# Challenges in predicting glioma survival time in multi-modal deep networks

1st Abdulrhman Aljouie
*Department of Biostatistics and Bioinfornatics*
*King Abdullah International Medical Research Center*
Riyadh, Saudi Arabia
aljouieab@ngha.med.sa

2nd Yunzhe Xue
*Department of Computer Science*
*New Jersey Institute of Technology*
Newark, USA
yx277@njit.edu

3rd Meiyan Xie
*Department of Computer Science*
*New Jersey Institute of Technology*
Newark, USA
mx42@njit.edu

4th Usman Roshan
*Department of Computer Science*
*New Jersey Institute of Technology*
Newark, USA
usman@njit.edu

***Abstract*—Prediction of cancer survival time is of considerable interest in medicine as it leads to better patient care and reduces health care costs. In this study, we propose a multi-path multi-modal neural network that predicts Glioblastoma Multiforme (GBM) survival time at the 14 months threshold. We obtained image, gene expression, and SNP variants from whole-exome sequences all from the The Cancer Genome Atlas portal for a total of 126 patients. We perform a 10-fold cross-validation experiment on each of the data sources separately as well as the model with all data combined. From post-contrast T1 MRI data, we used 3D scans and 2D slices that we selected manually to show the tumor region. We find that the model with 2D MRI slices and genomic data combined gives the highest accuracies over individual sources but by a modest margin. We see considerable variation in accuracies across the 10 folds and that our model achieves 100% accuracy on the training data but lags behind in test accuracy. With dropout our training accuracy falls considerably. This shows that predicting glioma survival time is a challenging task but it is unclear if this is also a symptom of insufficient data. A clear direction here is to augment our data that we plan to explore with generative models. Overall we present a novel multi-modal network that incorporates SNP, gene expression, and MRI image data for glioma survival time prediction.**

***Index Terms*—TCGA GBM, MRI, SNPs, CNN, mRNA expression.**

## I. Introduction

Predicting glioma survival time helps patients and their clinicians evaluate available treatment plans and make informed choices. Glioblastoma Multiforme (GBM) is the most common and lethal glioma type in adult [19]. In GBM, less than 5% of patients reach 5 years survival threshold after diagnosis with a median survival of 15 months [18]. Most advanced cancer patients prefer to know their estimated prognostic information [10]. However, clinicians' survival time estimates are inaccurate, and often optimistic [10], [17].

Many studies have devised 3D convolutional neural networks (CNNs) to improve the accuracy of structural MRI scans to classify glioma patients into survival categories [3], [5], [15], [22]. In this paper we look into a heterogeneous combination of somatic and germline genetic single variations, messenger RNA expressions, and post-contrast T1 MRI modality data that show the malignancy. We obtained whole exome sequencing data (WES) from the The Cancer Genome Atlas (TCGA) portal (https://www.cancer.gov/tcga), messenger RNA, and post-contrast axial T1 MRI sequences from The Cancer Imaging Archive (TCIA [6]) for all European ancestry individuals with GBM. For this study we only include samples for which all three data types are available, giving us a total of126. For each sample we assign a label of 0 if their survival time is below14 months and 1 otherwise (to obtain balanced sets), thus converting our survival time prediction problem into a classification one.

We design a multi-path neural network that takes as input all three data sources and evaluate its accuracy in a 10-fold cross-validation experiment. We ran the Genome Analysis Toolkit (GATK4) pipeline to obtain single mutations with exhaustive site-level and sample-level quality controls to eliminate sequencers artifacts and false-positive SNPs. We included biallelic and multiallelic loci and converted the two allele copies of each SNP into a numerical format using an in-house python script. Then, we ranked SNPs on each training split to select the best 100 SNPs to use as predictive markers. We obtained TCGA-GBM samples from mRNA expression information after Robust Multi-array Analysis (RMA) normalization from the Broad Institute TCGA Genome Data Analysis Center Firehose.

We downloaded MRI sequences in Digital Imaging and Communications in Medicine (DICOM) format from TCIA. From the 3D axial T1 MRI sequences. We explored both 3D volumes and 2D slices. For 3D scans, we converted the DICOM images to Neuroimaging Informatics Technology Initiative (NIfTI) format. We extracted non-brain tissue with FSL BET, and registered the images to T1 MRI MNI152 reference with FSL FLIRT. We train a model with 3D U-Net [14] separately as well as simultaneously with SNPs and mRNA data.

