End-to-End Evidential-Efficient Net for Radiomics Analysis of Brain MRI to Predict Oncogene Expression and Overall Survival

Yingjie Feng[1], Jun Wang[1], Dongsheng An[2], Xianfeng Gu[2], Xiaoyin Xu[3], and Min Zhang[4] (✉)

[1] School of Software Technology, Zhejiang University, Hangzhou, China
[2] Department of Computer Science, Stony Brook University, Stony Brook, NY, USA
[3] Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA
[4] College of Computer Science and Technology, Zhejiang University, Hangzhou, China
min_zhang@zju.edu.cn

Abstract. We presented a novel radiomics approach using multimodality MRI to predict the expression of an oncogene (O6-Methylguanine-DNA methyltransferase, MGMT) and overall survival (OS) of glioblastoma (GBM) patients. Specifically, we employed an EffNetV2-T, which was down scaled and modified from EfficientNetV2, as the feature extractor. Besides, we used evidential layers based to control the distribution of prediction outputs. The evidential layers help to classify the high-dimensional radiomics features to predict the methylation status of MGMT and OS. Tests showed that our model achieved an accuracy of 0.844, making it possible to use as a clinic-enabling technique in the diagnosing and management of GBM. Comparison results indicated that our method performed better than existing work.

Keywords: Evidential Deep Learning · EfficientNet-V2 · Radiomics · MGMT Promoter Methylation Prediction · Brain Tumor.

1 Introduction

GBM is the most lethal type of brain tumor, constituting 60% of malignant adult brain tumors [24]. Diverse MRI modalities are sensitive to different tissue and thus can provide rich information about GBM, including shedding new insight into the oncogenetic status of GBM. With the rapid development of deep learning techniques, good performance in classification [8, 16] and regression tasks [13, 4, 7, 22, 10] on MRI to address clinical questions has been achieved. In this study, we designed an end-to-end deep learning model for predicting OS and the status of MGMT methylation using multimodality MRI.

1.1 Importance of MGMT Prediction

One important oncogenetic characteristic of GBM is the expression of MGMT. The methylation status of MGMT can be used not only as a diagnostic ba-

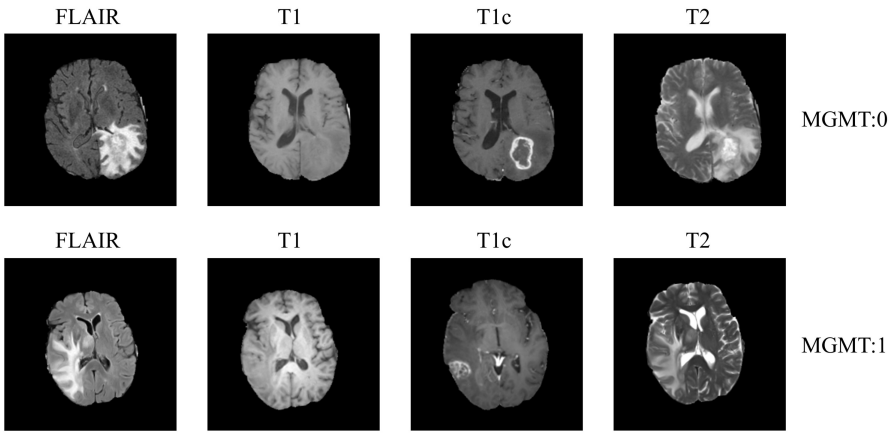

Fig. 1. Four MRI modalities of patients with different MGMT promoter methylation status.

