

Old Dominion University

ODU Digital Commons

Electrical & Computer Engineering Faculty
Publications

Electrical & Computer Engineering

2020

Efficacy of Radiomics and Genomics in Predicting TP53 Mutations in Diffuse Lower Grade Glioma

Zeina A. Shboul

Old Dominion University, zshbo001@odu.edu

Khan Iftekharuddin

Old Dominion University, kiftekha@odu.edu

Follow this and additional works at: https://digitalcommons.odu.edu/ece_fac_pubs



Part of the [Electrical and Computer Engineering Commons](#), [Nervous System Commons](#), and the [Neurology Commons](#)

Original Publication Citation

Shboul, Z., & Iftekharuddin, K. (2020). Efficacy of radiomics and genomics in predicting TP53 mutations in diffuse lower grade glioma. In Krol, A. and Gimi, B.S. (Eds.), *Medical Imaging 2020: Biomedical Applications in Molecular, Structural and Functional Imaging*, Houston, Texas, U.S.A. 113170Y (1-7). SPIE. <https://doi.org/10.1117/12.2550764>

This Conference Paper is brought to you for free and open access by the Electrical & Computer Engineering at ODU Digital Commons. It has been accepted for inclusion in Electrical & Computer Engineering Faculty Publications by an authorized administrator of ODU Digital Commons. For more information, please contact digitalcommons@odu.edu.

PROCEEDINGS OF SPIE

SPIDigitalLibrary.org/conference-proceedings-of-spie

Efficacy of radiomics and genomics in predicting TP53 mutations in diffuse lower grade glioma

Shboul, Zeina, Iftkharuddin, Khan

Zeina A. Shboul, Khan M. Iftkharuddin, "Efficacy of radiomics and genomics in predicting TP53 mutations in diffuse lower grade glioma," Proc. SPIE 11317, Medical Imaging 2020: Biomedical Applications in Molecular, Structural, and Functional Imaging, 113170Y (28 February 2020); doi: 10.1117/12.2550764

SPIE.

Event: SPIE Medical Imaging, 2020, Houston, Texas, United States

Efficacy of Radiomics and Genomics in Predicting *TP53* Mutations in Diffuse Lower Grade Glioma

Zeina A. Shboul and Khan M. Iftekharuddin

The authors are with the Vision Lab in the Department of Electrical and Computer Engineering, Old Dominion University, Norfolk, VA 23529

ABSTRACT

An updated classification of diffuse lower-grade gliomas is established in the 2016 World Health Organization Classification of Tumors of the Central Nervous System based on their molecular mutations such as *TP53* mutation. This study investigates machine learning methods for *TP53* mutation status prediction and classification using radiomics and genomics features, respectively. Radiomics features represent patients' age and imaging features that are extracted from conventional MRI. Genomics feature is represented by patients' gene expression using RNA sequencing. This study uses a total of 105 LGG patients, where the patient dataset is divided into a training set (80 patients) and testing set (25 patients). Three *TP53* mutation prediction models are constructed based on the source of the training features; *TP53*-radiomics model, *TP53*-genomics model, and *TP53*-radiogenomics model, respectively. Radiomics feature selection is performed using recursive feature selection method. For genomics data, EdgeR method is utilized to select the differentially expressed genes between the mutated *TP53* versus the non-mutated *TP53* cases in the training set. The training classification model is constructed using Random Forest and cross-validated using repeated 10-fold cross validation. Finally, the predictive performance of the three models is assessed using the testing set. The three models, *TP53*-Radiomics, *TP53*-RadioGenomics, and *TP53*-Genomics, achieve a predictive accuracy of 0.84 ± 0.04 , 0.92 ± 0.04 , and 0.89 ± 0.07 , respectively. These results show promise of non-invasive MRI radiomics features and fusion of radiomics with genomics features for prediction of *TP53*.

Keywords: Glioma grading, *TP53* mutation prediction and classification, Radiomics, Genomics, Random Forest.

