



## Classification of MGMT Promoter Methylation in Glioblastoma Patients Using EfficientNet-RNN

---

S Manoj, T V Raghavasimhan, S Umamaheswari and D Sangeetha

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

January 22, 2023

# Classification of MGMT Promoter Methylation in Glioblastoma Patients Using EfficientNet-RNN

Manoj S

*Department of Information Technology  
Madras Institute of Technology – Anna University  
manoj.selvam312@gmail.com*

Dr S Umamaheswari

*Department of Information Technology  
Madras Institute of Technology – Anna University  
uma\_sai@mitindia.edu*

Raghavasimhan T V

*Department of Information Technology  
Madras Institute of Technology – Anna University  
raghavtv@outlook.com*

Dr D Sangeetha

*Department of Information Technology  
Madras Institute of Technology – Anna University  
dsangeetha@mitindia.edu*

**Abstract**—Glioblastoma (GBM) is a deadly malignant brain tumor. The biggest threat is the very low survival time, as it reduces the chances of administering the right treatment. For patients with GBM, time is extremely precious and an accurate prognosis is essential. The problem that is prevalent is the invasiveness in the procedures required to identify if the standard treatment will be effective for the patient. Due to lack of a direct indication in the MRI, doctors are forced to perform surgery to identify the genotypic indication of the prognosis. Promoter methylation of the MGMT biomarker indicates a better response to chemotherapy and longer survival times in patients with GBM. While MGMT itself is not identifiable from images, this project attempts to predict its presence from multimodal MRI data acquired from Radiological Society of North America (RSNA). The proposed method trains an EfficientNet-RNN model for each of the 4 modalities of the MRI and fuses their individual outputs to produce an AU-ROC score of 0.5876, which is an improvement on the individual values.

**Index Terms**—Glioblastoma, brain, genotype, MGMT, MRI, multimodal, deep learning

## I. INTRODUCTION

Glioblastoma is one of the deadliest cancers due to the very low median survival time of the patients affected by this condition[1]. It is essential to identify and begin the best mode of treatment quickly to improve survival time. It is found that administering temozolomide(TMZ) can prove beneficial to patients and could contribute to long term survival[2]. However, TMZ and all forms of chemotherapy are highly toxic and have severe side effects such as multiple organ failure. It is not humane to begin treatment without any assurance of the efficacy of the treatment. The efficacy of the accepted treatment methods could be hindered by the DNA repair enzyme O(6)-methylguanine-DNA methyltransferase (MGMT) when it is active[3]. Promoter methylation of MGMT is shown to favour long-term survival in about 71% of patients[4]. Therefore, it would be safer to identify the status of the MGMT promoter before starting treatment. Standard methods for identifying MGMT promoter status such as RT-PCR tests are invasive, expensive and time-consuming. Hence, it is essential to have a method that is safe, quick and reliable while also being

economically viable. This paper proposes a method to use multimodal MR Image data to predict the promoter status.

### A. MGMT Promoter Methylation

O6-methylguanine-DNA methyltransferase, known as MGMT is a DNA repair enzyme. In the case of glioblastoma patients undergoing alkylating agent chemotherapy, the active MGMT gene leads to poor results as it leads to repair of cancer cells. This is highly unfavourable and counter-productive. This also leads us to ponder if the inactivity of MGMT could offer considerable improvement in the treatment of patients with glioblastoma. Studies have shown that the methylation of MGMT promoter leads to selective gene silencing, subsequently improving the efficacy of the medication[3].

### B. Magnetic Resonance Imaging

MR Imaging is widely used for diagnosis and post-treatment inspection of patients with glioblastoma. The main modalities that are available include T1-weighted (T1w) and contrast-enhanced T1 (T1wCE), T2-weighted (T2w) and Fluid-Attenuated Inversion Recovery (FLAIR). Each of these provide essential information about the tumor and its environment. T1w and T1wCE are preferable for harder material such as muscle; T2w offers a clearer picture of soft tissues; FLAIR is similar to T2 but also enhances the cerebrospinal fluid (CSF) in the image[5].