sis but also for prognostic evaluation, predicting the sensitivity of radiotherapy and chemotherapy, and providing effective information for precise treatment plans [11]. Though the status of MGMT can be assessed by biopsy, it is often beneficial to predict whether a GBM expresses MGMT at the earliest time point for better design of treatment and predicting the progress of the tumor. There has been great interest in using radiomics methods to infer the methylation status of MGMT in GBM [21, 17, 14]. Recently, several studies have shown appealing results in identifying the methylation status of MGMT promoters using deep learning methods [8, 27, 16]. Chang et al. designed a CNN based on ResNet to classify the methylation status of MGMT on multimodality MRI scans (T1c, T2, FLAIR) and reached a mean accuracy score of 83% on 5-fold cross-validation [8]. In a work by Yogananda et al., researchers proposed a MGMT-net based on 3D-Dense-UNets to use T2 MRI only for predicting the status of MGMT [27] and achieved a mean accuracy of 94.73% with an AUC of 0.93 on 3-fold cross-validation. Using T2 MRI, Korfiatis et al. found that ResNet50, with an accuracy of 94.90%, gave the most accurate results in this task.

While most methods predict MGMT status by focusing on features extracted from the tumor region, studies using whole brain MRI are not widely reported. On the whole brain MRI, Han and Kamdar designed a new model to predict the status of MGMT without tumor segmentation [11]. Their results showed an accuracy of 67% on the validation data and 62% on the test data [11], indicating the usefulness of whole brain analysis in such applications. In 2021, the Brain Tumor Radiogenomic Classification challenge [3], organized by the Radiological Society of North America (RSNA) and the Medical Image Computing and Computer-Assisted Interventions (MICCAI) conference, contributed the most extensive image dataset for MGMT prediction without tumor segmentation.

Participants were not able to achieve more than 0.62 AUC on the validation set [19], pointing to the challenge in accurately predicting MGMT status.

1.2 Importance of OS Prediction

Accurate estimate of OS is important for assessing the prognosis of GBM as the estimate is used to design appropriate treatment [18]. Most previous methods [1, 25, 26, 9, 28] for OS prediction employ a two-step pipeline that includes: 1) segmenting whole tumor region into necrotic, edema, and enhancing tumor regions; and 2) extracting radiomics features to train a prediction model. However, these approaches have some obvious shortcomings. On the one hand, the majority of datasets from hospitals do not contain segmentation maps, which are time-consuming and labor-intensive to acquire. On the other hand, annotation disagreements among experts can also cause inconsistency in tumor segmentation. Therefore, existing methods that predict OS are limited by the requirement for segmentation maps.

1.3 Challenges

In predicting MGMT status and OS, challenges experienced by existing works are: 1) the predicted values of the output are concentrated near the mean of ground truth, and the models do not have a good discriminating ability; and 2) given an input, a model must and can only create a single predicted value, and it is not clear that the uncertainty in prediction is sufficiently incorporated into the model. In other words, a model does not know its own limitation in dealing with uncertainty.

1.4 Contribution of This Work

A novel contribution of this work is that we proposed an EDL-based approach as a classifier for predicting the methylation status of MGMT. We used EfficientNet as a feature extractor and reached the best performance on the task of predicting the methylation status of MGMT. Our model also achieved similar performance on predicting patients' OS, demonstrating its broad applicability. Novelties of this work are as follows.

1. In general, when applying radiomics analysis with deep-learning methods, the learning models do not know the uncertainty of the prediction. To address this problem, we used Evidential-Regression to implement the final prediction, with uncertainty information attached to the prediction.
2. In Evidential-Regression, we used *a priori* NIG (Normal Inverse-Gamma) of a Gaussian distribution to fit the data. And we implemented a network structure similar to a multilayer perceptron to enhance the performance of Evidential-Regression.

2 Methods

2.1 Overview

Evidential Deep Learning As an important branch of prediction uncertainty modeling, EDL builds on previous works of uncertainty estimation and modeling probability distributions using neural networks. Unlike BNNs, which indirectly infer prediction uncertainty through weight uncertainties, EDL employs the theory of subjective logic to explicitly estimate the uncertainty [20]. EDL treats the prediction as subjective opinions and uses a deterministic neural network to accumulate evidence that leads to these opinions. Though EDL is usually employed to address the "know unknown" flaws, for its powefulness, as shown in [20], EDL can be used for handling uncertainties in classification. In this work, we used EDL to produce evidential distributions that better separate features arisen from binary or multi-nary populations. Then starting with the EDL-generated distributions, the learner can achieve higher performance in classification. In our case, we consider MGMT prediction as a binary classification task such that, given a GBM case, a regression model assigns probabilities to whether the case is MGMT mutant or not. So we use Evidential-Regression in EDL to perform classification for both MGMT and OS tasks.