1. INTRODUCTION

Diffuse or infiltrative lower-grade gliomas (LGG) are a Central Nervous System (CNS) brain tumor that include the World Health Organization (WHO) Grade II and III gliomas [1, 2]. Infiltrative LGG arise from the glial cells in the nervous tissues of the CNS [3, 4]. The histology classes of diffuse LGG include astrocytoma, oligodendroglioma, and oligoastrocytoma. However, an updated classification of diffuse LGG is established based on their genetic mutations (such as isocitrate dehydrogenase (IDH) mutation and tumor protein 53 (TP53) mutation) and their histological type [1]. TP53 mutation is a molecular mutation that usually occurs in a younger patient with astrocytoma [5, 6]. TP53 is a well-known tumor suppressor gene that encodes p53 protein and involves in tumor survival, proliferation, and invasion [7]. A study by Ohgaki et. al, [8] with 715 diagnosed cases of glioblastoma (WHO Grade IV glioma) reveals that mutated TP53 is predictive of longer survival.

Identifying the presence of certain mutations such as TP53 requires molecular profiling using invasive methods by obtaining tumor tissue sample with an increase in the proliferation and neovascularization [9]. Then, Immunohistochemistry (IHC) staining is used to detect missense mutations. Recently, several studies in the literature have proposed the use of RNA sequencing to determine TP53 functional status [10]. However, such invasive methods may also be accompanied by high cost, morbidity, and different difficulties [11]. Therefore, in this work, we investigate different machine learning methods for classification and prediction of TP53 mutation in diffuse LGG based on radiomics, genomics, and fused radiogenomics and compare the performance of these prediction models.

Radiomics are quantitative features that are extracted from volumes of interest in radiology imaging. These quantitative features offer valuable information regarding intensity, shape, size, volume, and texture, and can be associated with glioma clinical outcomes and survival prediction [12-17]. For example Gutman et al. [18], associate three volumetric imaging features with a genetic mutation, survival prediction and Verhaak subtypes [19]. In another study by Shboul et al. [17], the authors investigate the efficacy of using non-invasive quantitative texture features extracted from structural MRI to predict the progression of lower-grade glioma. Compared with the radiomic features, RNA counts data (i.e. genomics) describe the gene expression by counting the relative frequency of the RNA reads

Medical Imaging 2020: Biomedical Applications in Molecular, Structural, and Functional Imaging, edited by Andrzej Krol, Barjor S. Gimi, Proc. of SPIE Vol. 11317, 113170Y · © 2020 SPIE · CCC code: 1605-7422/20/\$21 · doi: 10.1117/12.2550764

mapping to one gene [20]. Because RNA is count data, in this study, we log-transformed the differentially expressed genes to normality to use them in machine learning models [21, 22].

The proposed work utilizes radiomics and genomics in machine learning modeling to predict *TP53* mutations. The proposed work ascertain the efficacy of non-invasive MRI radiomics to predict *TP53* mutation status when compared to genomics data. The dominant approach in identifying *TP53* mutations besides RNA sequencing in literature requires viable tumor tissue that involves invasive tissue sampling. Our results show the efficacy of using either radiomics or radiomics fused with the genomics in predicting *TP53* mutation status when compared to genomics model alone.

2. METHOD

In this study, we investigate *TP53* mutation prediction based on a) radiomics (patient age at the diagnosis and imaging features extracted from Magnetic Resonance Imaging (MRI)), b) genomics (gene expression using RNA sequencing), c) and fused radiogenomics (illustrated in Figure 1). First, optimum radiomics are selected using recursive feature selection (RFS), and the top five differentially expressed genes are selected using EdgeR method [23]. Then, the optimum radiomics and the top differentially expressed genes are utilized along using a Random Forest (RF) classifier to construct and train the three different *TP53* mutations models: *TP53*-Radiomics, *TP53*-Genomics, and *TP53*-RadioGenomics model, respectively. The *TP53* mutation models are evaluated using repeated 10-fold Cross-Validation (CV). Finally, we assess the predictive performances of these models in predicting *TP53* mutation status using the testing dataset.

