifies the predicted IDH wild-type tumor voxels. A straightforward dual-volume fusion (DVF) approach was developed to combine the 2 segmentation volumes. Both the volumes were combined and the largest connected component was obtained using a 3D connected component algorithm in MATLAB(R). Majority voting over the voxel-wise classes of IDH-mutated or IDH-wild-type provided a single IDH classification for each subject. The IDH classification process is fully automated, and a tumor segmentation map is a natural output of the voxel-wise classification approach. The combined volumes provided a single tumor segmentation map. Testing was performed on either K80 or P40 NVIDIA-GPUs.

WITHDRAWN
see manuscript DOI for details

Cross-validation

To evaluate the reliability of the networks, a 3-fold cross-validation was performed on the 214 subjects by randomly shuffling the dataset and distributing it into training, validation and testing (approximately 70 subjects in training, 70 subjects in validation, and 70 subject in testing). Majority voting over the voxel-wise classes of IDH-mutated or IDH-wild-type provided a single IDH classification for each subject.

Statistical Analysis

Statistical analysis was performed in MATLAB(R) and R for T2-net and TS-net separately. The accuracy of the two networks was evaluated with majority voting (*i.e.* voxel-wise cutoff of 50%). This threshold was then used to calculate the accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of the model. A logistic regression model was applied to the subject IDH classification and the Receiver Operating Characteristic (ROC) curve was derived. To evaluate the

bioRxiv preprint doi: https://doi.org/10.1101/757385; this version posted September 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

performance of the networks for tumor segmentation, the Dice-score was used. The Dice-score determines the amount of spatial overlap between the ground truth segmentation and the network segmentation.

RESULTS

T2-net

The T2-net achieved a mean testing accuracy of 96.1% by majority voting with sensitivity of 97.8%, specificity of 95.1%, PPV of 91.8%, NPV of 98.8% and an area under the curve (AUC) of 0.973 (goodness of fit p-value $8.9*10^{-30}$). T2-net segmented the whole tumor with a Dice-score of 0.84. The confusion matrix is provided in Table 1. T2-net misclassified 1 IDH-mutated subject (out of 46) and 4 IDH wild-type subjects out of 81.

WITHDRAWN
see manuscript DOI for details

Multi-contrast TS-net

The multi-contrast TS-net achieved a mean testing accuracy of 96.8% by majority voting with sensitivity of 100%, specificity of 95.1%, PPV of 92.0%, NPV of 100% with AUC of 0.975 (goodness of fit p-value: $1.2*10^{-28}$). TS-net segmented the whole tumor with a Dice-score of 0.89. The confusion matrix for TS-net is provided in Table 1. TS-net classified all IDH-mutated subjects correctly. It misclassified 4 IDH wild-type subjects out of 81. These 4 subjects were not the same 4 subjects that T2-net misclassified.

ROC analysis

The network output classifies voxels in the tumor as IDH mutated or IDH wild-type. The percent of IDH-mutated voxels was computed for the network output for each subject in the test set by dividing the predicted IDH mutated voxels by the total number of predicted voxels in each tumor. The percent mutated voxels can be viewed as a network output prediction likelihood of the tumor being IDH-mutated. Note that, the majority voting (the 50% threshold) was used to determine IDH prediction. For the ROC analysis, the percent of IDH mutated voxels was sorted and used as separate thresholds (cut-points) to determine

bioRxiv preprint doi: https://doi.org/10.1101/757385; this version posted September 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

IDH mutation status for the subjects across the test set for each new cut-point. The resulting predicted IDH class membership was compared to the ground truth values to determine sensitivity (true positive rate) and 1- specificity (false positive rate) at each threshold. The resulting values were plotted using Matlab to obtain an ROC curve (true positive rate against false positive rate). Matlab routines were used to fit the curves and determine the area under the curve (AUC).

Voxel-wise classification

Since these networks are voxel-wise classifiers, they perform a simultaneous tumor segmentation. Figures 3A and 3B show examples of the voxel-wise classification for an IDH wild-type, and IDH mutated case using T2-net. The DVF procedure was effective in removing false positives to increase accuracy.