### C. EfficientNet

EfficientNet[13] is a family of models that implements a uniform scaling method with the help of compound scaling coefficients. It performs well on common transfer learning datasets such as ImageNet, CIFAR-100 among others, and produces state-of-art accuracy while being computationally less intensive than standard CNNs.

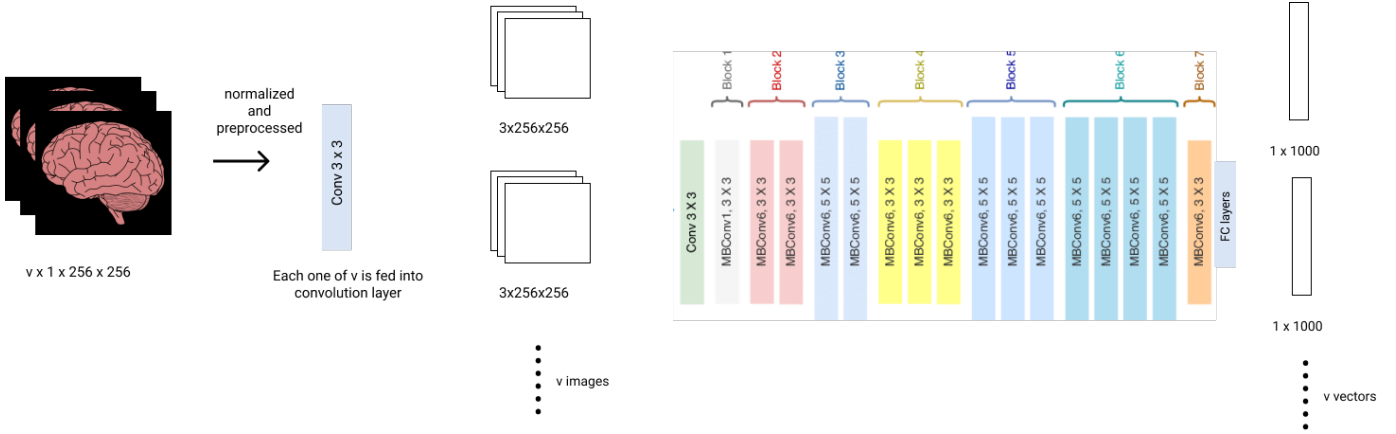


Fig. 1: Backbone Feature Extractor.

## II. RELATED WORK

### A. Video Resnet

It uses a full convolution layer which predicts by combining all the slices trying to extract features without localizing specific slice information. Here the layers are similar to resnet but the kernel is 3D instead of the 2D kernel as in resnet. Video resnet is used to analyse video files and predict output. To train using this technique, we fix the number of slices. The model does not treat slices with and without tumors as different entities, leading to unnecessary training.

### B. Resnet

Resnet was the first model to implement a residual layer in which the output of the previous layer is added to the current layer which creates an effect of regularization within the network thereby giving the ability to generate models with more layers without overfitting and vanishing gradient problem. We used Resnet 18 (18 layers) as feature extractor which was pre trained on image net dataset (having 30.24% as Top-1 error). We used the full model which gave us output of a 2048 dimension vector which then is fed into another layer of CNN and finally passed into RNN. Resnet does not efficiently scale as Efficient Net.

### C. Unet RNN

Unet is an image segmentation model which has two parts: an encoder and a decoder. The encoder tries to extract the features and reduce the dimension of input and decoder tries to generate the segmentation mask of the actual image. We trained the standard Unet model with BraTS[7][8] segmentation data. The model predicts tumor segments with a DICE loss of 0.24. The idea was to use the encoder as our feature extractor so that model would be able to extract features which were related to the tumor. By this way we could give a single channel (corresponding to the specific modal) input to the pretrained single channel input receiving UNet. The accuracy was less compared to the feature extractor we are using currently.