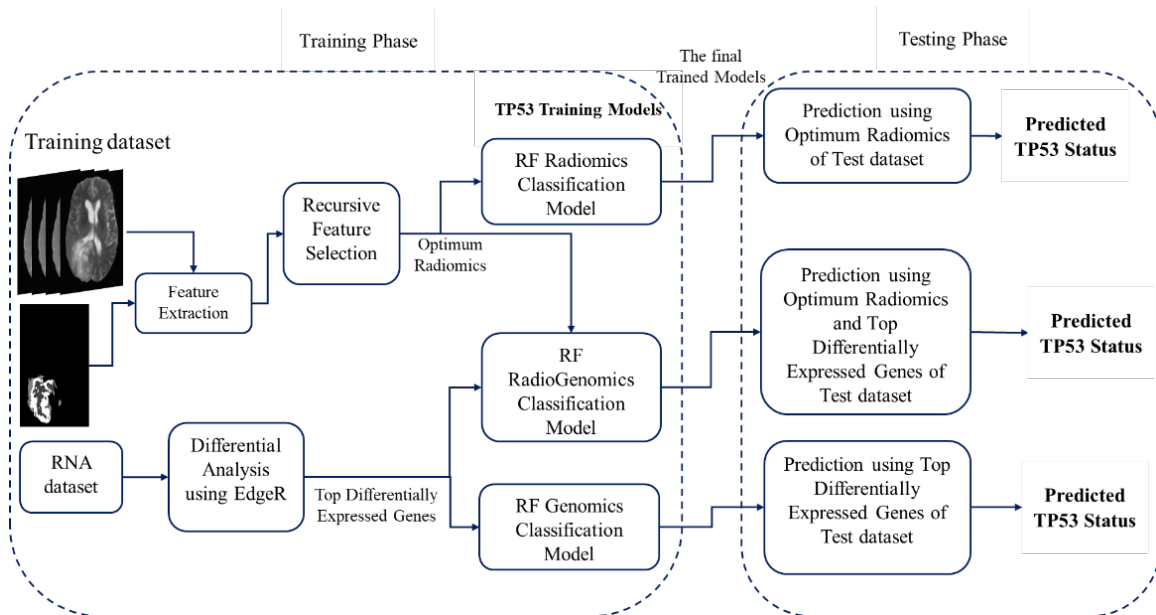


Figure 1. Flow Diagram of the proposed LGG *TP53* prediction model. The training and testing analysis are repeated 10 times independently.

2.1 Dataset

A total of 105 pre-operative LGG patients described in [24-26] is used to evaluate and assess the proposed models in Figure 1. Four modalities of the MRI are provided with the dataset (T1, T1Gd, T2, FLAIR) along with their segmented tumor. Clinical data and RNA counts data are downloaded from the Genomic Data Commons Data Portal (<https://portal.gdc.cancer.gov/>). Clinical data are de-identified following the HIPAA standard. The range of the patients' age at the diagnosis is 20 – 75 years and the median and the mean age are 46 years. Fifty-five cases harbor *TP53* mutations while fifty patients are without *TP53* mutations (i.e. wild-type).

2.2 Extraction of radiomics

Radiomics are quantitative imaging texture features that are extracted from raw MRI (T1, T1Gd, T2, and FLAIR) sequences of the tumor, five Texton features [27], and multifractal features [28, 29], and the age of the patient at the

diagnosis. Radiomics features include the histogram, the co-occurrence matrix, the neighborhood gray-tone difference matrix, and the Size Zone Matrix [17, 30].

Also, we utilize volumetric features that are computed as follows: the volume of the whole tumor, the volume of the whole tumor divided by the volume of the brain, the volume of sub-regions (edema, enhance tumor, and necrosis) divided by the volume of the whole tumor, and divided by the volume of the brain, the volumes of the enhance tumor and necrosis divided by the volume of the edema, the combined volume of the edema and enhance tumor, the volume of the edema divided by the combined volume of enhancing tumor and necrosis, and the volume of the necrosis from the combined volume of the edema and enhance tumor.

Additionally, another nine area property features are extracted from the tumor from x , y , and z -axes. The properties are area, centroid, perimeter, major axis length, minor axis length, eccentricity, orientation, solidity, and extent.

2.3 Machine learning models for *TP53* prediction in LGG

Three *TP53* mutation prediction models are constructed based on the source of the training features; *TP53*-radiomics model, *TP53*-genomics model, and *TP53*-radiogenomics model, respectively. Figure 1 illustrates the complete flow diagram of the three *TP53* mutation ML prediction models.