WITHDRAWN
see manuscript DOI for details

Cross validation results

The cross-validation subject-wise accuracy for T2-net and TS-net were nearly identical. T2-net achieved a mean cross-validation testing accuracy of 97.14% across the 3 folds (97.18%, 97.14%, and 97.10%, standard dev=0.04). Mean cross-validation sensitivity, specificity, PPV, NPV and AUC for T2-net was 0.97 ±0.03, 0.98 ±0.01, 0.98 ±0.01, 0.97 ±0.01 and 0.98 ±0.01, respectively. The mean cross-validation Dice-score for tumor segmentation was 0.85 ± 0.009. (Table 2). The multi-contrast TS-net achieved a mean cross-validation testing accuracy of 97.12% across the 3 folds (97.22%, 97.10%, and 97.05%, standard dev=0.09). Mean cross-validation sensitivity, specificity, PPV, NPV and AUC for TS-net was 0.98 ±0.02, 0.97 ±0.001, 0.97 ±0.002, 0.97 ±0.001 and 0.99 ±0.01, respectively. The mean cross-validation Dice-score for tumor segmentation was 0.89 ± 0.006 (Table 2).

Training and segmentation times:

Each network took approximately 2 weeks to train. The trained networks took approximately three minutes to segment the whole tumor, implement DVF and predict the IDH mutation status for each subject.

bioRxiv preprint doi: https://doi.org/10.1101/757385; this version posted September 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

DISCUSSION:

We developed two deep-learning MRI networks for IDH-classification of gliomas. Both our T2-network and the multi-contrast network outperformed IDH classification algorithms previously reported in the literature [14,15,27,28]. When comparing the T2-network with the multi-contrast network, our results suggest that similar performance can be achieved by using T2-weighted images only. The ability to use only T2-weighted images makes clinical translation much more straightforward and less prone to failures from image acquisition artifacts. The preprocessing used to prepare the data is also minimal. The time required for T2-net to segment the whole tumor, implement DVF, and predict the IDH mutation status for one subject is approximately 3 minutes on a K80 or P40 NVIDA-GPU.

WITHDRAWN
see manuscript DOI for details

There are several factors that may explain the higher performance achieved by our networks when compared to previously published results. First and foremost is the use of 3D networks, compared to previously reported 2D networks. Additionally, the 3D network architecture is advantageous as the dense connections carry information from all the previous layers to the following layers [24]. These types of networks are easier to train and can reduce over-fitting [29]. The Dual Volume Fusion (DVF) post-processing step also helps in effectively eliminating false positives while increasing the segmentation accuracy. It essentially excludes extraneous voxels that are not connected to the tumor. The 3D networks interpolate between slices to maintain inter-slice information more accurately. The networks use minimal preprocessing without any requirement for extraction of pre-engineered features from the images or histopathological data [27].

The 3D networks used here are voxel-wise classifiers, providing a classification for each voxel in the image. This provides a simultaneous single-label tumor segmentation (e.g. the sum of voxels classified as IDH mutated and non-mutated provide the tumor label). The single label whole tumor segmentation performance for these networks provided excellent Dice-scores of 0.84 and 0.89 for T2-net and the multi-

bioRxiv preprint doi: https://doi.org/10.1101/757385; this version posted September 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

contrast TS-net, respectively. These tumor segmentation Dice-scores are similar to the top performers from BraTS2017 tumor segmentation challenge [29].

Both T2-net and TS-net achieved similar overall subject classification accuracies. For the IDH wild-type tumors, both networks incorrectly classified 4 subjects from the test set. These 4 subjects were not the same between the networks. In reviewing these cases, there were no discriminating imaging features. All these cases were enhancing tumors, with mixed T2 and FLAIR signal, and with significant edema.