## III. ARCHITECTURE

The  $v$  here (in Fig. 1) represents the number of slices. The proposed architecture accommodates a varying number of slices per image. Each slice is an image of dimension  $256 \times 256$ . This tensor is then normalised and fed into a  $3 \times 3$  kernel convolution layer which treats each slice as separate images. This produces  $v$  images each with dimensions  $3 \times 256 \times 256$ . The output is passed to an efficient net for feature extraction. An extra full convolution layer is added so that the extracted features are in the required dimensions. Instance normalisation is applied on each convolution layer.

The  $v$  vectors are fed into a LSTM layer for processing all the slices combined to give a 1024 dimension output vector. Another fully connected layer and a sigmoid activation layer are added to give a single class output probability. The weighted average of the four modalities trained separately on the same architecture is calculated to give the final prediction.

### A. Efficient Net

Different variations of efficient net were considered. EfficientNet-b4 was chosen as it produces optimum results. We tried incrementing the number of hidden layers in LSTM and the number of parameters in the b4 model balanced with the optimum number of hidden layers considering the RAM space we had.

### B. Instance Normalisation

Normalisation maintains the intermediate output towards unit gaussian distribution thereby solving the vanishing gradient problem. Batch normalisation is the most commonly used technique which normalises the output per batch. Here, the batch size is the number of slices of the MR image. Normalisation across the batch is less insightful and hence, instance normalisation was used. This can normalise slices independent of the entire batch. This provides improved accuracy for shorter epochs.

### C. LSTM

LSTM layer is used in speech processing to process a sentence word-by-word. It is able to relate words that occur early in the sentence to those that appear later. We relate the vectorised feature extracted slice information to a word and feed it into a LSTM layer. This provides an efficient way to distinguish slices with and without the favourable features and passes on the necessary information for the final prediction.

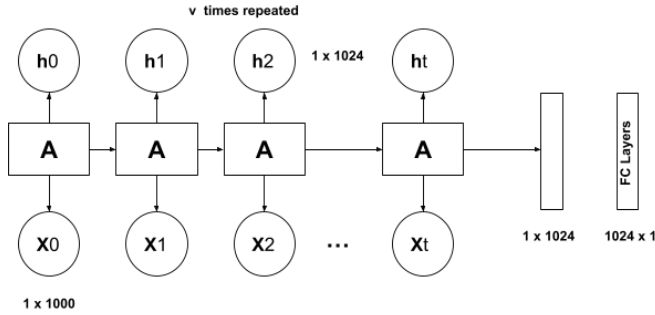


Fig. 2: LSTM layer.

### D. Multimodal Fusion

Often, any one modality of the MRI is used for training the model due to availability constraints. To exploit all four modes, we could combine the 4 modalities and feed into a single neural network thereby enabling it to correlate features of different modalities. In this technique, we would be passing each modality of the MR image as a channel and the feature extractor would be able to predict output based on combining the information in all 4 channels. An analysis of the dataset (shown in Fig. 3) showed that each mode had varying numbers of slices for different patients.

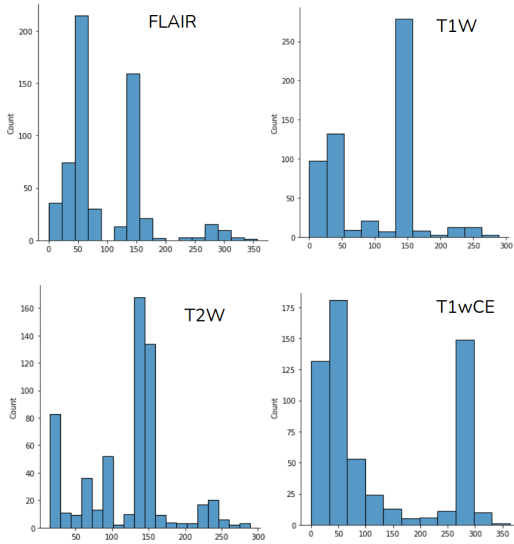


Fig. 3: Plot of number of slices for image vs number of images with slices in the specified range.