Workflow of Our Algorithm Our algorithm consists of several major steps, namely, feature extraction, EDL for generating evidential distributions, and classifiers. After being pre-processed, data are input into feature extractor to generate high dimensional feature maps. Evidential predictor outputs the evidential distribution based on the feature maps. Last, the algorithm computes the result from the distribution parameters. For MGMT classification, the result is a bi-

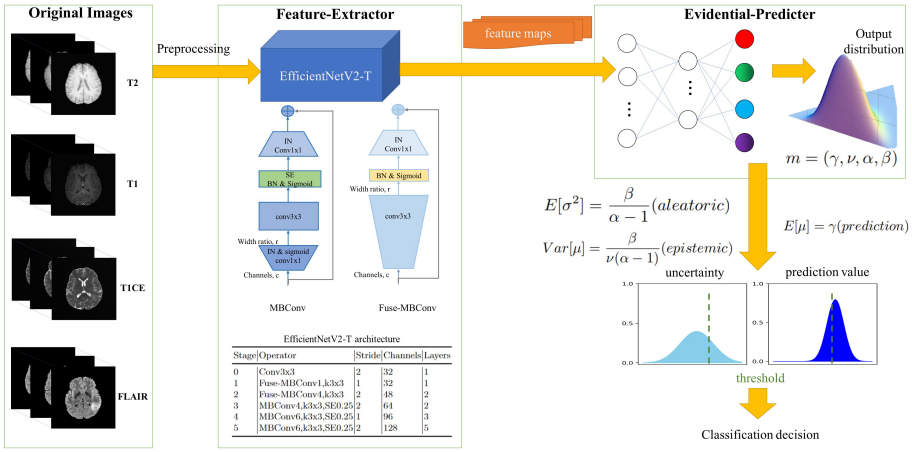

Fig. 2. Architecture of our proposed Evidential-Efficient-Net.

nary assignment. For OS prediction, the result is a probability on whether the GBM patient will have a short-, medium-, or long-term survival.

2.2 Evidential Regression

For regression problems, generally, loss function $L_i(w) = \frac{1}{2}||y_i - f(x_i; w)||^2$ is used for optimization, but such a loss function can only characterize how close the predicted value is to the data. In other words, it can only be used to represent uncertainty in the data, also known as the aleatoric uncertainty [15].However, using evidential regression with uncertainty in the prediction results, it is possible to explicitly calculate the part of uncertainty caused by the model's predictions.

Although the ground-truth labels only have binary value of 0 or 1, combined with the prediction of the NormalCNN shown in Fig. 3 and other related knowledge, it is expected that the classification probability values should be close to the normal distribution. To reach the approximation for the true posterior which is close to the normal distribution, the evidential distribution takes the form of the Gaussian conjugate prior – NIG distribution – such that

$$P(\mu, \sigma^2|\gamma, \nu, \alpha, \beta) = \frac{\beta^\alpha \sqrt{v}}{\Gamma(\alpha) \sqrt{2\pi\sigma^2}} (\frac{1}{\sigma^2})^{\alpha+1} \exp\left\{-\frac{2\beta + \nu(\gamma - \mu)^2}{2\sigma^2}\right\}.$$

With the NIG distribution, we can calculate the prediction, aleatoric uncertainty, also known as statistical or data uncertainty, and epistemic uncertainty, which presents the estimated uncertainty in the prediction, as follows:

$$E[\mu] = \gamma \text{ (prediction)}, E[\sigma^2] = \frac{\beta}{\alpha - 1} \text{ (aleatoric)}, Var[\mu] = \frac{\beta}{\nu(\alpha - 1)} \text{ (epistemic)}.$$