In this study, the dataset is divided into training and testing datasets. Eighty patients are used in the training dataset and the remaining (twenty-five patients) are used in the testing dataset. Feature selection and Random Forest [31] are utilized to construct and train the three different *TP53* mutation prediction models. Repeated 10-fold cross-validation is utilized to cross-validate and tune the RF parameters. Radiomics feature selection is performed using recursive feature selection (RFS). RFS is constructed based on RF and is performed by maximizing the ROC metric. In RFS, radiomics are trained using RF, ranked bases on their importance, and then the least important radiomics are removed. The step of training RF using the remaining radiomics and ranking them is repeated recursively until the radiomics' set that maximizes the ROC is reached.

Genomics represented by RNA counts data is the other source of features in this work. EdgeR [23] is exploited to select the top differentially expressed genes (i.e. most important). In EdgeR, the negative binomial distribution is used to model RNA counts. EdgeR assumes that the variance of RNA counts depends on two dispersion parameters; common and individual. Once the negative binomial is fitted and dispersion parameters are obtained, differential expression analysis is performed using either the likelihood ratio test or F-test to find the top differentially expressed genes. Then the top differentially expressed genes are log-transformed into normality and fed into RF to build *TP53*-Genomics and *TP53*-RadioGenomics.

Moreover, the prediction performance is tested using the testing datasets. The prediction performance is assessed using the accuracy (Acc), balanced accuracy (B. Acc), sensitivity (Sens.), specificity (Spec.), negative predictive value (NPV), and positive predictive value (PPV).

Finally, a comparison of the performance between the three *TP53* mutation models is conducted using Analysis of Variance (ANOVA) method. If the p - value of the ANOVA test is less than 0.05, a statistically significant difference is accepted. Association between *TP53* mutation status and the optimum radiomics and the top differentially expressed genes is assessed using ANOVA.

3. RESULTS

The RFS method offers the most frequent optimum radiomics features as follows: the age, information correlation, and the low-gray level zone emphasis. EdgeR offers the top differentially expressed genes, and the gene that is the most occurred (in the ten training splits) is *TEC* gene. Figure 2 illustrates the distribution of the most frequent optimum features and the distribution of the *TEC* gene expression. Younger ages, low values of low-gray level zone emphasis, and high values of information correlation are significantly associated with mutated *TP53* with p - value < 0.05. Low values of *TEC* RNA expression are significantly correlated with mutated *TP53* with p - value < 0.05.

Moreover, our analysis shows that the patients' age (at the diagnosis) carries a Hazard Ratio [HR] of 1.07 (95% CI, 1.03-1.11, likelihood ratio test p -value < 0.05). Besides, patients age greater than the median age of 46 years carries a HR of 3.15 (95% CI, 1.39-7.16, likelihood ratio test p -value < 0.05) and a median survival of 62 months vs. 94.5 months for patients with age less than 46 years as illustrated in Figure 3a. Unlike patients age at the diagnosis, the optimum radiomics of information correlation and the low gray level zone emphasis and the *TEC* gene do not show an association with overall survival. Additionally, our analysis reveals that *TP53* mutation status does not associate

with the overall survival (likelihood ratio test p -value = 0.1) as illustrated in Figure 3b. However, as shown in Table 1, a chi-square test reveals that mutated *TP53* and patients with ages less than 46 are significantly associated with p -value < 0.05, whereas WT *TP53* status is significantly associated with patients with ages greater than 46.