WITHDRAWN
see manuscript DOI for details

Since these networks are voxel-wise classifiers, there are portions within each tumor that are classified as IDH mutated, and other areas as IDH-wild-type. Heterogeneous genetic expression can occur in gliomas over time and result in varied tumor biology [15,30]. Immunohistochemistry (IHC) evaluations use monoclonal antibodies to detect the most frequent IDH mutations (e.g. IDH1-R132H). Different cutoff values have been proposed to determine the IDH status of a tissue sample. While some advocate staining of more than 10% of tumor cells to confer IDH positivity, others suggest that one "strongly" staining tumor cell is sufficient[31]. Heterogeneity of staining with IHC has been reported where up to 46% of subjects showed partial uptake[32]. In 2011 Perusser et al. reported that IDH1-R132H expression may occur in only a fraction of tumor cells[33]. Heterogeneity of the sample can also affect the sensitivity of genetic testing [34]. IDH heterogeneity and reported false negativity in some gliomas have been explained by monoallelic gene expression, wherein only one allele of a gene is expressed even though both alleles are present. According to Horbinski, sequencing may not always be adequate to identify tumors that are functionally IDH1/2 mutated [33,35]. Although heterogeneity of IDH status has been reported in histochemical and genomic evaluations of gliomas, we do not make the claim that the deep learning networks are detecting heterogeneous IDH mutation status in these tumors. Rather, the morphologic expression of the IDH mutation status is likely heterogeneous and reflected in the mixed classification outputs of IDH-mutated

bioRxiv preprint doi: https://doi.org/10.1101/757385; this version posted September 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

and IDH-wild-type within a particular tumor. Regardless, the accuracies using this voxel-wise approach well outperform other methods.

The multi-contrast input required by previous approaches can be compromised due to patient motion from lengthier examination times, and the need for gadolinium contrast. High quality T2-weighted images are almost universally acquired during clinical brain tumor diagnostic evaluation. Clinically, T2w images are typically acquired within 2 minutes at the beginning of the exam and are relatively resistant to the effects of patient motion. On modern MRI scanners, high quality T2w images can even be obtained in the presence of patient motion [18]. As such, the ability to use only T2w images is a significant advantage when considering clinical translatability. This method was inspired by a similar approach used for the identification of the O6–methylguanine-DNA methyltransferase (MGMT) methylation status and prediction of 1p/19q chromosomal arm deletion (15). Furthermore, our preprocessing steps preserve native image information without the need for any region-of-interest or tumor pre-segmentation procedures. Previous deep learning algorithms for MRI-based IDH classification use explicit tumor presegmentation steps. These were accomplished either by manual delineation of the tumor, or by adding a separate deep learning tumor segmentation network. The use of these presegmentation steps adds unnecessary complexity to the classification process, and in the case of manual presegmentation, makes them unworkable as a robust automated clinical workflow. Our network uniquely performs a simultaneous tumor segmentation as a natural consequence of the voxel-wise segmentation process.

WITHDRAWN
see manuscript DOI for details

LIMITATIONS

Deep learning studies typically require a very large amount of data to achieve good performance. The number of subjects available from the TCIA dataset is relatively small compared to the sample-sizes typically required for deep learning. Despite this caveat, the data are representative of real-world clinical experience, with multi-parametric MR images from multiple institutions and represents one of the largest publically available brain tumor databases. Additionally, the acquisition parameters and imaging vendor

bioRxiv preprint doi: https://doi.org/10.1101/757385; this version posted September 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

platforms are diverse across the imaging centers contributing data to the TCIA dataset. Although our results show promise for expeditious clinical translation, our algorithms performance will need to be replicated in an independent dataset.

CONCLUSION

We developed two deep-learning MRI networks for IDH-classification of gliomas: i) a T2-network and ii) a multi-contrast network with high accuracy. Both networks outperformed the state-of-the-art algorithms. We also demonstrate similar performance when comparing the T2-network with the multi-contrast network. The high accuracy and use of only T2-weighted images will facilitate imminent clinical translation for this approach.

WITHDRAWN
see manuscript DOI for details

Funding

Support for this research was provided by NIH/NCI U01CA207091 (AJM, JAM).

Disclosures

No conflicts of interest

Acknowledgments

None

bioRxiv preprint doi: https://doi.org/10.1101/757385; this version posted September 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

REFERENCES

1. Louis DN, Perry A, Reifenberger G, et al. The 2016 World Health Organization classification of tumors of the central nervous system: a summary. *Acta neuropathologica.* 2016; 131(6):803-820.

2. Yan H, Parsons DW, Jin G, et al. IDH1 and IDH2 mutations in gliomas. *New England Journal of Medicine.* 2009; 360(8):765-773.