For 2D slices, we manually selected one slice that shows the tumor for each sample. Then we used these 2-D image slices to train a 2D CNN with ResNet18 [11] encoder and measured the accuracy of predicting test samples in 10-fold cross-validation. We compared the accuracy of predicting survival time with SNPs, mRNA expressions, and MRI scans separately as well as when combining the three data sources. For SNPs and mRNA expressions we used separate multi-layer neural nets and for images we used 2D and 3D convolutional neural networks.

We see a slight improvement with in combined model with 2D images over the individual data sources but considerable variation in test accuracy across different train test folds. We conjecture this may be due to our small training set of 126 individuals. By synthetically augmenting the data with a generative model we may improve sample representation and consequently model accuracy.

## II. Methods

### A. Data

Our data is composed of TCGA-GBM (https://www.cancer.gov/tcga) European ancestry individuals that have all of the following data: 1) Survival time (days from diagnosis to death), we also performed right censoring to increase the dataset size where we included samples for which days to the last follow-up are above the 14 months threshold, 2) WES data, 3) mRNA expressions information, and 4) post-contrast T1 axial MRI sequence. We excluded samples that don't meet the inclusion criteria. The total number of samples included in the analysis is 126, Table I shows the clinical characteristic of these samples.

TABLE I
SAMPLES CLINICAL CHARACTERISTIC

| Clinical characteristic | TCGA-GBM (n=126) |
|---|---|
| Ancestry (European) | 126 |
| Ethnicity (not reported/not hispanic or latino) | 25/101 |
| Gender (male/female) | 76/51 |
| Average age | 60.38 $\pm$13.37 |
| Vital status (dead/alive) | 123/3 |
| Average survival (days) excluding censored | 483.44 $\pm$ 431.95 |
| # of samples in each class (short-term/long-term) | 63/63 |

*1) SNPs:* We obtained TCGA-GBM 126 European ancestry individuals pre-aligned WES for each sample that met the inclusion criteria from TCGA (https://www.cancer.gov/tcga) through the NCI Genomic Data Commons (GDC) data portal (https://gdc.cancer.gov/). We ran a GATK (version 4) HaplotypeCaller [12], [8], [2] on each sample. We then pooled all samples together for joint genotyping utilizing a computing cluster in a scatter and gather approach on each chromosome to expedite variant discovery process. To filter out low-quality SNPs, we applied the GATK variant quality recalibration score (VQSR), which uses a machine learning trained on external datasets to assign a quality score to each site-level variant. We used a truth sensitivity of a 99% as a threshold. Those SNPs that passed VQSR are further interrogated on sample-level genotype quality (GQ) and depth (DP). SNPs that passed VQSR at the site-level and GQ $>$ 20 and DP $\geq$ 5 at the sample-level are included for further analysis. We performed the typical widely used multi-allelic encoding of SNPs shown in Figure 1.

We calculated the chi-squared statistic [7] between each SNP and the binary class label and ranked SNPs based on the test statistics. The higher the statistic the more important the SNP in its predictive ability. We included top-ranked 100 SNPs for further analysis.

*2) mRNA expressions:* We downloaded gene expression information for the samples that were normalized with Robust Multi-array Analysis (RMA) from the Broad Institute TCGA Genome Data Analysis Center Firehose [1].

*3) 3D MRI scans:* We obtained axial T1 MRI sequences in DICOM format from The Cancer Imaging Archive (TCIA). We converted DICOM images to NIfTI format with dcmtonii software and removed non-brain tissue with FSL BET [21] with option -B (an option that leads to overall better performance in skull-stripping [20]). We then aligned images to T1 axial MNI152 reference with FSL FLIRT.

*4) 2D MRI slices:* For each subject, we manually selected an image slice that best shows the tumor and its surrounding tumor enhancing-area. Table II shows the vector and matrix dimensions for the three data sources we used in our analysis.

TABLE II
SNPs, mRNA, AND T1 MRI DATA

| Data set | Vector (matrix) dimension |
|---|---|
| SNPs (passed filtering) | 79980 |
| mRNA expressions | 12042 |
| 3D post-contrast T1 MRI scans | (182, 218, 182) |
| 2D post-contrast T1 MRI slices | (256,256) |

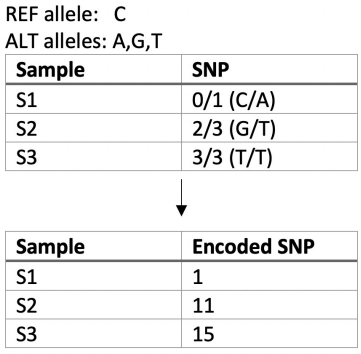


Fig. 1. A toy example for encoding a multialleic SNP into a numerical format

### B. Network architecture and training

We constructed three separate neural nets using PyTorch [13] to evaluate the predictive power of each data source alone. We concatenated the output of these nets to evaluate combining SNPs, mRNA expressions, and images predictive ability.