On analysis of the dataset all the channels had varying numbers of slices for different images. Thereby we were not able to fix the number of slices (which should be constant for combining different channels). If we take random slices to fix the number of slices we would compromise highly critical data for prediction.

### IV. IMPLEMENTATION AND RESULT

This project uses a deep learning based approach to extract features from the MR images and also to classify the images into two classes - MGMT Promoter Methylated and MGMT Promoter Unmethylated. The architecture is explained in detail in the previous section.

#### A. Training and Validation

Binary cross entropy loss was used in training the model. AdamW optimiser was used and the model was trained with a learning rate of 10e-6. The model was trained for 50-60 epochs with early stopping as a regularisation method to avoid the possibility of overfitting. The model was trained separately on each modality of the MRI. The final output was calculated by calculating a weighted average of the individual models' outputs. The weights can be assigned based on knowledge of the domain or can be learned with the help of machine learning algorithms.

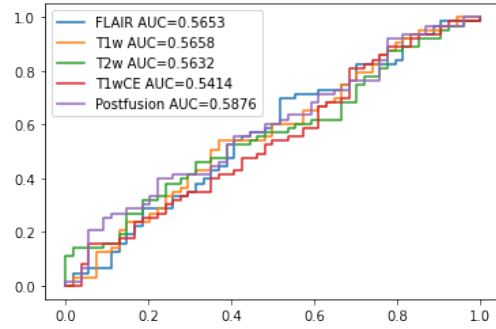


Fig. 4: ROC curve.

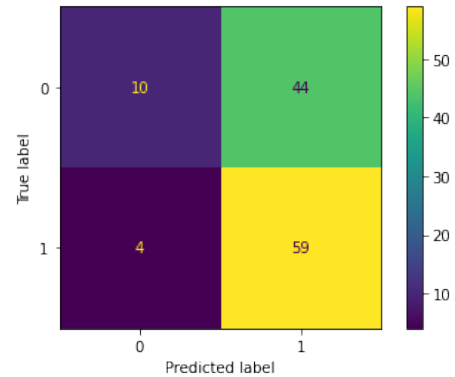


Fig. 5: Confusion Matrix after outputs are fused.

TABLE I: Study of pre-existing methods.

Scan Mode	val_acc	Model	train_loss	val_loss
Without T2w	0.6059	r2plus1d_18	0.6885	0.6837
Without T2w	0.5547	resNext	0.2669	3.6044
FLAIR	0.5324	monai	0.6905	0.6990
All	0.5312	efficientnet-b0	0.6365	0.7217
T2w	0.5154	unet	0.689	0.6952
Without T2w	0.4740	resNext50	0.5940	0.7598

TABLE II: Efficient Net metrics

Name	Scan Mode	val_acc	val_loss	rnnh	rnnl
efficientnet-b4-T2w	T2w	0.6323	0.6852	1024	1
efficientnet-b5-T2w	T2w	0.6488	0.6858	512	2
efficientnet-b4-T1wCE	T1wCE	0.6309	0.6855	512	2
efficientnet-b6-T1w	T1w	0.6501	0.6871	512	2
efficientnet-b6-T1wCE	T1wCE	0.6507	0.6845	512	2
efficientnet-b4-FLAIR	FLAIR	0.6504	0.6829	512	2
efficientnet-b6-FLAIR	FLAIR	0.6387	0.677	1024	1