3. Pope WB, Prins RM, Thomas MA, et al. Non-invasive detection of 2-hydroxyglutarate and other metabolites in IDH1 mutant glioma patients using magnetic resonance spectroscopy. *Journal of neuro-oncology.* 2012; 107(1):197-205.

4. Choi C, Ganji SK, DeBerardinis RJ, et al. 2-hydroxyglutarate detection by magnetic resonance spectroscopy in IDH-mutated patients with gliomas. *Nature medicine.* 2012; 18(4):624.

5. de la Fuente MI, Young RJ, Rubel J, et al. Integration of 2-hydroxyglutarate-proton magnetic resonance spectroscopy into clinical practice for disease monitoring in isocitrate dehydrogenase-mutant glioma. *Neuro-oncology.* 2015; 18(2):283-290.

6. Tietze A, Choi C, Mickey B, et al. Noninvasive assessment of isocitrate dehydrogenase mutation status in cerebral gliomas by magnetic resonance spectroscopy in a clinical setting. *Journal of neurosurgery.* 2017; 128(2):391-398.

7. Suh CH, Kim HS, Paik W, et al. False-Positive Measurement at 2-Hydroxyglutarate MR Spectroscopy in Isocitrate Dehydrogenase Wild-Type Glioblastoma: A Multifactorial Analysis. *Radiology.* 2019; 291(3):752-762.

8. SongTao Q, Lei Y, Si G, et al. IDH mutations predict longer survival and response to temozolomide in secondary glioblastoma. *Cancer Sci.* 2012; 103(2):269-273.

9. Okita Y, Narita Y, Miyakita Y, et al. IDH1/2 mutation is a prognostic marker for survival and predicts response to chemotherapy for grade II gliomas concomitantly treated with radiation therapy. *Int J Oncol.* 2012; 41(4):1325-1336.

10. Mohrenz IV, Antonietti P, Pusch S, et al. Isocitrate dehydrogenase 1 mutant R132H sensitizes glioma cells to BCNU-induced oxidative stress and cell death. *Apoptosis.* 2013; 18(11):1416-1425.

11. Molenaar RJ, Botman D, Smits MA, et al. Radioprotection of IDH1-Mutated Cancer Cells by the IDH1-Mutant Inhibitor AGI-5198. *Cancer Res.* 2015; 75(22):4790-4802.

12. Sulkowski PL, Corso CD, Robinson ND, et al. 2-Hydroxyglutarate produced by neomorphic IDH mutations suppresses homologous recombination and induces PARP inhibitor sensitivity. *Sci Transl Med.* 2017; 9(375).

13. Beiko J, Suki D, Hess KR, et al. IDH1 mutant malignant astrocytomas are more amenable to surgical resection and have a survival benefit associated with maximal surgical resection. *Neuro Oncol.* 2014; 16(1):81-91.

14. Zhang M, Pan Y, Qi X, et al. Identification of New Biomarkers Associated With IDH Mutation and Prognosis in Astrocytic Tumors Using NanoString nCounter Analysis System. *Appl Immunohistochem Mol Morphol.* 2018; 26(2):101-107.

15. Chang P, Grinband J, Weinberg BD, et al. Deep-Learning Convolutional Neural Networks Accurately Classify Genetic Mutations in Gliomas. *AJNR Am J Neuroradiol.* 2018; 39(7):1201-1207.

16. Wegmayr V AS, Buhmann J, Nicholas Petrick; Kensaku Mori, Editor(s). Classification of brain MRI with big data and deep 3D convolutional neural networks. *Published in SPIE Proceedings, Medical Imaging 2018: Computer-Aided Diagnosis.* 2018; 1057501.

WITHDRAWN
See manuscript DOI for details

bioRxiv preprint doi: https://doi.org/10.1101/757385; this version posted September 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

17. Xinyang Feng JY, 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.

18. Nyberg E, Sandhu GS, Jesberger J, 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.* 2012; 33(1):77-82.

19. Clark K, Vendt B, Smith K, et al. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. *J Digit Imaging.* 2013; 26(6):1045-1057.

20. Ceccarelli M, Barthel FP, Malta TM, et al. Molecular Profiling Reveals Biologically Discrete Subsets and Pathways of Progression in Diffuse Glioma. *Cell.* 2016; 164(3):550-563.

