POST-HOC OVERALL SURVIVAL TIME PREDICTION FROM BRAIN MRI

Renato Hermoza[†]   *Gabriel Maicas*[†]   *Jacinto C. Nascimento*[‡]   *Gustavo Carneiro*[†]

[†]Australian Institute for Machine Learning, The University of Adelaide
[‡]Institute for Systems and Robotics, Instituto Superior Tecnico, Portugal

ABSTRACT

Overall survival (OS) time prediction is one of the most common estimates of the prognosis of gliomas and is used to design an appropriate treatment planning. State-of-the-art (SOTA) methods for OS time prediction follow a pre-hoc approach that require computing the segmentation map of the glioma tumor sub-regions (necrotic, edema tumor, enhancing tumor) for estimating OS time. However, the training of the segmentation methods require ground truth segmentation labels which are tedious and expensive to obtain. Given that most of the large-scale data sets available from hospitals are unlikely to contain such precise segmentation, those SOTA methods have limited applicability. In this paper, we introduce a new post-hoc method for OS time prediction that does not require segmentation map annotation for training. Our model uses medical image and patient demographics (represented by age) as inputs to estimate the OS time and to estimate a saliency map that localizes the tumor as a way to explain the OS time prediction in a post-hoc manner. It is worth emphasizing that although our model can localize tumors, it uses only the ground truth OS time as training signal, i.e., no segmentation labels are needed. We evaluate our post-hoc method on the Multimodal Brain Tumor Segmentation Challenge (BraTS) 2019 data set and show that it achieves competitive results compared to pre-hoc methods with the advantage of not requiring segmentation labels for training. We make our code available at https://github.com/renato145/posthocOS.

Index Terms— Glioma, Overall Survival Prediction, Saliency Maps, Brats, post-hoc, lesion localization.

1. INTRODUCTION

Gliomas are the most common brain tumors in adults, where the more severe form of the disease, high-grade gliomas, can cause a short life expectancy with a median of two years or less [1]. Treatment planning for gliomas aims to increase the overall survival (OS) time. Thus, accurate estimate of the OS time is important because it affects treatment decision and patient prognosis.

Recent methods [2, 3, 4, 5, 6] employed for OS time prediction follow a pre-hoc pipeline [7] that consists of two

stages. In the first stage, tumors are segmented into the following tissues: necrotic, edema tumor or enhancing tumor. Next, hand-designed radiomic features are extracted to train a OS time prediction model. However, the dependence on manually [6] or automatically obtained tumor segmentation maps [2, 3, 4, 5] represents an important weakness of pre-hoc approaches because not only are these segmentation maps tedious and expensive to be acquired, but they can also contain noise due to annotation disagreements between experts [8]. Furthermore, the majority of data sets available from hospitals and clinics do not contain such segmentation maps. Hence, requiring segmentation labels limits the applicability of pre-hoc methods that predict patient OS time.

In this paper, we propose a new post-hoc method [7] for OS time prediction using as inputs medical image and patient demographics (represented by age) that is trained with only the OS time label, without requiring segmentation labels. Our model consists of two parts that run in parallel. One part automatically estimates the OS time, while the other part highlights the relevant image region that localizes the tumor, which is used to explain the OS time estimation in a post-hoc manner. Contrary to the pre-hoc pipeline, our method can estimate the OS time and provide the tumor localization without the use of ground truth segmentation labels and of hand-designed radiomic features.

We evaluate our model on the Multimodal Brain Tumor Segmentation Challenge (BraTS) 2019 data set. Results show that our post-hoc model trained with no segmentation labels achieves comparable results to state-of-the-art (SOTA) pre-hoc models that rely on segmentation labels during training.

2. RELATED WORKS

OS time prediction is a key task that guides treatment plan for patients suffering from gliomas. With the aim to foster research in OS time prediction, BraTS [1, 8, 9] is a publicly available data set that contains the ground truth of OS time and tumor segmentation from brain magnetic resonance images (MRI).

