scientific reports

Check for updates

OPEN

Deep learning classification of MGMT status of glioblastomas using multiparametric MRI with a novel domain knowledge augmented mask fusion approach

İlker Özgür Koska[1,2]✉ & Çağan Koska[3]

We aimed to build a robust classifier for the MGMT methylation status of glioblastoma in multiparametric MRI. We focused on multi-habitat deep image descriptors as our basic focus. A subset of the BRATS 2021 MGMT methylation dataset containing both MGMT class labels and segmentation masks was used. A comprehensive mask fusion approach was developed to select relevant image crops of diseased tissue. These fusion masks, which were guided by multiple sequences, helped collect information from the regions that seem disease-free to radiologists in standard MRI sequences while harboring pathology. Integrating the information in different MRI sequences and leveraging the high entropic capacity of deep neural networks, we built a 3D ROI-based custom CNN classifier for the automatic prediction of MGMT methylation status of glioblastoma in multi-parametric MRI. Single sequence-based classifiers reached intermediate predictive performance with 0.65, 0.71, 0.77, and 0.82 accuracy for T1W, T2W, T1 contrast-enhanced, and FLAIR sequences, respectively. The multiparametric classifier using T1 contrast-enhanced and FLAIR images reached 0.88 accuracy. The accuracy of the four-input model that used all sequences was 0.81. The best model reached 0.90 ROC AUC value. Integrating human knowledge in the form of relevant target selection was a useful approach in MGMT methylation status prediction in MRI. Exploration of means to integrate radiology knowledge into the models and achieve human-machine collaboration may help to develop better models. MGMT methylation status of glioblastoma is an important prognostic marker and is also important for treatment decisions. The preoperative non-invasive predictive ability and the explanation tools of the developed model may help clinicians to better understand imaging phenotypes of MGMT methylation status of glial tumors.

Keywords MGMT methylation, Glioblastoma, Artificial intelligence, Deep learning, Model explanation

Abbreviations

TMZ	Temozolomide
MGMT	Methyl guanine methyl transferase
DL	Deep Learning
GBM	Glioblastoma multiforme
BRATS	Brain tumor segmentation challenge
ANN	Artificial neural network

Standard treatment of high-grade glial tumors is radiotherapy coupled with chemotherapy after the widest possible surgical removal. Studies have shown that adding temozolomide (TMZ) to the treatment may double the average 2-year survival in certain conditions[1,2].

Methyl-guanine-methyltransferase (MGMT) is the most important DNA repair protein that protects the cellular genome from the mutagenic effects of the alkylating agents. In the presence of MGMT, the lethal effect

[1]Department of Radiology, Behçet Uz Children's Hospital, Izmir, Turkey. [2]Department of Biomedical Technologies, Dokuz Eylül Universtiy The Graduate School of Natural and Applied Sciences, Buca, Izmir, Turkey. [3]Department of Electrical Electronical Engineering, Yaşar University, Bornova, İzmir, Turkey. ✉email: ozgurkoska@yahoo.com

of alkylating agents decreases. However, MGMT is in certain situations epigenetically silenced by methylation of its promoter region[2].

In patients who are candidates for TMZ treatment, the determination of the drug-sensitive group before the administration of the treatment would be a better strategy. The standard way for this is surgically sampling the tissue and then applying molecular-genetic methods to the sample. This is an expensive method that carries many risks due to the operation and has low reproducibility[2]. Moreover, the tumor tissue may be heterogeneous regarding the MGMT methylation status, leading to erroneous decisions since the sampled part may not represent the entire tumor environment. Different cellular populations that are packed together to form different tumoral microenvironments are called habitats. These habitats might be responsible for differential treatment responses and require different treatment strategies. Furthermore, the epigenetic status of these tumors may exhibit temporal heterogeneity as well[2–4]. Therefore, determining the MGMT status of the tumor with imaging instead of surgery has great potential for healthcare delivery efficiency. Various groups researched to determine the MGMT status of glial tumors in MR images by using the professional knowledge of the radiologist, radiomics, or deep learning (DL) methods[5]. Most of these studies were carried out with around 100 samples and with such small datasets, radiomics methods were the natural choice.

In this research paper, we aimed to develop an AI-based model that can determine MGMT methylation status by using brain MRIs of GBM patients in the BRATS 2021 dataset.