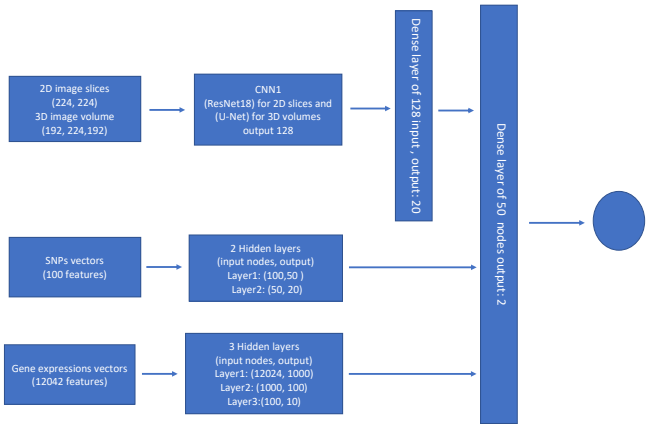


Fig. 2. Our multi-modal deep neural network. We see three paths each for SNP, gene expression, and images. We train the network as one model instead of training the three paths separately.

For SNPs we constructed a net with 2 hidden layers. For activation function, we used Relu, and 0.01 learning rate, batch size of 5 and 30 epochs with Stochastic Gradient Descent (SGD) [4] and Nesterov Momentum update. For mRNA expressions, we set the parameters to exactly what we used for SNPs but with 3 hidden layers (1000, 100, 10). For 2D T1 MRI sequence slices, we used ResNet18 convolutional neural network [11] which has 18 hidden layers and has 18 output nodes. Because the ResNet18 input size shape is (244,244), we resized all slice images to (256,256) dimensions and randomly center cropped (224, 224), we use the cropped images as an input for the ResNet18 convolutional neural network. We used the following parameters with ResNet18: learning rate of 0.01, batch size of 6, 15 epochs. For 3D volumes, we employed 3D U-Net [14] where we padded the original images with zero to fit the network input dimensions of (192, 224, 192). We used the same parameters here that we used to train the 2D slices in ResNet18.

To combined each of SNPs and mRNA dataset with MRI slices, we add one more dense layer to the end of ResNet18. After ReLu activation we concatenate the output to the network's output. Then we feed it into a dense layer with 50 input nodes and 2 output nodes. For combining the three data sources, we concatenate the three outputs of each network. Figure 2 shows our network architecture for combined data sources.

## III. Results

We report on our 10-fold cross-validation results on all three data sources combined as well as individual data sources with both 2D and 3D images. We evaluate accuracy as the sum of correct predictions over the total number of the test set. We selected our survival threshold at 14 months intentionally so that our data is balanced: we have equal number of samples on both classes.

### A. Combined data with 3D volumes

In Figure 3 we show the mean accuracy of our model across 10-folds and 15 epochs for each of the three data sources separately and the combined data model. We see that our model can achieve a 100% accuracy on the individual and combined data models. In the test, however, the accuracies are lower. We see that the combined data model does not perform better than the individual ones. In fact here the gene expression data source gives the best test accuracy of 62.4% at epoch 13.

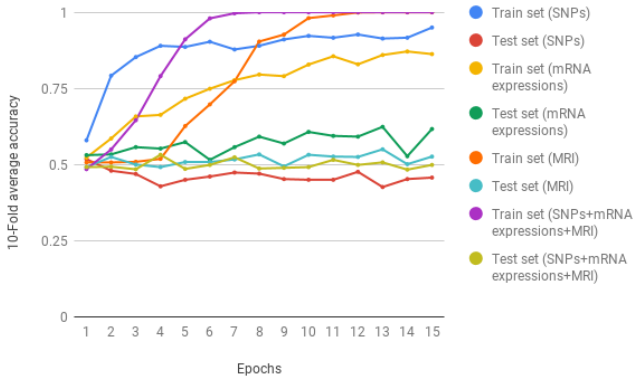


Fig. 3. Mean 10-fold accuracy of our network across 15 epochs for training and test sets with 3D volumes as the image data.

In In Figure 6 we see the test accuracy of each of the ten folds of the combined model. We see in some folds the test accuracy goes to 75% whereas in other as low as 25%. This suggests that in some in some folds we have a diverse enough training set that captures the distribution of test datapoints, whereas in other folds the training and test image datasets are very different.