21. Bakas S, Akbari H, Sotiras A, et al. Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. *Sci Data.* 2017; 4:170117.

22. Tustison NJ, Cook PA, Klein A, et al. Large-scale evaluation of ANTs and FreeSurfer cortical thickness measurements. *Neuroimage.* 2014; 99:166-179.

23. Avants BB, Tustison NJ, Song G, Cook PA, Klein A, Gee JC. A reproducible evaluation of ANTs similarity metric performance in brain image registration. *Neuroimage.* 2011; 54(3):2033-2044.

24. Jegou S, Drozdzal, Michal., Vazquez, David., Romero, Adriana., Bengio, Yoshua. The One Hundred Layers Tiramisu: Fully Convolutional DenseNets for Semantic Segmentation. 2017.

25. Abadi M, et al.,. Tensorflow: A system for large-scale machine learning. *OSDI.* 2016; pp. 265-284.

26. Kingma DP, Ba, J.L.,. Adam: A method for stochastic optimization. . *ICLR.* 2015.

27. Delfanti RL, Piccioni DE, Handwerker J, 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.* 2017; 135(3):601-609.

28. Chang K, Bai HX, Zhou H, et al. Residual Convolutional Neural Network for the Determination of IDH Status in Low- and High-Grade Gliomas from MR Imaging. *Clin Cancer Res.* 2018; 24(5):1073-1081.

29. Wang GL, Wenqi & Ourselin, Sebastien & Vercauteren, Tom. Automatic Brain Tumor Segmentation using Cascaded Anisotropic Convolutional Neural Networks. *MICCAI Brats Challenge 2017.* 2017.

30. Pusch S, Sahm F, Meyer J, Mittelbronn M, Hartmann C, von Deimling A. Glioma IDH1 mutation patterns off the beaten track. *Neuropathol Appl Neurobiol.* 2011; 37(4):428-430.

31. Lee D, Suh YL, Kang SY, Park TI, Jeong JY, Kim SH. IDH1 mutations in oligodendroglial tumors: comparative analysis of direct sequencing, pyrosequencing, immunohistochemistry, nested PCR and PNA-mediated clamping PCR. *Brain Pathol.* 2013; 23(3):285-293.

32. Agarwal S, Sharma MC, Jha P, et al. Comparative study of IDH1 mutations in gliomas by immunohistochemistry and DNA sequencing. *Neuro Oncol.* 2013; 15(6):718-726.

33. Preusser M, Wohrer A, Stary S, Hoftberger R, Streubel B, Hainfellner JA. Value and limitations of immunohistochemistry and gene sequencing for detection of the IDH1-R132H mutation in diffuse glioma biopsy specimens. *J Neuropathol Exp Neurol.* 2011; 70(8):715-723.

34. Tanboon J, Williams EA, Louis DN. The Diagnostic Use of Immunohistochemical Surrogates for Signature Molecular Genetic Alterations in Gliomas. *Journal of Neuropathology & Experimental Neurology.* 2015; 75(1):4-18.

35. Horbinski C. What do we know about IDH1/2 mutations so far, and how do we use it? *Acta Neuropathol.* 2013; 125(5):621-636.

WITHDRAWN
see manuscript bioRxiv/medRxiv

bioRxiv preprint doi: https://doi.org/10.1101/757385; this version posted September 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

TABLES

Table 1: Confusion matrix for T2-net and TS-net.

No. of testing subjects = 127	**Ground Truth IDH-mutated**	**Ground Truth IDH Wild-type**
T2-net IDH-mutated	**45**	**4**
T2-net IDH wild-type	**1**	**77**
TS-net IDH-mutated	**46**	**4**
TS-net IDH wild-type	**0**	**77**

WITHDRAWN
see manuscript DOI for details

Table 2: T2-net and TS-net Cross validation results.