The contribution of our study to the literature is threefold. First, our study analyses 577 patients. It is one of the largest sample sizes for this problem. Second, DL studies generally feed the whole image to the model in an end-to-end fashion. This is partially due to the knowledge gap of the machine learning researchers about the pathology and partially due to the strive of the developers to fully automate the process. However, most of the pixels across the image are repetitive, reflecting normal anatomy. The descriptive pixels reside only in a small fraction of the image. Since the useless information to perform a task is just noise, most of the pixels in the image can be assumed as noise considering the classification task. We eliminated this noise and enhanced the driving signal by masking out the region of interest (ROI), keeping the 3D context, and adding multiparametric contributions of different sequences to help the model converge better. This is a kind of human-machine collaboration. Third, as far as we know, there is no study in the literature that tackles the problem as a signal agnostic problem by analyzing the normal appearing pixels in a particular sequence, if they exhibit pathologic signal in any other sequence, thus benefiting from the guidance of different MRI sequences.

Method

Institutional Review Board approval was obtained from the local ethics committee for this retrospective study. Informed patient consent was waived by the Katip Çelebi University local ethical board. This study has been performed in accordance with the Declaration of Helsinki.

Dataset

The MR images in the BRATS dataset included T1W, T1CE, T2W, and FLAIR sequences. Task 1 of the competition was a segmentation task, and Task 2 was a classification task. Both datasets were partially overlapping, with 1251 patients in Task 1 and 585 patients in Task 2. We selected patients who had both segmentation maps and information about the MGMT methylation status (Fig. 1). The images had 240*240*155 isotropic voxels, and segmentation maps included GBM habitat labels (Fig. 2).

During annotation, the strengths of four distinct MR sequences were emphasized, and the annotation was carried out accordingly. The T1W and T1CE sequences were used to label the contrast-enhancing viable tumor zone and central necrosis. On the other hand, edema and non-enhancing tumor zones were labeled in T2W and FLAIR images. Each mask label corresponded to a different glial tumor habitat (Fig. 2)[11].

Preprocessing

BRATS images were presented as skull stripped, co-registered, and bias field corrected. Therefore, most of the preprocessing was already done. On the other hand, the pixel distribution was original. So, as a preprocessing step, we applied the [0–1] normalization process to prevent the higher values from dominating the calculations[12,13].

Relevant region extraction

Medical images exhibit repetitive patterns that show great similarity to each other. Therefore, radiologists can analyze the deviations from the normal anatomy and diagnose the diseases. The contribution of normal-appearing brain regions in a GBM patient for determining the MGMT class is limited. In contrast, the appearance of the tumor, which reflects its biology, can make a difference in this distinction. On the other hand, artificial neural networks (ANN) are extremely flexible models that can analyze many images to determine the relevant patterns for classification tasks. However, the more unnecessary information in the image is discarded before the classification, the easier they can do their job. Therefore, we thought that using segmentation maps of the tumor to mask out the significant regions for classification, thus narrowing the search space of the model, would be more effective than using the whole image. Furthermore, we hypothesized that a region exhibiting a pathological signal in a sequence should harbor pathology in all other co-registered sequences at the same region, even if it may appear normal in those sequences (Fig. 3). Therefore, we combined the mask labels and obtained a single fused general pathology detection mask (Fig. 4). Then, we obtained the ROI crops by multiplying the images of all sequences with these single fusion masks (Fig. 5). This approach effectively combined the information of different habitats in different MRI sequences. In this way, the standard 240*240*155 dimensions of the images changed and became different. However, the input size of ANN should be fixed[12,13]. Therefore, we set a fixed size for all these cropped images as 128*128*64 and resized the cropped images accordingly.