After obtaining an evidential distribution expression that captures both uncertainties at the same time, model training becomes a process of accumulating evidence on the model that supports our observations or maximizing the ability of the model to fit and reduce the impact of erroneous evidence on the model. In terms of accumulating evidence, we use the Student-t distribution derivation to obtain the loss function of the negative log likelihood part:

$$L_i^{NLL}(w) = \frac{1}{2}\log(\frac{\pi}{\nu}) - \alpha\log(\Omega) + (\alpha + \frac{1}{2})\log((y_i - \gamma)^2\nu + \Omega) + \log(\frac{\Gamma(\alpha)}{\Gamma(\alpha + \frac{1}{2})})$$

where $\Omega = 2\beta(1 + \nu)$. This loss provides an objective target for training a neural network to output parameters of an NIG distribution by maximizing the model evidence to fit with the observations. In terms of reducing the impact of evidence on errors, we use the evidence regularizer proposed by Amini et al. [2]: $L_i^R(w) = |y_i - \gamma| \cdot (2\nu + \alpha)$. The total loss $L_i(w)$ is the sum of two losses and a regularization coefficient λ to adjust their relative importance:

$$L_i(w) = L_i^{NLL}(w) + \lambda L_i^R(w)$$

2.3 Model

The core component of our method is the EfficientNetV2, which, in our method, was scaled down to a tiny net that we named EfficientNetV2-T. It preserved the MBConv and Fused-MBConv block in the original architecture searching space of EfficientNet. We adjusted the channels and the SE (squeeze and excitation) layer inside the blocks to accommodate four MRI modalities. As the original EfficientNetV2-S proves to be too deep for our task since it caused overfitting and reached a low bottleneck, we introduced dropout layers and adjusted regularization in our design to address this problem. With the extracted feature map, the evidential predictor exports the parameters $m = (\gamma, \nu, \alpha, \beta)$ of the evidential distribution. In the end, we used output distribution to calculate predicted values, classification results, aleatoric and epistemic uncertainties.

3 Experiments and Results

3.1 Datasets and Implementation

Dataset and Evaluation Metrics for MGMT Predication This study used the dataset from the Brain Tumor Radiogenomic Classification challenge [3]. Multimodality MRI scans (T1, T1c, T2, FLAIR) of 585 GBM patients were provided in DICOM format. Pre-processing included skull stripping, isotropic resolution uniformization, and co-registration to the same anatomical template (SRI24). The dataset consists of 307 methylated cases and 278 unmethylated cases. Three cases were removed due to data quality issue. We then randomly separated the cases into a training group of 466 patients and a test group of 116 patients. For each modality, we converted and resampled DICOM to 3D $16 \times 256 \times 256$ NIFITI data. Four modalities data constitute the four input channels to our model. To evaluate the performance of our model, we implemented 5-fold cross-validation on the dataset and calculated several performance metrics for each patients. In this study, six widely used performance metrics are used to compare our model with several state-of-the-art models for predicting MGMT methylation, including overall accuracy (OA), sensitivity (SN), specificity (SP), positive predictive value (PPV), negative predictive value (NPV), the area under the receiver operating characteristic (ROC) curve (AUC).

Dataset and Evaluation Metrics for OS Prediction We trained and tested the model for predicting OS of GBM patients with BraTS2019 datasets[18, 5, 6], which contain four MRI modality scans and survival labels of 210 patients. Like with BraTS2021, we resampled scans of each modality to $128 \times 128 \times 128$ and sent them to four channels of the feature-extractor. The ages of patients were appended with feature maps and sent to the evidential predictor. For this task, we calculated MSE (mean squared error) and classification accuracy for each patient. Classification is based on an official evaluation setup that categorizes lengths of OS into three groups: 1) short-term survivors (less than 300 days), 2) mid-term survivors (between 300 and 450 days), and 3) long-term survivors (more than 450 days).

Table 1. Results on MGMT Prediction.