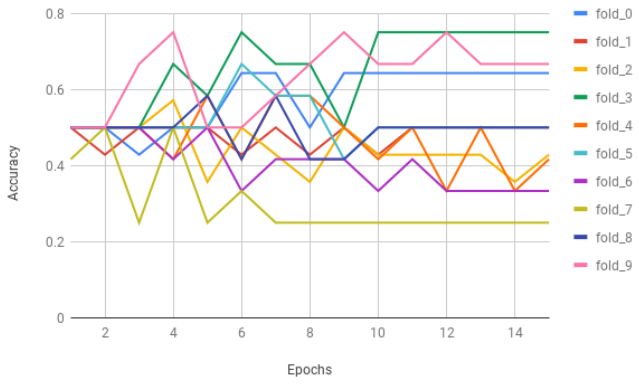


Fig. 4. Test accuracy of each of the 10-folds of our network across 15 epochs on all three data sources combined with 3D volumes as the images.

### B. Combined data with 2D slices

In Figure 5 we show the mean 10-fold accuracy on training and test sets across 15 epochs of our model with 2D slices as images. Here we see an overall better test accuracy with

the combined model but by a small margin. At epoch 11 the combined data gives 63% accuracy whereas the gene expression alone gives 62.4% at epoch 13. This difference however is not statistically significant.

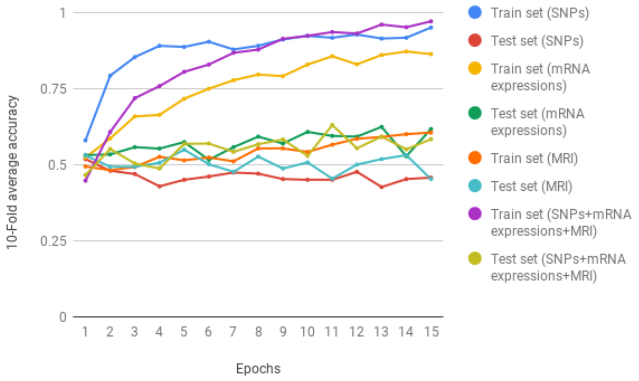


Fig. 5. Mean 10-fold accuracy of our network across 15 epochs for training and test sets with handpicked 2D slices (that manifest the tumor) as the image data.

When we see the test accuracy of each of the 10 folds on the combined data we see considerable variation as shown in Figure 6. Again this shows that in some folds our train and test distributions are likely to be the same and in others very different, thus making it hard to classify.

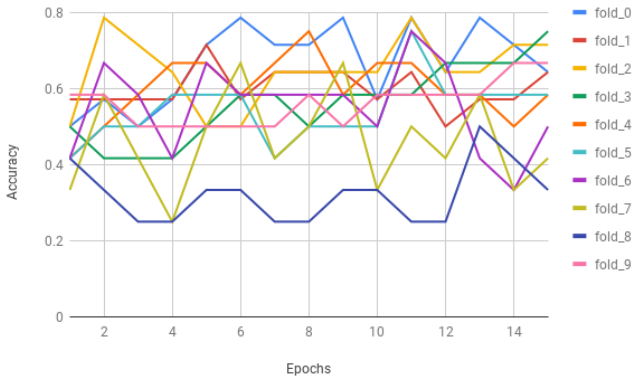


Fig. 6. Test accuracy of each of the 10-folds of our network across 15 epochs on all three data sources combined with 2D slices as the images.

## IV. Discussion

We see that combining data into three sources yields a slight improvement that is not statistically significant. Clearly simply by combining more data we cannot expect to predict survival time accurately but we also possibly need to enlarge our training set size. The variation in test accuracy across the folds suggests model instability that we attribute to insufficient data. One possible avenue to solve this is to generate artificial samples for all three sources with a generative model like a generative adversarial network [9].

Another thing in our results is the 100% accuracy in training on the combined data in both 2D and 3D. Could our model be overfitting? We add dropout [16] that is a popular and powerful method to reduce overfitting. It reduces the training accuracy all the way down to in the 50-60% range and does not improve test accuracy. This suggests we may need a richer model with dropout since even fitting training samples becomes very hard with this method.

Finally we see that the 2D combined model performs better than the 3D. A 3D model in general requires much more data than a 2D which is one likely reason for the 3D model's poorer performance. We fine-tuned our 3D U-Net but it did not improve accuracy. Again we conjecture here that additional data points via generative modeling may increase accuracy.

## V. Conclusion

We show that integrating genomic and neuroimages in multi-path neural network slightly improves glioma survival time prediction at the 14 month threshold. We see instability in test accuracy in our model and conjecture that a larger sample size produced via a generative model may improve stability and overall accuracy.