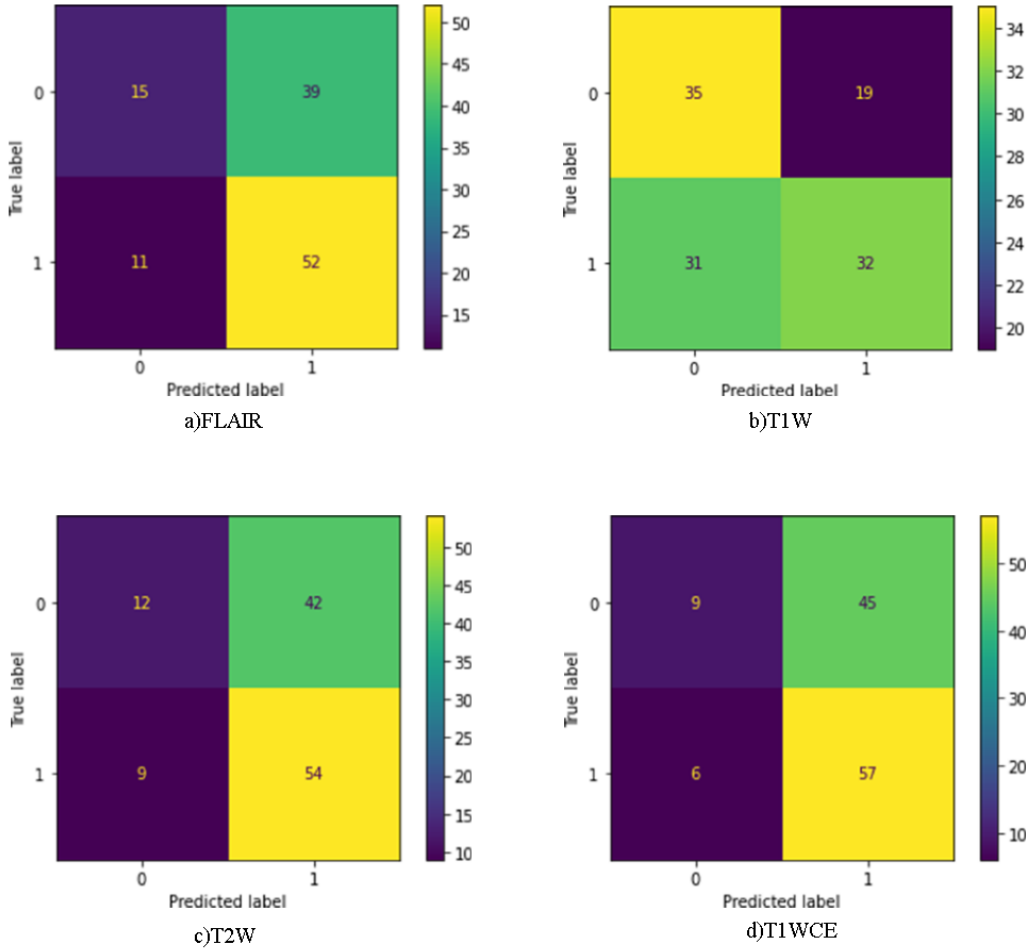


Fig. 6: Confusion matrices for model trained on 4 modalities.

### B. Implementation Environment

The model was trained on the NVIDIA TESLA P100 Cloud GPU provided by Kaggle on the predefined environment. WB was used to log the results and visualise the data to draw meaningful conclusions from various metrics.

### C. Performance

The curve in Fig. 4 shows that Postfusion performs significantly better than the individual models with an **AU-ROC**

**score of 0.5876**. Table 1 provides a comparison of various standard techniques and their performance on this data.

### V. CONCLUSION

This work predicts the methylation of MGMT promoter using Efficient net as feature extractor and LSTM to process the features to arrive at the prediction. The difference from past attempts at similar problems is very significant in the way the modalities are fused. “Post-fusion” provides a way to handle medical data in deep learning when the 3D image slices are not the same in different modalities.

While the focus has been on optimising the deep learning model, this work remains faithful to the goal of helping patients diagnosed with glioblastoma by providing a quick and non-invasive solution and in the future, would proceed in the direction of improving feature extraction techniques.