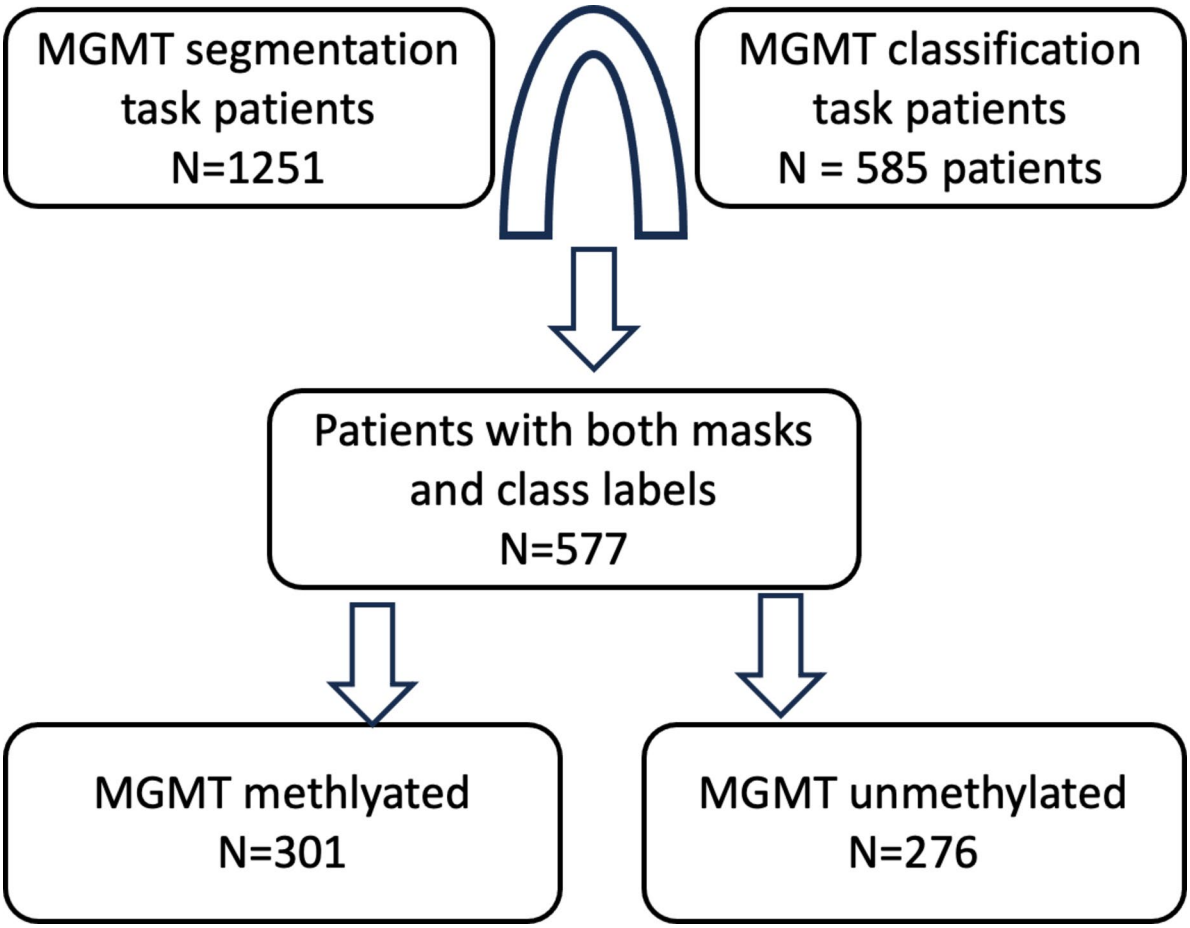


Fig. 1. Patient selection.

Input strategy

Three different input strategies were applied. All the strategies were based on masked-out isolated tumor pixels. Initially, we fed a single sequence as the input to the model. Then, we used multiple sequences as inputs to multi-input single-output models. FLAIR and T1CE sequences were used for the two-input model, and all sequences were used for the four-input model.

Model architecture

We built a 3D CNN model with four down-sampling blocks using the Keras library[14] (Fig. 6). We obtained the feature vector by applying global average pooling to the output of the feature extractor part and connected two fully connected layers, one with 512 nodes and the other with 256 nodes. Finally, we used a single node with a sigmoid activation function for binary output. ADAM with the default 1e-3 learning rate as the optimizer, binary cross entropy as the loss function, accuracy as the evaluation metric, and rectified linear unit as the activation function were the other important hyperparameters of the model. The model was trained for 100 epochs with 16 batches. 77 randomly selected class-balanced patients were reserved for validation and 100 for testing on a holdout training scheme.

Statistics

Python scripting language (version 3.7.0) and the Scipy package (version 1.4.0) were used for statistical analysis.

Predictions and ground truths were used to build contingency tables. Recall, precision, accuracy, and f1 scores were calculated using these contingency tables. Confidence intervals were calculated using 95% as the confidence level.