Fold Description	**T2-net**			**TS-net**	
Fold Number	**%Accuracy**	**Dice-score**		**%Accuracy**	**Dice-score**
Fold 1	97.18	0.843		97.22	0.88
Fold 2	97.14	0.86		97.10	0.883
Fold 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**

bioRxiv preprint doi: https://doi.org/10.1101/757385; this version posted September 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

FIGURES

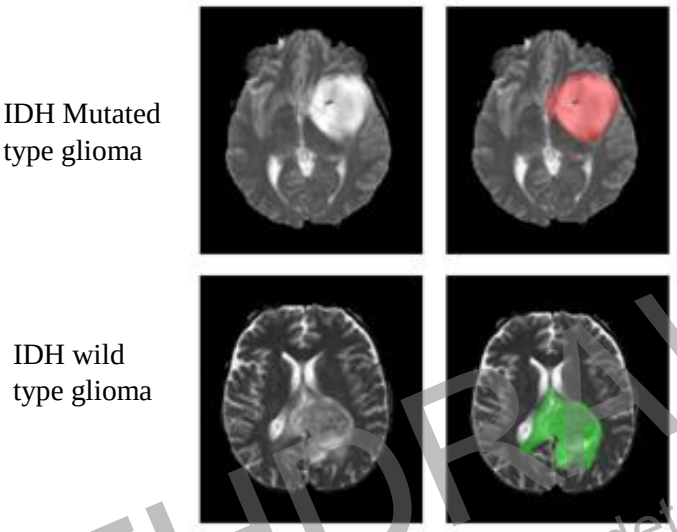
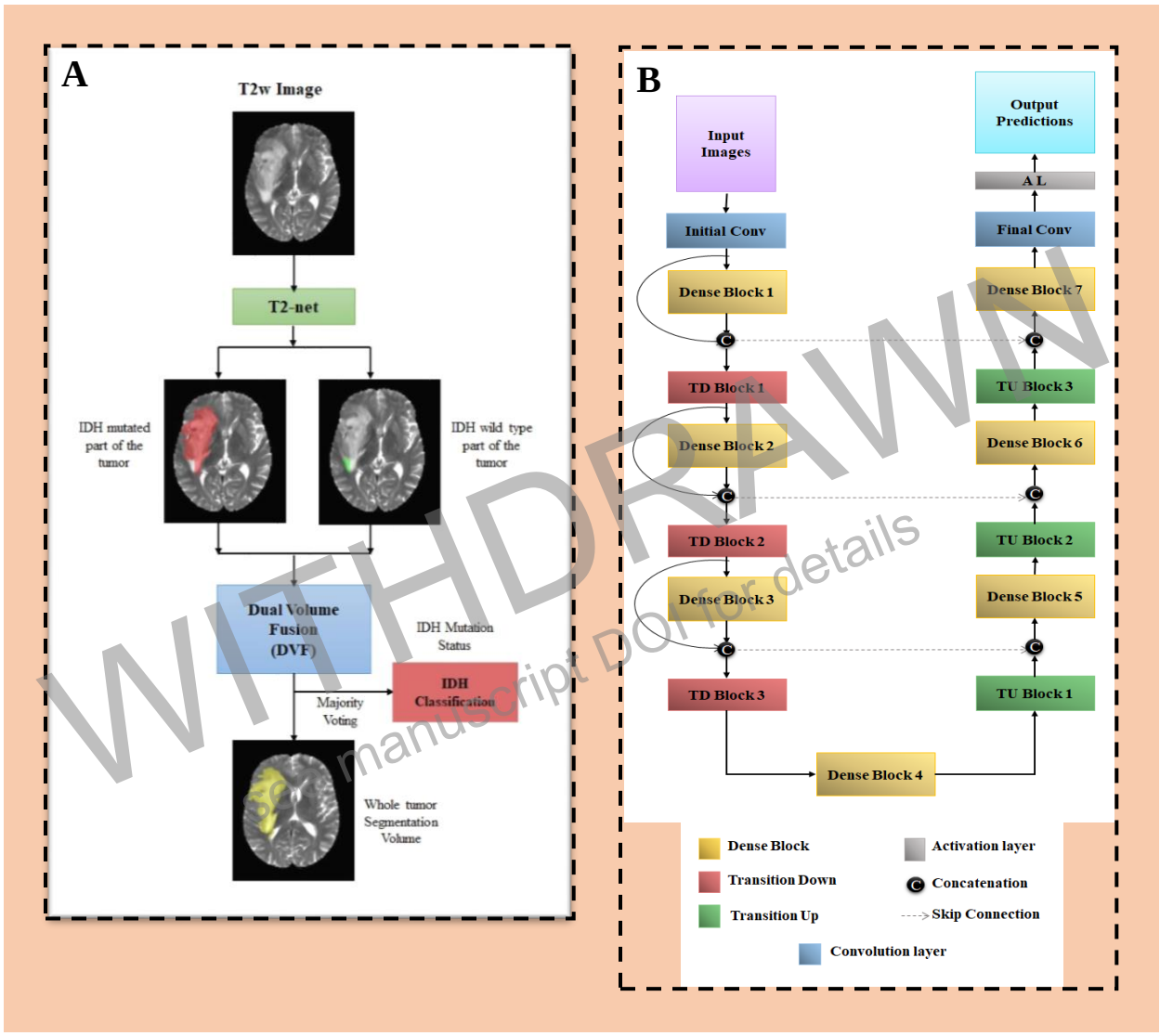