Current SOTA methods for OS time prediction [2, 3, 4, 5, 6] follow a pre-hoc pipeline consisting of two stages: 1) segmentation of the tumor sub-regions; and 2) use of the tumor segmentation map from 1) to train a OS time prediction

model. In the first stage, the tumor segmentation maps are obtained manually [6] or automatically [2, 3, 4, 5]. However, both manual and automated approaches require ground truth segmentation labels, which are costly to obtain. In the second stage, most prediction models use hand-crafted radiomics features [2, 3, 4, 5]. These features include shape, intensity and texture features, extracted from the brain image and the tumor sub-regions segmentation. For the OS time prediction task, [2, 3] train a random forest regressor, [4] rely on a linear model and [5] uses a neural network with two hidden layers. More recently, Zhou *et al.* [6] used image and radiomics features for the OS prediction model, but still relied on segmentation labels to allow the extraction of patches centered at the tumor.

Although the pre-hoc approaches above have advanced the field of OS time prediction from brain MRI images, they require ground truth tumor segmentation labels for training. Such requirement is a major weakness of these approaches because the segmentations labels are costly to obtain and are, in general, fairly unreliable given that annotations disagreements can be common among experts. On the contrary, post-hoc methods have been shown [7] to be competitive with the performance of pre-hoc methods for prediction tasks without requiring ground truth segmentation for training. In this paper, we propose a post-hoc approach that performs OS time prediction after being trained using only OS time ground truth. Additionally, our post-hoc method provides a saliency map that shows the relevant area of the image for the OS prediction, which is expected to be the tumor localization – such saliency map can be used to explain the OS time prediction.

3. METHOD

The proposed post-hoc method estimates the number of days of OS time and provides the tumor location from pre-operative brain MRI. For risk prediction, we split the OS time into N bins, where each bin represents a particular range of days for the estimation. Note that the OS time estimation will steadily decrease for consecutive bins because if the estimation for 100 days is $X\%$, then the risk for 200 days must be equal or smaller. The model outputs a survival probability associated with each bin, and the final OS time prediction is computed by summing the product between the number of days for each bin and the survival probability produced by the model. For model explainability, we use the saliency maps obtained during the process of estimating the risk. See Fig. 1 for a graphical summary of our method. We describe the data set in Sec. 3.1, and our method in Sec. 3.2.

3.1. Data Set

The data set is defined by $\mathcal{D} = \{(\mathbf{x}_i, \mathbf{y}_i, \mathbf{s}_i, \mathbf{a}_i)\}_{i=1}^{|\mathcal{D}|}$, where $\mathbf{x}_i : \Omega \to \mathbb{R}^4$ denotes a brain MRI image with 4 modalities (Flair, T1, T1ce and T2), and Ω is the image lattice,
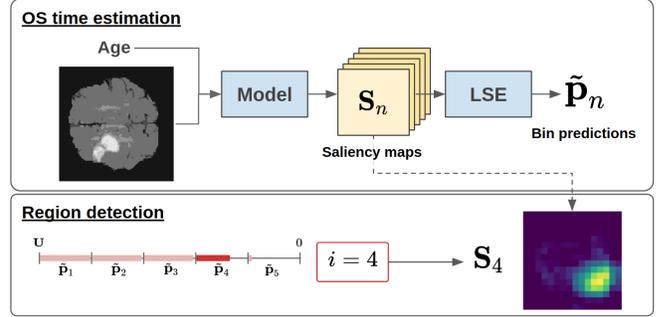
Fig. 1. The proposed post-hoc model consists of two parts: 1) OS time estimation, where the model splits the prediction into bins representing an increasing OS time estimation (in days); and 2) region detection, where the saliency maps $\{\mathbf{S}_n\}_{n=1}^N$ extracted in 1) are used to highlight relevant regions of the image used for OS time estimation. The final OS time prediction is obtained from the sum of $\{\tilde{\mathbf{p}}_n\}_{n=1}^N$, with each $\tilde{\mathbf{p}}_n$ denoting the product of the number of days in bin n and the survival probability – see Eq. 4.

$\mathbf{y}_i \in \mathbb{N}$ indicates the OS time defined in days, $\mathbf{s}_i : \Omega \to \mathcal{Y}$, $\mathcal{Y} = \{Background, Tumor\}$ is a binary map that represents the whole tumor segmentation annotation, and $\mathbf{a}_i \in \mathbb{N}$ indicates the age in years. Note that $\mathbf{s}_i$ is only used for evaluating our proposed weakly supervised tumor segmentation method – we never use those segmentations for training.