Results

Total number of samples was 577 with 276 MGMT unmethylated (276/577, 48%) and 301 MGMT methylated tumors (301/577, 52%). To the best of our knowledge, this is one of the largest sample sizes for this problem. On the other hand, since the patients were de-identified and anonymized, which was a mandatory step to make the dataset available to the public, it was not possible to compare demographic or clinical distributions. However, considering this dataset was prepared with the hope of drawing the attention of many researchers to the topic to find a solution to the problem, we can assume that it was prepared in a balanced manner between the two groups.

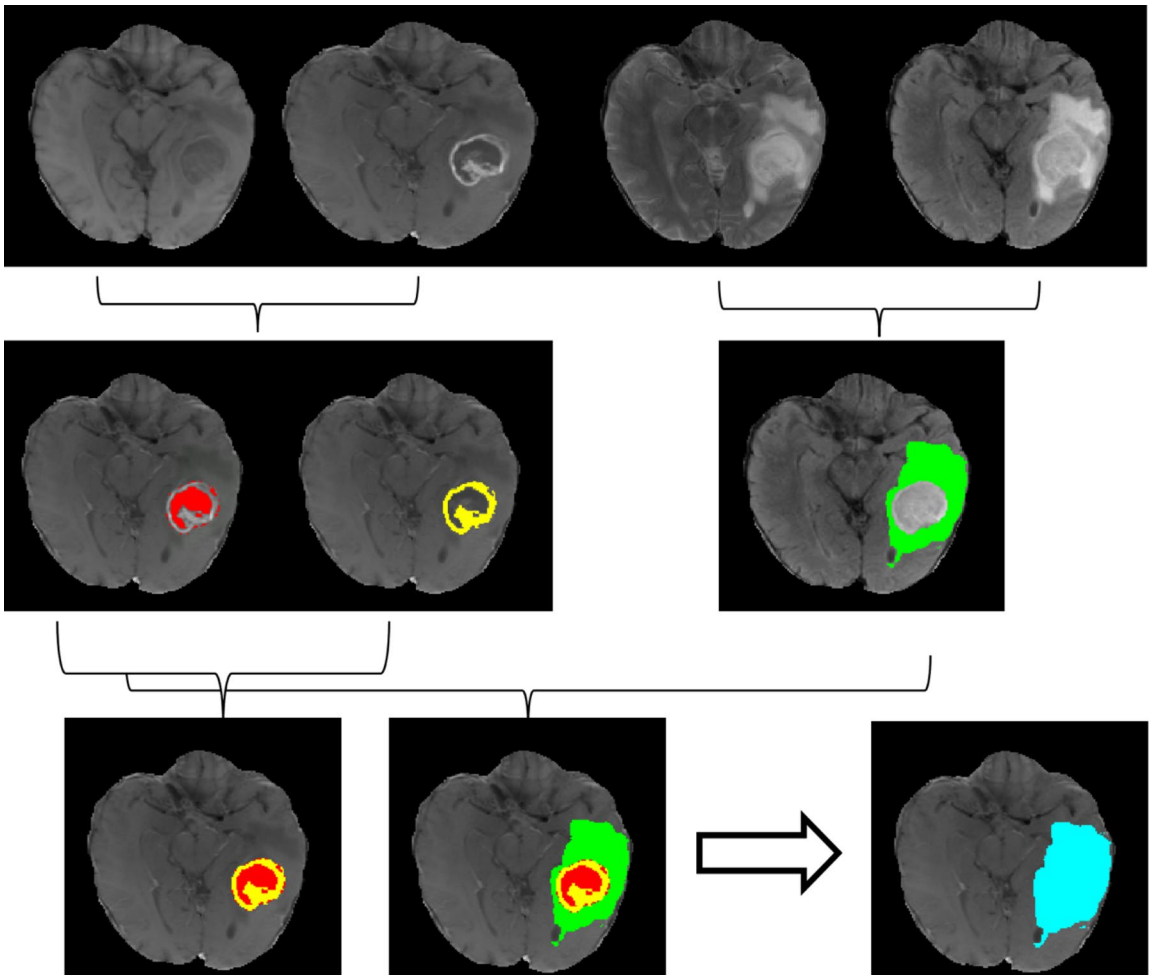


Fig. 2. Four MRI modalities and related masks that were used for this study. Enhancing tumor region (Yellow mask) and central necrosis zone (Red mask) were determined in T1W and T1CE sequences. Both comprise together the tumor core. Peritumoral edema and non-enhancing tumor (Green mask) were determined in T2W and FLAIR sequences. The information gained from all sequences was combined to construct the fusion mask, which combined all pathology-harboring regions from different sequences (Blue mask). All sequences were cropped based on this fusion mask.