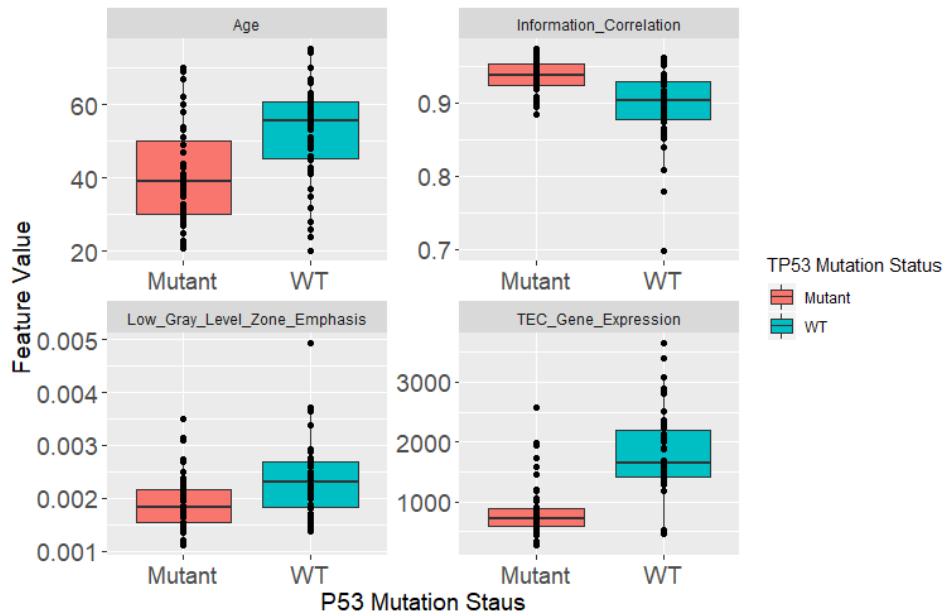


Figure 2. Feature distribution of the optimum radiomics and the top differentially expressed gene in the 10 training sets in discriminating mutated *TP53* (Mutant) and not mutated *TP53* (i.e. Wild Type (WT)).

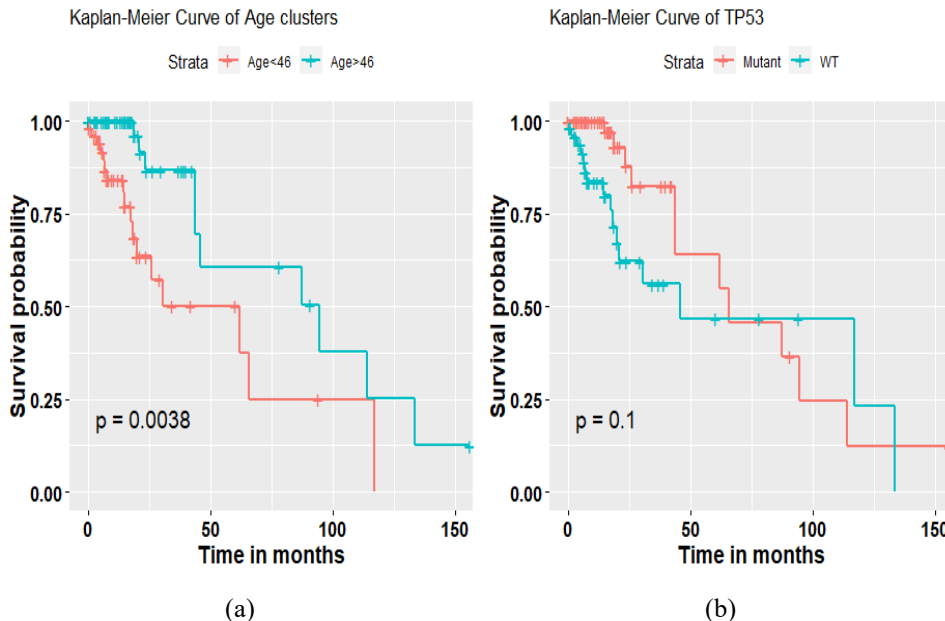


Figure 3. Kaplan Meier curve of a) patients' age clustered into two groups (age less than 46 years and age greater than 46 years), and b) *TP53* mutation status.

Table 1. TP53 mutation distribution based on the patients' age in the dataset.

Age groups	TP53 mutation Status	
	Mutant	WT
Patients' age less than 46	39	14
Patients' age greater than 46	16	36

Figure 4 demonstrates the performance of the three TP53 mutation prediction models using the 10 testing sets. These values represent the mean \pm standard deviation of the performance of the 10 repetitions using the testing sets. As illustrated in Figure 4, TP53-Radiomics model achieves an accuracy of 0.84 ± 0.04 , a sensitivity of 0.87 ± 0.08 , a specificity of 0.82 ± 0.09 , and balanced accuracy of 0.84 ± 0.04 . The performance of TP53-RadioGenomics model (with an accuracy of 0.92 ± 0.04 , a sensitivity of 0.92 ± 0.05 , a specificity of 0.93 ± 0.08 , and balanced accuracy of 0.92 ± 0.05) outperforms the performance of the TP53-Radiomics models significantly (ANOVA test, p - value < 0.05) as shown in Table 2 and Figure 4. Note the performance of the TP53-RadioGenomics model is better than the performance of the TP53-Genomics models (with an accuracy of 0.89 ± 0.07 , a sensitivity of 0.89 ± 0.07 , a specificity of 0.89 ± 0.13 , and balanced accuracy of 0.89 ± 0.07), however, the improvements of the TP53-RadioGenomics is not significant (ANOVA test, p - value > 0.05) as shown in Table 2 and Figure 4.