Methods	AUC	OA	SN	SP	PPV	NPV
Kaggle Winner	0.605	0.610	0.702	0.509	0.615	0.618
ResNet [12]	0.611	0.64	-	-	-	-
Saeed et al. [19]	0.630	-	-	-	-	-
EfficientNetV2-S [23]	0.637	0.67	-	-	-	-
Proposed Methods	**0.809**	**0.844**	**0.819**	**0.886**	**0.921**	**0.750**
Our Method (without EDL)	0.613	0.62	-	-	-	-
EfficientNetV2-S (with EDL)	0.729	0.73	-	-	-	-

Implementation Our model was implemented in a PyTorch 16.1 environment. The training and testing process was performed on a PC equipped with eight NVIDIA GTX 2080Ti GPUs. In training, we chose the Adam optimizer. The learning rate was 0.01, the batch-size was set to 16, and the model parameters were randomly initialized.

3.2 Results of Predicting MGMT

For MGMT prediction, the test result of our model is shown in Table 1. Compared with two state-of-the-art methods, it is seen that our method had significantly better performance across all metrics. Also, we designed ablation experiments to compare and verify the effectiveness of our components. The EfficentNetV2-S with EDL gave an increase of 0.092 on AUC and 0.06 on accuracy, respectively, as compared to the original network. Together with the comparison between proposed methods and our method without EDL, we observed that the effect of EDL is significant. Besides, the better results of proposed methods over EfficientNetV2-S with EDL proves that our modification of EfficientNet leads to improved performance.

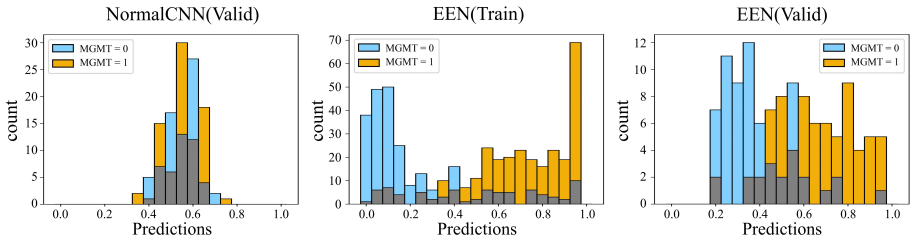

Fig. 3. Histograms of the predicted classification probabilities made by our EEN and ordinary CNN on training and validation data in predicting MGMT.

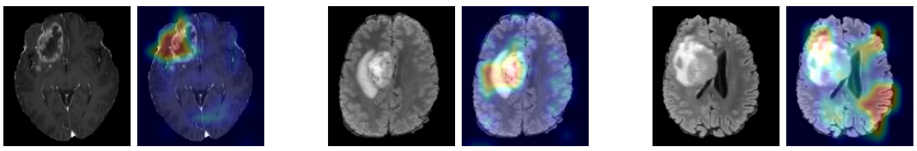

Fig. 4. Visualization of heat-maps generated by gradCAM in MGMT prediction.

On the predicted value distribution of the final output, as shown in Fig. 3, we compared the results of our method with the predicted output value of our network without EDL. Although the distribution of the predicted values after using EDL still resembles a normal distribution, it had better differentiation ability than networks that do not use EDL. In Fig. 4, we present heatmaps generated by the gradCAM algorithm to show that our model learns more information from tumor areas. Specifically, tumor areas in most cases contribute mainly to classification while other regions provide supplementary support to the classification process like the case a and b. However, in a few cases (e.g., Fig. 4(c)), the non-tumor area also contributes a lot to the final classification results, which may reveal some limitations of the model.

Table 2. Comparison and Ablation Results on OS Prediction.

Method	Accuracy	MSE
Kaggle first winner [1]	0.586	105,062
Kaggle second winner [25]	0.488	100,000
Post-hoc [13]	0.517	105,746
EfficientNetV2-S [23]	0.421	129,547
Proposed Methods	**0.513**	**79,265**
Regression (using age only)	0.387	152,619
Ours (without EDL)	0.409	136,475

3.3 Results of Predicting OS

Our model achieves the results of Table 2 on the test dataset. Compared with two state-of-the-art methods, it is seen that our method had better performance. Currently, on this dataset, most high-performance models use segmented tumor regions for analysis, such as [1, 25]. While our method does not require tumor segmentation and directly processes MRI of the whole brain still achieves good performance close to the champion [1] in terms of accuracy. Compared with other models that do not require segmentation [13, 23], our performances also

have certain advantages. It is worth noting that, thanks to the good fitting ability of EDL, our experimental results are particularly better as measured by MSE. The ablation experiments also demonstrate the effectiveness of our method.