There were no statistically significant differences between habitat volumes of the MGMT-methylated and the unmethylated tumors. However, enhancing tumor volume ($p = 0.14$) and necrosis volume/tumor core volume ratio ($p = 0.15$) are promising. Volume statistics of each habitat are presented in Fig. 7.

The built DL model achieved an accuracy of 0.77 (CI 95% 0.73–0.80) with T1CE images, 0.71 (CI 95% 0.67–0.74) with T2W images, 0.82 (CI 95% 0.78–0.85) with FLAIR and 0.65 (CI 95% 0.61–0.68) with non-contrast T1W images (Table 1).

We hypothesized that incorporating signal information from different sequences would enhance the model's performance. For this purpose, we gave the T1CE and FLAIR sequences together to the model and jointly trained the model with them (Fig. 8). The accuracy and the f1 score of that model were 0.88 (CI 95% 0.85–0.90 and CI 95% 0.85–0.90, respectively). We believe that this approach allowed the model to focus directly on the target by not wasting resources trying to classify unnecessary pixels and increased performance. Additionally, we used all the sequences as different inputs into a four-input model. The accuracy of the four-input model was 0.81 (CI 95% 0.77–0.84), and the f1 score was 0.80 (0.76–0.83). Many groups have tried to enhance the model performance in the literature by using multiple sequences as input. However, our approach differed from the others in two main points (Fig. 8). Other groups combined the features that they obtained using radiomics or adopted an ensemble approach. However, we gave the set of images confined to the tumor region directly, allowing the model to learn from the image. The second key difference was the approach for assembling the image crop as input to the model (Fig. 9). The other groups used crops that appeared pathologic in the corresponding sequence. However, while some pathologies can be distinguished in some sequences, the signal change in the pathological region may not be distinguishable in other sequences. This does not change the fact that there is disease in that region. Using this information, we used the segmentation maps of all sequences together and converted them into a single mask. All sequences contributed to the establishment of this mask. This fusion mask helped to isolate the regions that