3.2. Model

Our proposed model uses a fully convolutional architecture [10] to extract features $\mathbf{F}_i \in \mathbb{R}^{Q \times V^3}$ from $\mathbf{x}_i$, which contains Q feature maps, and a simple linear layer to extract $\tilde{\mathbf{s}}_i \in \mathbb{R}^Q$ from $\mathbf{a}_i$. Then $\mathbf{F}_i$ and $\tilde{\mathbf{s}}_i$ are combined as in:

$$\tilde{\mathbf{F}}_i = \mathbf{F}_i + b(\tilde{\mathbf{s}}_i; V^3), \tag{1}$$

where $b(\tilde{\mathbf{s}}_i; V^3) \in \mathbb{R}^{Q \times V^3}$ takes each $\tilde{\mathbf{s}}_{i,q}$, for $q \in \{1, ..., Q\}$, and broadcasts to the respective V^3 dimensions. A final convolution layer takes $\tilde{\mathbf{F}}$ from (1) as input to produce N saliency maps $\{\mathbf{S}_{i,n}\}_{n=1}^N$, each of size V^3, where N is the number of bins used to discretize $\mathbf{y}_i$. More specifically, each bin $\mathbf{b}_n$ represents a number of survival days $\{\mathbf{b}_n\}_{n=1}^N$ using equidistant values between 0 and the maximum number of survival days in the data set, which is $U = 1800$ days. In fact, the maximum number of survival days represents an upper-limit that is obtained by summing the values for each bin: $U = \sum_{n=1}^N \mathbf{b}_n$, with $\mathbf{b}_n = U/N$ for all $n \in \{1, ..., N\}$. We apply a global pooling on the saliency maps $\mathbf{S}_{i,n}$ to obtain the predictions for each bin $\{\tilde{\mathbf{p}}_{i,n}\}_{n=1}^N$ as follows:

$$\tilde{\mathbf{p}}_{i,n} = \mathbf{b}_n \cdot \mathbf{p}_{i,n}, \tag{2}$$

where

$$\mathbf{p}_{i,n} = \sigma(LSE(\mathbf{S}_{i,n})), \tag{3}$$
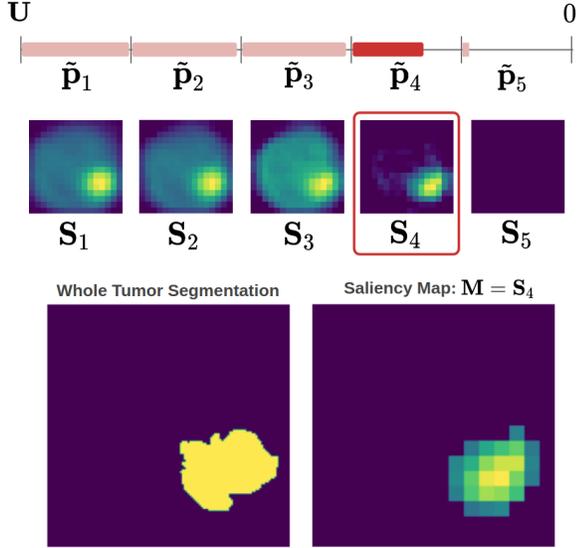
Fig. 2. Comparison between the true label segmentation (bottom-left) for the whole tumor and the saliency map extracted from the bin predictions (bottom-right). The selected saliency map corresponds to the last bin responsible for the prediction (see Eq. 7).

with σ denoting the sigmoid activation function, and LSE the log-sum-exp pooling function.

We compute the final OS time prediction $\tilde{\mathbf{y}}_i$ by subtracting the predictions for each bin $\tilde{\mathbf{p}}_{i,n}$ from the upper limit U:

$$\tilde{\mathbf{y}}_i = f(\mathbf{x}_i, \mathbf{a}_i; \theta) = U - \sum_{n=1}^{N} \tilde{\mathbf{p}}_{i,n}, \qquad (4)$$

where $f(\mathbf{x}_i, \mathbf{a}_i; \theta)$ represents the whole model parameterized by θ (θ denotes all weights of the deep learning model), taking image $\mathbf{x}_i$ and age $\mathbf{a}_i$ as inputs. The rationale behind (4) is that we expect the saliency maps to be active on the regions responsible for OS time survival. Thus a higher $\mathbf{p}_{i,n}$ indicates a lower OS time prediction by the number of days in bin n.