#### REFERENCES

- [1] Bi, W. and Beroukhi, R., 2014. Beating the odds: extreme long-term survival with glioblastoma. *Neuro-Oncology*, 16(9), pp.1159-1160.
- [2] Hegi ME, Dierens AC, Gorlia T, et al. MGMT gene silencing and benefit from temozolomide in glioblastoma. *N Engl J Med*. 2005;352(10):997-1003.
- [3] Weller M, Stupp R, Reifenberger G, Brandes AA, van den Bent MJ, Wick W, Hegi ME. MGMT promoter methylation in malignant gliomas: ready for personalized medicine? *Nat Rev Neurol*. 2010 Jan;6(1):39-51. doi: 10.1038/nrneuro.2009.197. Epub 2009 Dec 8. PMID: 19997073.
- [4] Gerber NK, Goenka A, Turcan S, et al. Transcriptional diversity of long-term glioblastoma survivors. *Neuro Oncol*. 2014;16(9):1186-1195. doi:10.1093/neuonc/nou043
- [5] Shukla G, Alexander GS, Bakas S, Nikam R, Talekar K, Palmer JD, Shi W. Advanced magnetic resonance imaging in glioblastoma: a review. *Chin Clin Oncol*. 2017 Aug;6(4):40. doi: 10.21037/cco.2017.06.28. PMID: 28841802.
- [6] Nguyen N, Redfield J, Ballo M, Michael M, Sorenson J, Dibaba D, Wan J, Ramos GD, Pandey M. Identifying the optimal cutoff point for MGMT promoter methylation status in glioblastoma. *CNS Oncol*. 2021 Sep 1;10(3):CNS74. doi: 10.2217/cns-2021-0002. Epub 2021 Sep 6. PMID: 34486380; PMCID: PMC8461752.
- [7] U.Baid, et al., "The RSNA-ASNR-MICCAI BraTS 2021 Benchmark on Brain Tumor Segmentation and Radiogenomic Classification", arXiv:2107.02314, 2021.
- [8] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, et al. "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)", *IEEE Transactions on Medical Imaging* 34(10), 1993-2024 (2015) DOI: 10.1109/TMI.2014.2377694
- [9] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J.S. Kirby, et al., "Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features", *Nature Scientific Data*, 4:170117 (2017) DOI: 10.1038/sdata.2017.117
- [10] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. Kirby, et al., "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection", *The Cancer Imaging Archive*, 2017. DOI: 10.7937/K9/TCIA.2017.KLXWJJ1Q
- [11] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. Kirby, et al., "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection", *The Cancer Imaging Archive*, 2017. DOI: 10.7937/K9/TCIA.2017.GJQ7R0EF MICCAI BRATS - The Multimodal Brain Tumor Segmentation Challenge. (2021). Retrieved 9 November 2021, from <http://braintumorsegmentation.org/>
- [12] lukemelas. *EfficientNet-PyTorch*. (2021). Retrieved 9 November 2021, from <https://github.com/lukemelas/EfficientNet-PyTorch>
- [13] Mingxing Tan, Quoc V. Le. [1905.11946] *EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks*. (2021). Retrieved 9 November 2021, from <https://arxiv.org/abs/1905.11946>
- [14] *Understanding LSTM Networks*. (2021). Retrieved 9 November 2021, from <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>
- [15] *U-NET FOR BRAIN MRI*. (2021). Retrieved 9 November 2021, from [https://pytorch.org/hub/mateuszbeda\\_brain-segmentation-pytorch\\_unet/](https://pytorch.org/hub/mateuszbeda_brain-segmentation-pytorch_unet/)
- [16] Tran, D., Wang, H., Torresani, L., Ray, J., LeCun, Y., Paluri, M. (2018). A closer look at spatiotemporal convolutions for action recognition. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition* (pp. 6450-6459).
- [17] Sak, H., Senior, A., Beaufays, F. (2014). Long short-term memory based recurrent neural network architectures for large vocabulary speech recognition. arXiv preprint arXiv:1402.1128.