## References


[1] Broad Institute TCGA Genome Data Analysis Center Firehose stddata_2016_01_28 run. Broad Institute of MIT and Harvard (2016). https://doi.org/10.7908/C11G0KM9

[2] Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., et al.: From fastq data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. Current protocols in bioinformatics **43**(1), 11–10 (2013)

[3] 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. Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries pp. 369–379 (2019). https://doi.org/10.1007/978-3-030-11726-9_33

[4] Bottou, L.: Large-scale machine learning with stochastic gradient descent. In: Proceedings of COMPSTAT'2010, pp. 177–186. Springer (2010)

[5] Chato, L., Latifi, S.: Machine Learning and Deep Learning Techniques to Predict Overall Survival of Brain Tumor Patients using MRI Images, pp. 9–14 (2017), https://ieeexplore.ieee.org/xpl/conhome/8215806/proceeding

[6] Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., Moore, S., Phillips, S., Maffitt, D., Pringle, M., et al.: The cancer imaging archive (tcia): maintaining and operating a public information repository. Journal of digital imaging **26**(6), 1045–1057 (2013)

[7] Cochran, W.G.: The $\chi 2$ test of goodness of fit. The Annals of Mathematical Statistics pp. 315–345 (1952)

[8] DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., Del Angel, G., Rivas, M.A., Hanna, M., et al.: A framework for variation discovery and genotyping using next-generation dna sequencing data. Nature genetics **43**(5), 491 (2011)

[9] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: Advances in neural information processing systems. pp. 2672–2680 (2014)

[10] Gwilliam, B., Keeley, V., Todd, C., Roberts, C., Gittins, M., Kelly, L., Barclay, S., Stone, P.: Prognosticating in patients with advanced cancer–observational study comparing the accuracy of clinicians' and patients' estimates of survival. Annals of Oncology **24**(2), 482–488 (2012). https://doi.org/10.1093/annonc/mds341

[11] 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)

[12] McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al.: The genome analysis toolkit: a mapreduce framework for analyzing next-generation dna sequencing data. Genome research **20**(9), 1297–1303 (2010)

[13] Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., Lerer, A.: Automatic differentiation in pytorch. In: NIPS-W (2017)

[14] Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)

[15] Soltaninejad, M., Zhang, L., Lambrou, T., Yang, G., Allinson, N., Ye, X.: Mri brain tumor segmentation and patient survival prediction using random forests andfully convolutional networks. Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries pp. 204–215 (2018). https://doi.org/10.1007/978-3-319-75238-9_18

[16] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: A simple way to prevent neural networks from overfitting. The Journal of Machine Learning Research **15**(1), 1929–1958 (2014)

[17] Stiel, S., Bertram, L., Neuhaus, S., Nauck, F., Ostgathe, C., Elsner, F., Radbruch, L.: Evaluation and comparison of two prognostic scores and the physicians estimate of survival in terminally ill patients. Supportive Care in Cancer **18**(1), 43–49 (2010). https://doi.org/10.1007/s00520-009-0628-0

[18] Tamimi, A.F., Juweid, M.: Epidemiology and Outcome of Glioblastoma, pp. 143–153. Codon Publications (2017), https://www.ncbi.nlm.nih.gov/books/NBK470003/

[19] Thakkar, J.P., Dolecek, T.A., Horbinski, C., Ostrom, Q.T., Lightner, D.D., Barnholtz-Sloan, J.S., Villano, J.L.: Epidemiologic and molecular prognostic review of glioblastoma. Cancer Epidemiology Biomarkers & Prevention **23**(10), 1985–1996 (2014). https://doi.org/10.1158/1055-9965.epi-14-0275

[20] Wang, Y., Nie, J., Yap, P.T., Li, G., Shi, F., Geng, X., Guo, L., Shen, D.: Knowledge-guided robust mri brain extraction for diverse large-scale neuroimaging studies on humans and non-human primates. PLoS ONE **9**(1), e77810 (2014). https://doi.org/10.1371/journal.pone.0077810

[21] Woolrich, M.W., Jbabdi, S., Patenaude, B., Chappell, M., Makni, S., Behrens, T., Beckmann, C., Jenkinson, M., Smith, S.M.: Bayesian analysis of neuroimaging data in fsl. NeuroImage **45**(1), S173–S186 (2009). https://doi.org/10.1016/j.neuroimage.2008.10.055

[22] Zong, W., Lee, J., Liu, C., Snyder, J., Wen, N.: Abstract 3351: Overall survival prediction of glioblastoma patients combining clinical factors with texture features extracted from 3-d convolutional neural networks. Science and Health Policy (2019). https://doi.org/10.1158/1538-7445.sabcs18-3351