The training of our model minimizes the loss function:

$$\ell_{Total}(\mathcal{D}, \theta) = \frac{1}{|\mathcal{D}|} \sum_{i=1}^{|\mathcal{D}|} \ell_{mae}(\tilde{\mathbf{y}}_i, \mathbf{y}_i) + \alpha \ell_{pen}(\mathbf{p}_i), \quad (5)$$

where $\ell_{mae}(\tilde{\mathbf{y}}_i, \mathbf{y}_i)$ is the mean absolute error (MAE) between the ground truth label $\mathbf{y}_i$ and the model prediction $\tilde{\mathbf{y}}_i$ for OS time in (4), $\ell_{pen}(\mathbf{p}_i)$ is a smoothness term that penalizes inconsistent predictions between neighboring bins which forces consecutive bins of the i-th training prediction $\mathbf{p}_i$ (defined in Eq. 2) to represent a risk estimation in descending order, with minimal amount of transition between neighboring bins. In particular, in this loss, a bin n of $\mathbf{p}_i$ should be active only when all previous bins from 1 to $n-1$ have become active. Therefore, the penalization term to force this
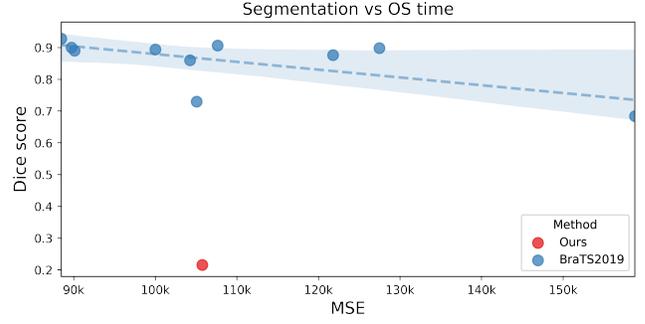
Fig. 3. Comparison of survival (MSE - smaller is better) and segmentation results (Dice - larger is better) on BraTS 2019 validation set [11]. Note the weak correlation between segmentation and OS results and that our method (red dot) has a competitive MSE even with a low Dice result.

constrain is defined as follows:

$$\ell_{pen}(\mathbf{p}_i) = \frac{1}{N-1} \sum_{n=1}^{N-1} \max(0, (\mathbf{p}_{i,n+1} - \mathbf{p}_{i,n})), \qquad (6)$$

where $\mathbf{p}_{i,n}$ represents the prediction of the i-th training sample at bin n. The weight α in (5) controls the smoothness strength.

For the explanation of the predictions of the post-hoc model, we observe that the saliency maps $\{\mathbf{S}_{i,n}\}_{n=1}^{N}$ should contain the interpretation of the OS time prediction, represented by the tumor localization. We use the saliency map $\mathbf{S}_{i,n^*}$, where n^* corresponds to the index of the last bin responsible for the prediction from the model (see Fig. 2):

$$n^* = \underset{n \in \{1,...,N\}}{\mathrm{argmax}} \{\min(\mathbf{p}_{i,n}, |1 - \mathbf{p}_{i,n})\} \qquad (7)$$

The inference of our method uses the forward pass of the trained model and gives the final OS time prediction using (4) and the saliency map from (7).

4. EXPERIMENTS

4.1. Data Set

We conduct the experiments in this paper using the data set from the Multimodal Brain Tumor Segmentation Challenge (BraTS) 2019 [1, 9, 8]. The data set contains pre-operative MRI scans of glioblastoma (GBM/HGG) and low-grade glioma (LGG) and focuses on the segmentation of brain tumors and the prediction of patient OS time. As the focus of this method is OS time prediction, we use a subset of 210 cases that includes survival labels for training, and we use 20% as internal validation for model selection (i.e., hyperparameter estimation). Also, a validation set of 29 cases is available for the OS time prediction task, while ground truth

annotations are not available for this set, it is possible to evaluate the results using a provided web service for the challenge. Each case includes MRI images for four modalities: Flair, T1, T1ce and T2. The ground truth for segmentation is marked with three labels: necrotic and non-enhancing tumor (label 1), edema tumor (label 2), and enhancing tumor (label 4). For evaluation purposes we use the complete tumor extent referred to as "whole tumor", as the saliency map we obtain does not distinguish between labels. For the task of OS prediction, each case includes the age of patient, a resection status: Gross Total Resection (GTR) or Subtotal Resection (STR), and the OS defined in days.