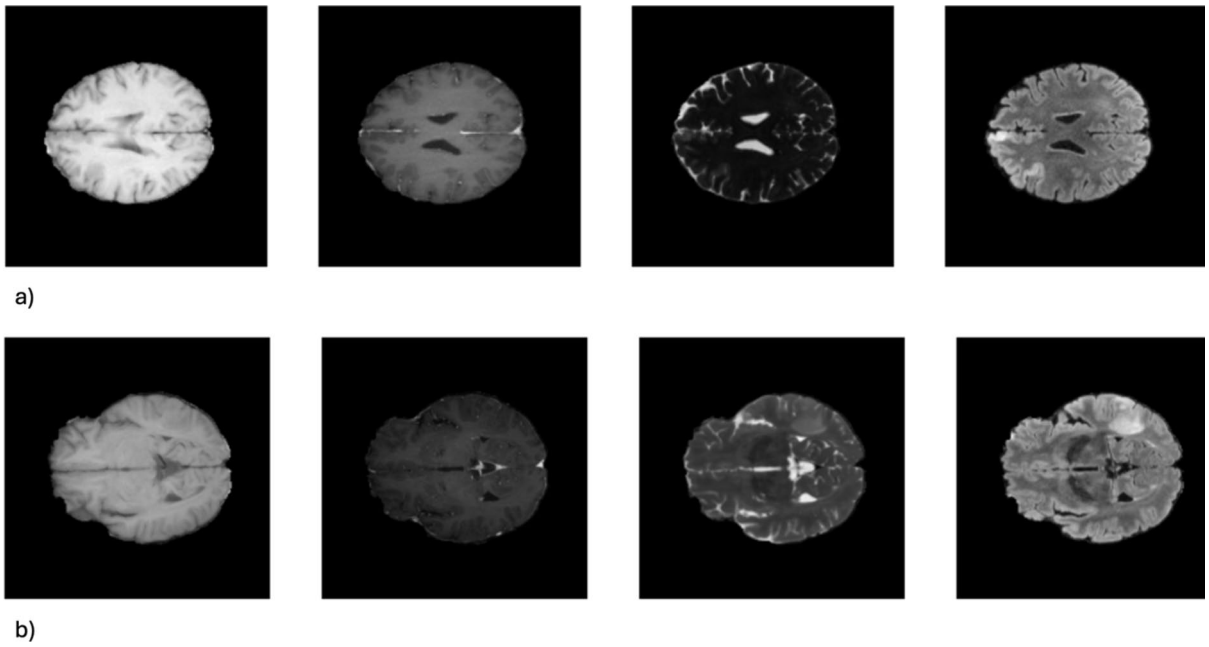


Fig. 3. The novelty of the proposed cropping strategy. Left to right T1W, T1CE, T2W, and FLAIR sequences. (**a**) Patient from MGMT unmethylated group (**b**) Patient from MGMT methylated group. In a) T1W, T1CE, and T2W appear almost normal. The masks obtained using these sequences will be empty, and any region from the T1W, T2W, and T1CE sequences will not be cropped by these masks. However, FLAIR depicts the right frontal cortical pathology. The mask obtained using the FLAIR image will mark the pathology harboring pixels. Using the combined mask and applying that mask to all sequences will crop the pathology harboring region in all sequences.

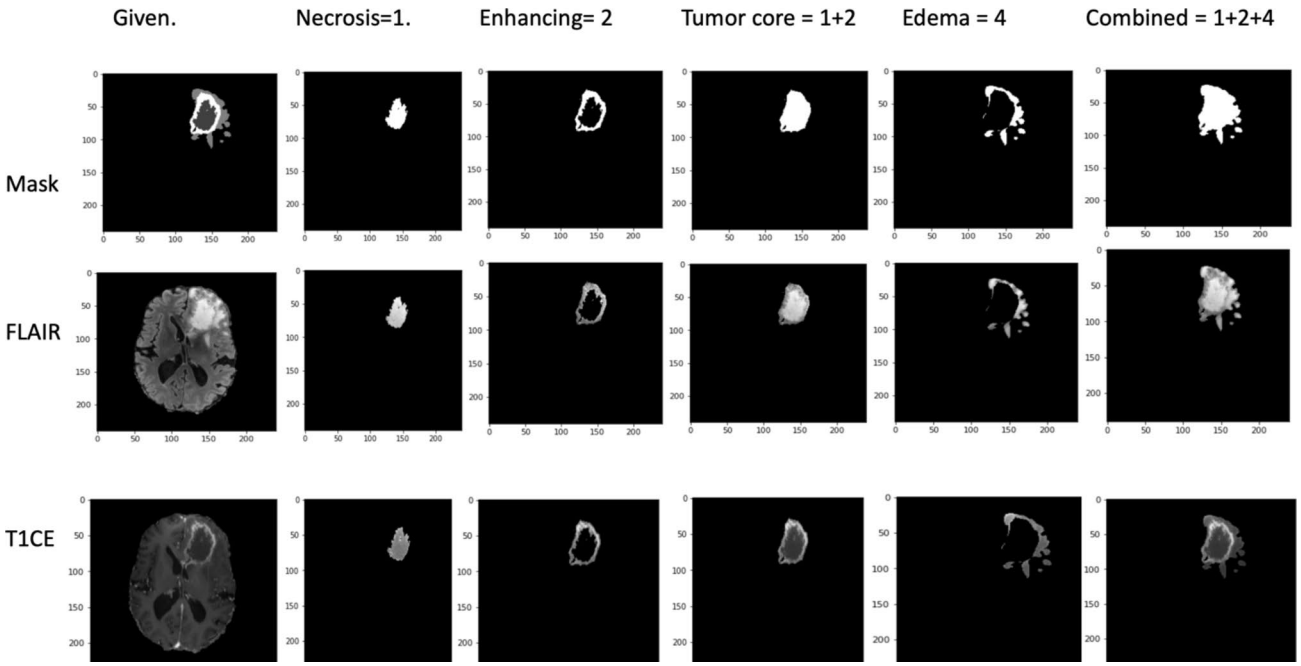


Fig. 4. Sample axial slices of FLAIR and T1CE sequences of a patient along with multilabel mask. The multilabel mask can be decomposed into individual components to obtain different habitats of tumors, which accentuate different microenvironments. We used this fusion mask, which was obtained by combining all labels, and cropped the underlying tumoral environment from each sequence with this mask (Right column).