4 Conclusions

We presented a novel End-to-End Evidential-Efficient Net for radiomics analysis that incorporates EDL layers as a classifier to predict MGMT expression and OS of brain tumor patients. We compared the proposed method to the state-of-the-art method and the results showed our model obtains better results. Our model also requires a short training time and demonstrates high stability.

References

1. Agravat, R.R., Raval, M.S.: Brain tumor segmentation and survival prediction. In: International MICCAI Brainlesion Workshop. pp. 338–348. Springer (2019)
2. Amini, A., Schwarting, W., Soleimany, A., Rus, D.: Deep evidential regression. Advances in Neural Information Processing Systems **33**, 14927–14937 (2020)
3. Baid, U., Ghodasara, S., Mohan, S., Bilello, M., Calabrese, E., Colak, E., Farahani, K., Kalpathy-Cramer, J., Kitamura, F.C., Pati, S., et al.: The RSNA-ASNR-MICCAI BraTS 2021 benchmark on brain tumor segmentation and radiogenomic classification. arXiv preprint arXiv:2107.02314 (2021)
4. Baid, U., Talbar, S., Rane, S., Gupta, S., Thakur, M.H., Moiyadi, A., Thakur, S., Mahajan, A.: Deep learning radiomics algorithm for gliomas (drag) model: a novel approach using 3d unet based deep convolutional neural network for predicting survival in gliomas. In: International MICCAI Brainlesion Workshop. pp. 369–379. Springer (2018)
5. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Freymann, J.B., Farahani, K., Davatzikos, C.: Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. Scientific data **4**(1), 1–13 (2017)
6. Bakas, S., Reyes, M., Jakab, A., Bauer, S., Rempfler, M., Crimi, A., Shinohara, R.T., Berger, C., Ha, S.M., Rozycki, M., et al.: Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the brats challenge. arXiv preprint arXiv:1811.02629 (2018)
7. Carver, E., Liu, C., Zong, W., Dai, Z., Snyder, J.M., Lee, J., Wen, N.: Automatic brain tumor segmentation and overall survival prediction using machine learning algorithms. In: International MICCAI Brainlesion Workshop. pp. 406–418. Springer (2018)
8. Chang, P., Grinband, J., Weinberg, B., Bardis, M., Khy, M., Cadena, G., Su, M.Y., Cha, S., Filippi, C., Bota, D., et al.: Deep-learning convolutional neural networks accurately classify genetic mutations in gliomas. American Journal of Neuroradiology **39**(7), 1201–1207 (2018)
9. Feng, X., Dou, Q., Tustison, N., Meyer, C.: Brain tumor segmentation with uncertainty estimation and overall survival prediction. In: International MICCAI Brainlesion Workshop. pp. 304–314. Springer (2019)