4.2. Experimental Set Up

Each modality of the MRI image and the age are normalized with mean value of zero and standard deviation of one, then, the image data of the four modalities for each case is down-sampled and stacked into a multi-channel data with size $128^3 \times 4$. During training we apply random scaling between 1 and 1.1 as data augmentation. The model consists of 4 blocks of: convolution, activation and batch normalization, where the activation function is a Leaky ReLU [12] with a 0.1 negative slope. The model is trained using the Adam [13] optimizer with a momentum of 0.9, weight decay of 0.001, a mini-batch size of 8 and α of 10,000.

We use the Mean Squared Error (MSE) and classification accuracy to evaluate our model. For the classification, we use the official evaluation setup [8] that splits the OS time into three groups: 1) short-survivors (less than 10 months), 2) mid-survivors (between 10 and 15 months), and 3) and long-survivors (more than 15 months). To assess the interpretability of our saliency maps, we evaluate the weakly supervised segmentation by using a dice score between the top 5% activations of the saliency map $\mathbf{S}_{i,n^*}$ from (7) and the ground truth segmentation for the whole tumor.

4.3. Results

In Table 1 we compare the OS time predictions of our method and the best four results from the BraTS 2019 competition [2, 3, 4, 5]. Note that our proposed post-hoc method only uses OS time labels during the training process, while the rest of methods follow a pre-hoc pipeline that requires segmentation labels and hand-designed radiomics features. We also compare segmentation versus OS time prediction results published on BraTS 2019 [11] in Fig. 3 and note that better segmentation accuracy does not imply better OS time prediction accuracy. This is an important point because even though the segmentation result produced by our weakly-supervised saliency map (with dice around 0.2) is much worse than the fully supervised SOTA results (dice > 0.9), our OS time prediction is competitive with the SOTA. In Tab. 2 we provide an ablation study to examine the effects of including the age of the patient as an input to the model. Moreover, in Tab. 2 we also study an

Table 1. Comparison on OS time prediction results on BraTS 2019 validation set. Note that our method uses only survival labels while the others use segmentation and survival labels for training. The evaluation metrics are: accuracy, mean squared error, median squared error, standard squared error and SpearmanR.

Paper	Acc	MSE	medianSE	stdSE	SpearmanR
[2]	0.586	105,062	16,461	188,752	0.404
[3]	0.590	89,724	36,121	-	0.360
[4]	0.310	107,639	77,906	109,587	0.204
[5]	0.448	100,000	49,300	135,000	0.250
Ours	0.517	105,746	51,962	181,311	0.248

Table 2. Ablation experiments on the use of age for OS time prediction and a simple regression method vs our proposed post-hoc method. Acc refers to accuracy and MSE to mean squared error.

Method	Train set		Validation set	
	Acc	MSE	Acc	MSE
Regression	0.286	73,878	0.207	223,604
Regression + Age	0.405	75,241	0.379	140,457
Post-hoc	0.405	77,439	0.379	134,213
Post-hoc + Age	**0.548**	**57,198**	**0.517**	**105,746**

alternative method for our proposed survival function, based on a simple regressor with a similar model architecture.

5. DISCUSSION AND CONCLUSION

In this paper, we proposed the use of a post-hoc method for OS time prediction from brain MRI and patient demographics (represented by age) of cases with high-grade gliomas. SOTA methods for OS time prediction follow a pre-hoc pipeline that requires for training the OS time labels and tumor segmentation annotations which are costly to obtain and may contain annotation inconsistencies. On the contrary, as shown in Table 1, our proposed post-hoc method is competitive to SOTA pre-hoc methods and only requires the OS time labels for training – in addition, it produces a saliency map localizing the tumor in the image for explaining the OS time estimation. Our post-hoc method also outperforms a regression model using a similar architecture, as shown in Table 2. Note that as shown in Fig. 3, extracting features using better segmentation results does not imply an increase on OS time prediction accuracy. We leave for future work the design of a post-hoc method that can improve both the OS time prediction accuracy as well as the precision of the saliency map segmentation.