Figure 1: Ground truth whole tumor masks. Red voxels represent IDH mutated (values of 1) and green voxels represent IDH wild-type (values of 2).

WITHDRAWN
see manuscript DOI for details

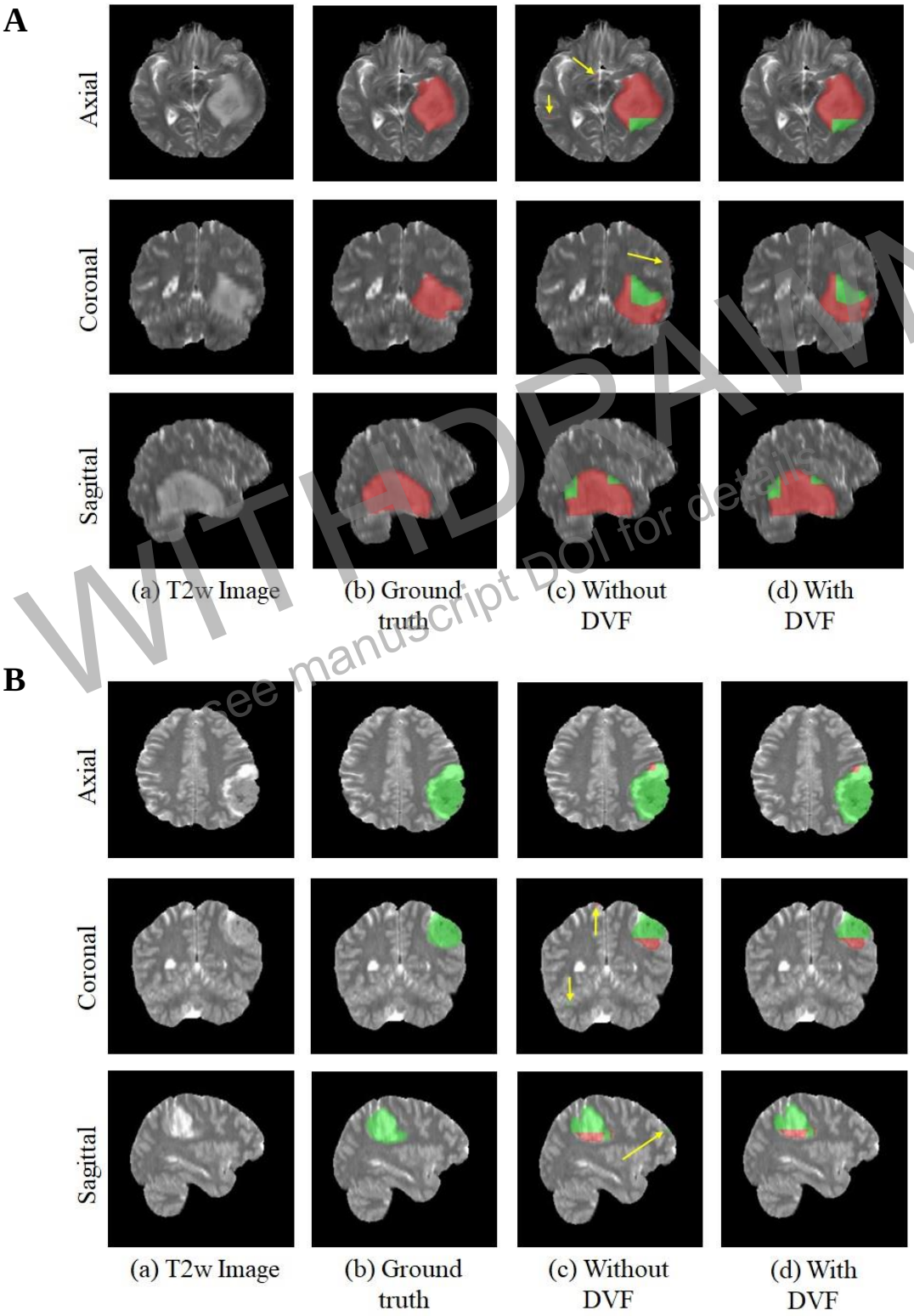
bioRxiv preprint doi: https://doi.org/10.1101/757385; this version posted September 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