Table 2. The ANOVA Probability value (p -value) of performance of the difference between TP53 mutation prediction models. Note, p -value < 0.05 indicates statistically significant difference.

TP53 models comparison	Accuracy	Balanced Accuracy	Negative Predictive Value	Positive Predictive Value	Sensitivity	Specificity
Radiomics VS. RadioGenomics	0.000	0.000	0.051	0.004	0.094	0.010
Radiomics VS. Genomics	0.085	0.083	0.418	0.083	0.517	0.147
RadioGenomic VS. Genomics	0.251	0.260	0.290	0.503	0.295	0.504

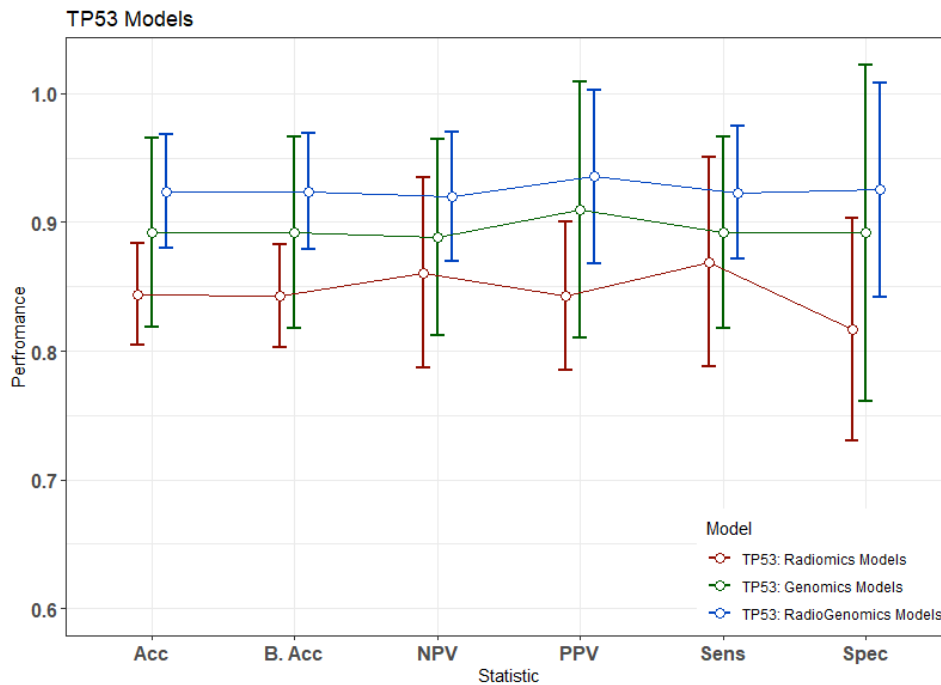


Figure 4. The performance of the TP53 mutation prediction models using 10 different testing splits. Error bar represents two standard deviations. "Acc" refers to accuracy, "B. Acc" refers to balanced accuracy, "NPV" refers to the negative predictive value, "PPV" refers to the positive predictive value, "Sens" refers to sensitivity, and "Spec" refers to specificity.

4. CONCLUSION

This study proposes multiple ML models for *TP53* mutation status prediction based on non-invasive radiomics when compared to invasive genomics. Radiomics represent the patients' age and extracted imaging features from the conventional raw and different characterization of MRI modalities (T1, T1Gd, T2, and Flair). Genomics is represented by RNA counts data in this study. Radiomics features selection is performed using RFS whereas the top differentially expressed genes are selected using EdgeR. A total of 105 LGG patients are randomly divided into 10 paired training (80 patients) and testing sets (25 patients). Based on the source of the training features, we construct three *TP53* mutation prediction models; the *TP53*-radiomics model, the *TP53*-genomics model, and *TP53*-radiogenomics model, respectively. The training classification of the three models is developed using the Random Forest. Ten-fold cross-validation is utilized to cross-validate and tune the RF parameters. Finally, the predictive performance of the three models is assessed using the testing set.