10. Feng, X., Tustison, N.J., Patel, S.H., Meyer, C.H.: Brain tumor segmentation using an ensemble of 3d u-nets and overall survival prediction using radiomic features. Frontiers in Computational Neuroscience **14**, 25 (2020)
11. Han, L., Kamdar, M.R.: MRI to MGMT: predicting methylation status in glioblastoma patients using convolutional recurrent neural networks. In: PACIFIC SYMPOSIUM ON BIOCOMPUTING 2018: Proceedings of the Pacific Symposium. pp. 331–342. World Scientific (2018)
12. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778 (2016)
13. Hermoza, R., Maicas, G., Nascimento, J.C., Carneiro, G.: Post-hoc overall survival time prediction from brain MRI. In: 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI). pp. 1476–1480. IEEE (2021)
14. Hu, L.S., Ning, S., Eschbacher, J.M., Baxter, L.C., Gaw, N., Ranjbar, S., Plasencia, J., Dueck, A.C., Peng, S., Smith, K.A., Nakaji, P., Karis, J.P., Quarles, C.C., Wu, T., Loftus, J.C., Jenkins, R.B., Sicotte, H., Kollmeyer, T.M., O'Neill, B.P., Elmquist, W., Hoxworth, J.M., Frakes, D., Sarkaria, J., Swanson, K.R., Tran, N.L., Li, J., Mitchell, J.R.: Radiogenomics to characterize regional genetic heterogeneity in glioblastoma. Neuro-Oncology **19**(1), 128–137 (08 2016). https://doi.org/10.1093/neuonc/now135
15. Kendall, A., Gal, Y.: What uncertainties do we need in bayesian deep learning for computer vision? Advances in Neural Information Processing Systems **30** (2017)
16. Korfiatis, P., Kline, T.L., Lachance, D.H., Parney, I.F., Buckner, J.C., Erickson, B.J.: Residual deep convolutional neural network predicts MGMT methylation status. Journal of Digital Imaging **30**(5), 622–628 (2017)
17. Levner, I., Drabycz, S., Roldan, G., Robles, P.D., Cairncross, J.G., Mitchell, R.: Predicting MGMT methylation status of glioblastomas from MRI texture. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 522–530. Springer (2009)
18. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al.: The multimodal brain tumor image segmentation benchmark (brats). IEEE transactions on medical imaging **34**(10), 1993–2024 (2014)
19. Saeed, N., Hardan, S., Abutalip, K., Yaqub, M.: Is it possible to predict MGMT promoter methylation from brain tumor MRI scans using deep learning models? Proceedings of Machine Learning Research **1**, 16 (2022)
20. Sensoy, M., Kaplan, L., Kandemir, M.: Evidential deep learning to quantify classification uncertainty. Advances in Neural Information Processing Systems **31** (2018)
21. Stupp, R., Mason, W.P., Van Den Bent, M.J., Weller, M., Fisher, B., Taphoorn, M.J., Belanger, K., Brandes, A.A., Marosi, C., Bogdahn, U., et al.: Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. New England Journal of Medicine **352**(10), 987–996 (2005)
22. Suter, Y., Jungo, A., Rebsamen, M., Knecht, U., Herrmann, E., Wiest, R., Reyes, M.: Deep learning versus classical regression for brain tumor patient survival prediction. In: International MICCAI Brainlesion Workshop. pp. 429–440. Springer (2018)
23. Tan, M., Le, Q.: Efficientnetv2: Smaller models and faster training. In: International Conference on Machine Learning. pp. 10096–10106. PMLR (2021)
24. Taylor, O.G., Brzozowski, J.S., Skelding, K.A.: Glioblastoma multiforme: an overview of emerging therapeutic targets. Frontiers in Oncology **9**, 963 (2019)

25. Wang, F., Jiang, R., Zheng, L., Meng, C., Biswal, B.: 3d u-net based brain tumor segmentation and survival days prediction. In: International MICCAI Brainlesion Workshop. pp. 131–141. Springer (2019)
26. Wang, S., Dai, C., Mo, Y., Angelini, E., Guo, Y., Bai, W.: Automatic brain tumour segmentation and biophysics-guided survival prediction. In: International MICCAI Brainlesion Workshop. pp. 61–72. Springer (2019)
27. Yogananda, C.G.B., Shah, B.R., Nalawade, S.S., Murugesan, G.K., Frank, F.Y., Pinho, M.C., Wagner, B.C., Mickey, B., Patel, T.R., Fei, B., et al.: MRI-based deep learning method for determining methylation status of the o6-methylguanine-DNA methyltransferase promoter outperforms tissue based methods in brain gliomas. bioRxiv (2020)
28. Zhou, T., Fu, H., Zhang, Y., Zhang, C., Lu, X., Shen, J., Shao, L.: 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. pp. 221–231. Springer (2020)