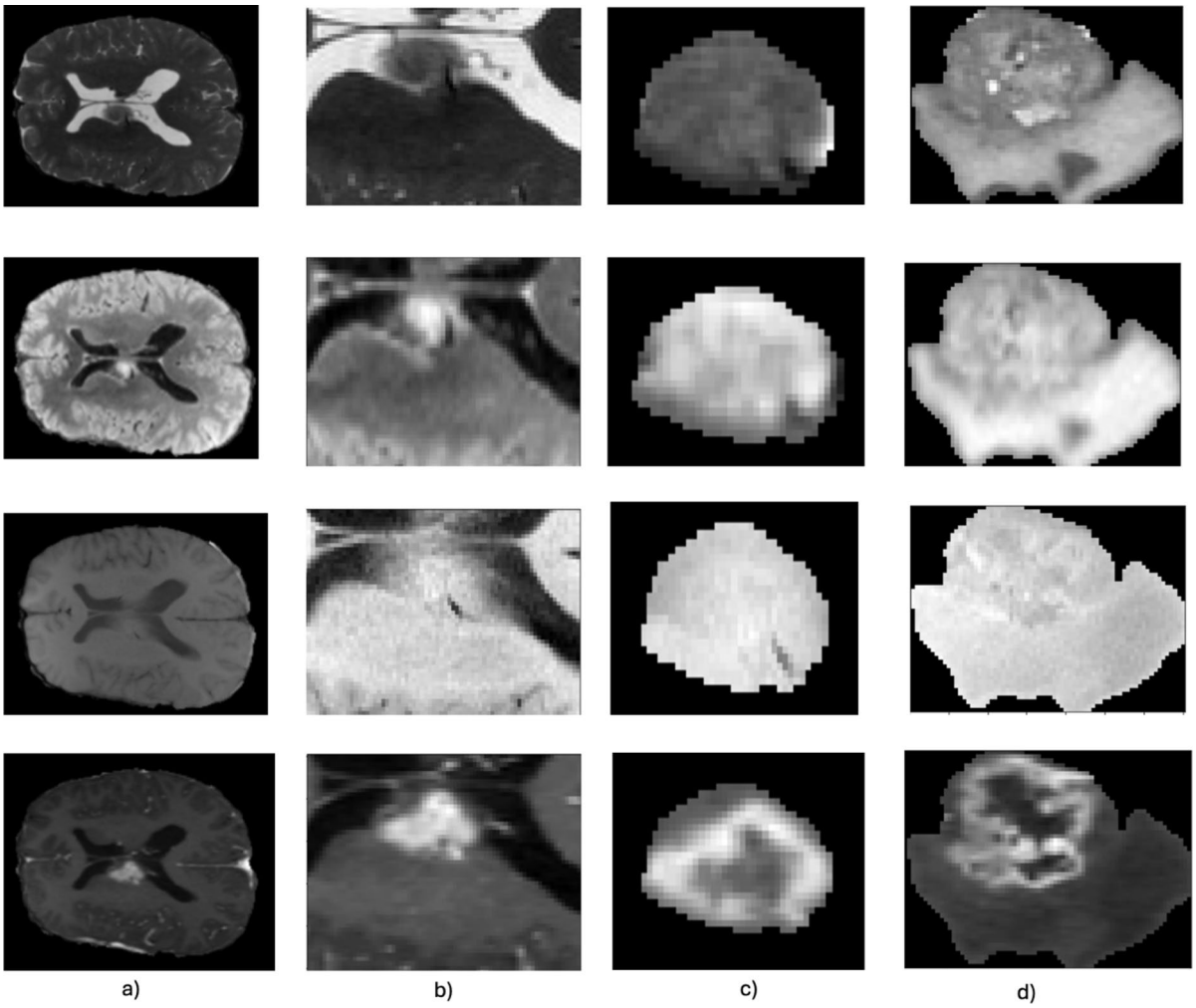


Fig. 5. Comparison of different input approaches. T2W, FLAIR, T1W, and T1CE sequences from upper to lower rows, respectively. (**a**) Whole slice, (**b**) ROI based on the maximum dimensions of the 3D mask, (**c**) and (**d**) Our approach. In our approach, the isolated tumor pixels that are completely cleaned from confounding normal anatomy pixels were used. Since the maximum sizes in three directions determine the sizes of the ROI box, in ROI-based approaches, the larger mask in a different region of the tumor size, such as seen in d), results in many nontumoral pixels in the crop, such as seen in b).