WITHDRAWN
See manuscript DOI for details

Figure 2: (A) T2-net overview. Voxel-wise classification of IDH mutation status is performed to create 2 volumes (IDH mutated and IDH wild-type). Volumes are combined using dual volume fusion to eliminate false positives and generate a tumor segmentation volume. Majority voting across voxels is used to determine the overall IDH mutation status. (B) Network architecture for T2-net and TS-net. 3D-Dense-UNets were employed with 7 dense blocks, 3 transition down blocks, and 3 transition up blocks.

bioRxiv preprint doi: https://doi.org/10.1101/757385; this version posted September 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



WITHDRAWN
see manuscript DOI for details

Figure 3: (A) Example voxel-wise segmentation for an IDH mutated tumor. Native T2 image (a). Ground truth segmentation (b). Network output without DVF (c) and after DVF (d). Yellow arrows in (C) indicate false positives. Red voxels correspond to IDH mutated class and green voxels correspond to IDH wild-type. (B) Example voxel-wise segmentation for an IDH wild-type tumor.

bioRxiv preprint doi: https://doi.org/10.1101/757385; this version posted September 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

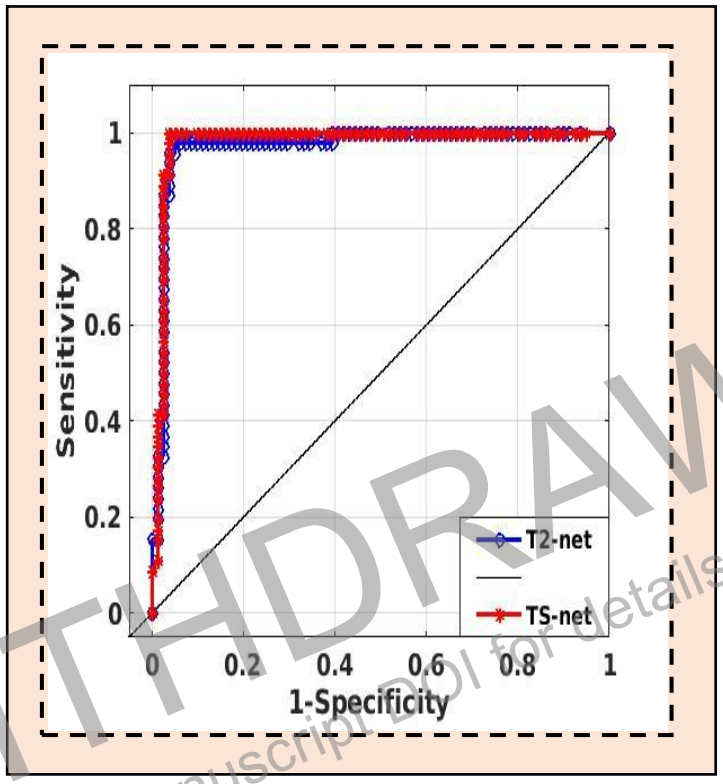


Figure 4: ROC analysis for T2-net and TS-net. Red line shows TS-net and Blue line shows T2-net ROC analysis results.

WITHDRAWN
see manuscript DOI for details