Compliance with Ethical Standards

This research study was conducted retrospectively using human subject data made available in open access by [14, 15, 16]. Ethical approval was not required as confirmed by the license attached with the open access data.

Acknowledgments

6. REFERENCES

[1] Bjoern H. Menze et al., "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)," *IEEE transactions on medical imaging*, vol. 34, no. 10, pp. 1993–2024, Oct. 2015.

[2] Rupal R Agravat and Mehul S Raval, "Brain tumor segmentation and survival prediction," in *International MICCAI Brainlesion Workshop*. Springer, 2019, pp. 338–348.

[3] Shuo Wang, Chengliang Dai, Yuanhan Mo, Elsa Angelini, Yike Guo, and Wenjia Bai, "Automatic brain tumour segmentation and biophysics-guided survival prediction," in *International MICCAI Brainlesion Workshop*. Springer, 2019, pp. 61–72.

[4] Xue Feng, Quan Dou, Nicholas Tustison, and Craig Meyer, "Brain tumor segmentation with uncertainty estimation and overall survival prediction," in *International MICCAI Brainlesion Workshop*. Springer, 2019, pp. 304–314.

[5] Feifan Wang, Runzhou Jiang, Liqin Zheng, Chun Meng, and Bharat Biswal, "3d u-net based brain tumor segmentation and survival days prediction," in *International MICCAI Brainlesion Workshop*. Springer, 2019, pp. 131–141.

[6] Tao Zhou, Huazhu Fu, Yu Zhang, Changqing Zhang, Xiankai Lu, Jianbing Shen, and Ling Shao, "M2net: Multi-modal multi-channel network for overall survival time prediction of brain tumor patients," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2020.

[7] Gabriel Maicas, Andrew P Bradley, Jacinto C Nascimento, Ian Reid, and Gustavo Carneiro, "Pre and post-hoc diagnosis and interpretation of malignancy from breast dce-mri," *Medical Image Analysis*, vol. 58, pp. 101562, 2019.

[8] Spyridon Bakas et al., "Identifying the Best Machine Learning Algorithms for Brain Tumor Segmentation, Progression Assessment, and Overall Survival Prediction in the BRATS Challenge," *arXiv:1811.02629 [cs, stat]*, Apr. 2019, arXiv: 1811.02629.

[9] Spyridon Bakas et al., "Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features," *Scientific Data*, vol. 4, pp. 170117, 2017.

[10] Maxime Oquab, Léon Bottou, Ivan Laptev, and Josef Sivic, "Is object localization for free? - Weakly-supervised learning with convolutional neural networks," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015, pp. 685–694, ISSN: 1063-6919, 1063-6919.

[11] Alessandro Crimi and Spyridon Bakas, *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 5th International Workshop, BrainLes 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 17, 2019, Revised Selected Papers, Part I*, vol. 11992, Springer Nature, 2020.

[12] Andrew L Maas, Awni Y Hannun, and Andrew Y Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proc. icml*, 2013, vol. 30, p. 3, Issue: 1.

[13] Diederik P. Kingma and Jimmy Ba, "Adam: A Method for Stochastic Optimization," *International Conference on Learning Representations (ICLR)*, 2015.

[14] Spyridon Bakas, Hamed Akbari, Aristeidis Sotiras, Michel Bilello, Martin Rozycki, Justin Kirby, John Freymann, Keyvan Farahani, and Christos Davatzikos, "Segmentation Labels for the Pre-operative Scans of the TCGA-GBM collection," 2017, type: dataset.

[15] Spyridon Bakas, Hamed Akbari, Aristeidis Sotiras, Michel Bilello, Martin Rozycki, Justin Kirby, John Freymann, Keyvan Farahani, and Christos Davatzikos, "Segmentation Labels for the Pre-operative Scans of the TCGA-LGG collection," 2017, type: dataset.

[16] Kenneth Clark, Bruce Vendt, Kirk Smith, John Freymann, Justin Kirby, Paul Koppel, Stephen Moore, Stanley Phillips, David Maffitt, Michael Pringle, Lawrence Tarbox, and Fred Prior, "The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository," *Journal of Digital Imaging*, vol. 26, no. 6, pp. 1045–1057, Dec. 2013.