Our analysis reveals that age, information correlation, and the low-gray level zone emphasis show significance in predicting *TP53* mutation status in the *TP53*-radiomics model. *TEC* gene is the most important gene that is used to predict *TP53* mutation status in the *TP53*-Genomics model. Moreover, our *TP53*-Radiomics model achieves an accuracy of 0.84 ± 0.04 , a sensitivity of 0.87 ± 0.08 , a specificity of 0.82 ± 0.09 , and balanced accuracy of 0.84 ± 0.04 . In comparison, in a recent study by Li et. al, [32], the authors propose a *TP53* mutation prediction model based on radiomics features. The authors use 180 LGG patients in the training set and 92 patients in the test set. Their proposed method utilize LASSO as a feature selection and SVM. Their proposed method has achieved an AUC of 76.3%, a sensitivity 62.2%, a specificity 85.1%, and an accuracy of 70.7%, in the validation set data. When comparing the performance of our *TP53*-Radiomics model and the performance of Li et. al, [32], our *TP53*-Radiomics model achieves better accuracy and sensitivity.

As expected, the *TP53*-Gadiomics model achieves better performance than the *TP53*-Radiomics model. However, the *TP53*-Radiomics model offers comparable performance and the *TP53*-RadioGenomics model outperforms the *TP53*-Radiomics model significantly. These results suggest efficacy of using conventional MRI as predictive biomarker for *TP53* mutation prediction.

ACKNOWLEDGMENTS

This work was partially funded by NIBIB/NIH grant# R01 EB020683.

REFERENCES

- [1] Louis, D. N., Perry, A., Reifenberger, G. et al., "The 2016 World Health Organization classification of tumors of the central nervous system: a summary," *Acta neuropathologica*, 131(6), 803-820 (2016).
- [2] Louis, D. N., Ohgaki, H., Wiestler, O. D. et al., "The 2007 WHO classification of tumours of the central nervous system," *Acta neuropathologica*, 114(2), 97-109 (2007).
- [3] Brat, D. J., Verhaak, R. G., Aldape, K. D. et al., "Comprehensive, integrative genomic analysis of diffuse lower-grade gliomas," *New England Journal of Medicine*, 372(26), 2481-2498 (2015).
- [4] Zong, H., Verhaak, R. G., and Canoll, P., "The cellular origin for malignant glioma and prospects for clinical advancements," *Expert review of molecular diagnostics*, 12(4), 383-394 (2012).
- [5] Sarkar, C., Karak, A. K., Nath, N. et al., "Apoptosis and proliferation: correlation with p53 in astrocytic tumours," *Journal of neuro-oncology*, 73(2), 93-100 (2005).
- [6] Watanabe, K., Tachibana, O., Sato, K. et al., "Overexpression of the EGF receptor and p53 mutations are mutually exclusive in the evolution of primary and secondary glioblastomas," *Brain pathology*, 6(3), 217-223 (1996).
- [7] Muller, P. A. J., and Vousden, K. H., "p53 mutations in cancer," *Nature Cell Biology*, 15, 2 (2013).
- [8] Ohgaki, H., Dessen, P., Jourde, B. et al., "Genetic pathways to glioblastoma: a population-based study," *Cancer research*, 64(19), 6892-6899 (2004).
- [9] Jalbert, L. E., Elkhalel, A., Phillips, J. J. et al., "Metabolic profiling of IDH mutation and malignant progression in infiltrating glioma," *Scientific reports*, 7, 44792 (2017).
- [10] Troester, M. A., Herschkowitz, J. I., Oh, D. S. et al., "Gene expression patterns associated with p53 status in breast cancer," *BMC cancer*, 6(1), 276 (2006).

- [11] Hakkarainen, T. W., Kopari, N. M., Pham, T. N. *et al.*, "Necrotizing soft tissue infections: review and current concepts in treatment, systems of care, and outcomes," *Current problems in surgery*, 51(8), 344 (2014).
- [12] Tsien, C., Galbán, C. J., Chenevert, T. L. *et al.*, "Parametric response map as an imaging biomarker to distinguish progression from pseudoprogression in high-grade glioma," *Journal of Clinical Oncology*, 28(13), 2293 (2010).
- [13] Hamstra, D. A., Galbán, C. J., Meyer, C. R. *et al.*, "Functional diffusion map as an early imaging biomarker for high-grade glioma: correlation with conventional radiologic response and overall survival," *Journal of clinical oncology*, 26(20), 3387 (2008).
- [14] Hamstra, D. A., Chenevert, T. L., Moffat, B. A. *et al.*, "Evaluation of the functional diffusion map as an early biomarker of time-to-progression and overall survival in high-grade glioma," *Proceedings of the National Academy of Sciences*, 102(46), 16759-16764 (2005).
- [15] Moffat, B. A., Chenevert, T. L., Lawrence, T. S. *et al.*, "Functional diffusion map: a noninvasive MRI biomarker for early stratification of clinical brain tumor response," *Proceedings of the national academy of sciences*, 102(15), 5524-5529 (2005).
- [16] Shboul, Z. A., Vidyaratne, L., Alam, M. *et al.*, "Glioblastoma and Survival Prediction." *International MICCAI Brainlesion Workshop* 358-368.
- [17] Shboul, Z. A., and Iftekharuddin, K. M., "Prediction of low-grade glioma progression using MR imaging." *Proc. SPIE* 10950, 1095049.
- [18] Gutman, D. A., Cooper, L. A., Hwang, S. N. *et al.*, "MR imaging predictors of molecular profile and survival: multi-institutional study of the TCGA glioblastoma data set," *Radiology*, 267(2), 560-569 (2013).
- [19] Verhaak, R. G., Hoadley, K. A., Purdom, E. *et al.*, "Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1," *Cancer cell*, 17(1), 98-110 (2010).
- [20] Kukurba, K. R., and Montgomery, S. B., "RNA sequencing and analysis," *Cold Spring Harbor Protocols*, 2015(11), pdb.top084970 (2015).
- [21] McDonald, J. H., [Handbook of biological statistics] sparky house publishing Baltimore, MD, (2009).
- [22] Law, C. W., Chen, Y., Shi, W. *et al.*, "voom: Precision weights unlock linear model analysis tools for RNA-seq read counts," *Genome biology*, 15(2), R29 (2014).
- [23] Robinson, M. D., McCarthy, D. J., and Smyth, G. K., "edgeR: a Bioconductor package for differential expression analysis of digital gene expression data," *Bioinformatics*, 26(1), 139-140 (2010).
- [24] Menze, B. H., Jakab, A., Bauer, S. *et al.*, "The multimodal brain tumor image segmentation benchmark (BRATS)," *IEEE transactions on medical imaging*, 34(10), 1993-2024 (2015).
- [25] Bakas, S., Akbari, H., Sotiras, A. *et al.*, "Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features," *Scientific data*, 4, 170117 (2017).
- [26] Bakas, S., Akbari, H., Sotiras, A. *et al.*, "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection," *The Cancer Imaging Archive*, (2017).
- [27] Leung, T., and Malik, J., "Representing and recognizing the visual appearance of materials using three-dimensional textons," *International journal of computer vision*, 43(1), 29-44 (2001).
- [28] Islam, A., Iftekharuddin, K. M., Ogg, R. J. *et al.*, "Multifractal modeling, segmentation, prediction, and statistical validation of posterior fossa tumors." *Proc. SPIE* 6915, 69153C.
- [29] Reza, S., and Iftekharuddin, K., "Multi-fractal texture features for brain tumor and edema segmentation." *Proc. SPIE* 9035, 903503.
- [30] Vallières, M., Freeman, C. R., Skamene, S. R. *et al.*, "A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities," *Physics in medicine and biology*, 60(14), 5471 (2015).
- [31] Breiman, L., [Random forests], (2001).
- [32] Li, Y., Qian, Z., Xu, K. *et al.*, "MRI features predict p53 status in lower-grade gliomas via a machine-learning approach," *NeuroImage: Clinical*, 17, 306-311 (2018).