appeared diseased in at least one sequence from the rest of the image, regardless of whether there was a signal change in the processed sequence. This gave the model a chance to analyze the signal of the tissue in the diseased area in other sequences as well. Thus, we reached 0.88 accuracy and 0.90 ROC-AUC with this approach.

Furthermore, we developed a Streamlit application for experimental use. This application had image and mask exploration tools, prediction outputs of the model, and prediction explanation tools (Fig. 10). We believe the ease of use and availability of this application could help the translation of the research into clinics.

Discussion

Using the information of all sequences in a 3D context, masking out the pathological region with isolation from the neighborhood normal anatomy, and enriching the data set by using data augmentation techniques, we were able to successfully classify the MGMT methylation status of GBM in DW MRI with a 0.88 accuracy with our best model. Detection of MGMT methylation status of GBM from MR images is an important problem that has not yet been solved and has the potential to create great value in patient management, proper treatment selection, and economic benefits. Although there are some controversies regarding the favorable survival expectation of MGMT methylated tumors, which was discussed in a recent article, we believe that their results should be validated[15]. During treatment, 2/3 of MGMT methylated tumors and just 11% of MGMT unmethylated tumors exhibit pseudo-progression, which can be considered as an exaggerated response to the treatment and a sign of better response[16]. This phenomenon may be another evidence of better outcomes related to MGMT methylated tumors.

Several studies tackled the problem using either radiomics or DL methods[5–10,15,17–19]. We can notice the highest accuracy rates in three studies in the literature[7,8,17]. In the first study, the T1CE series and T2W series

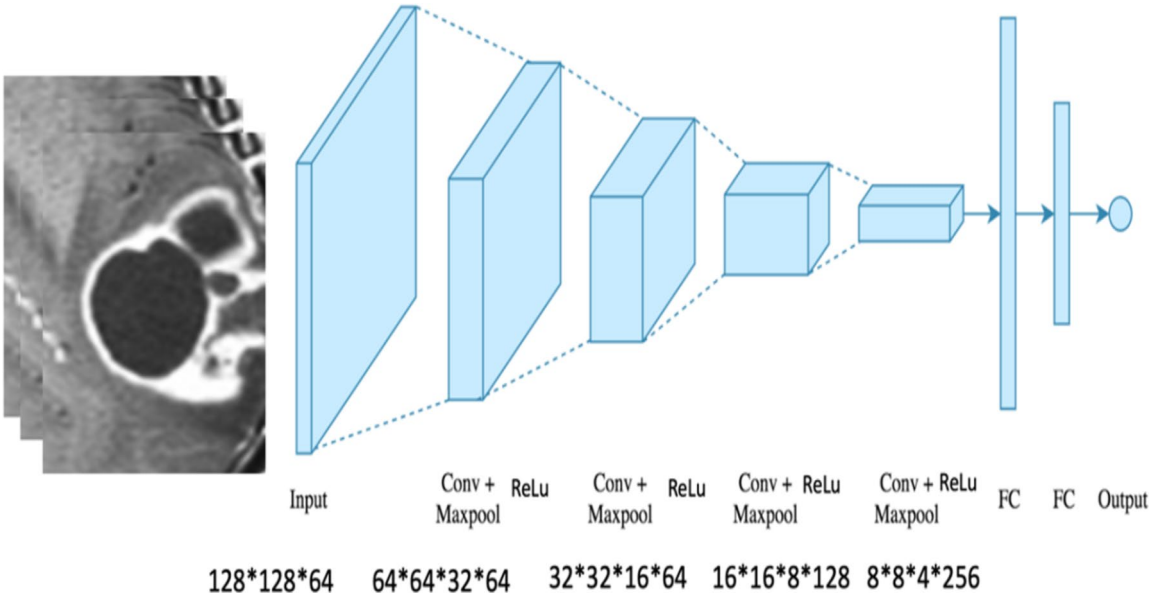
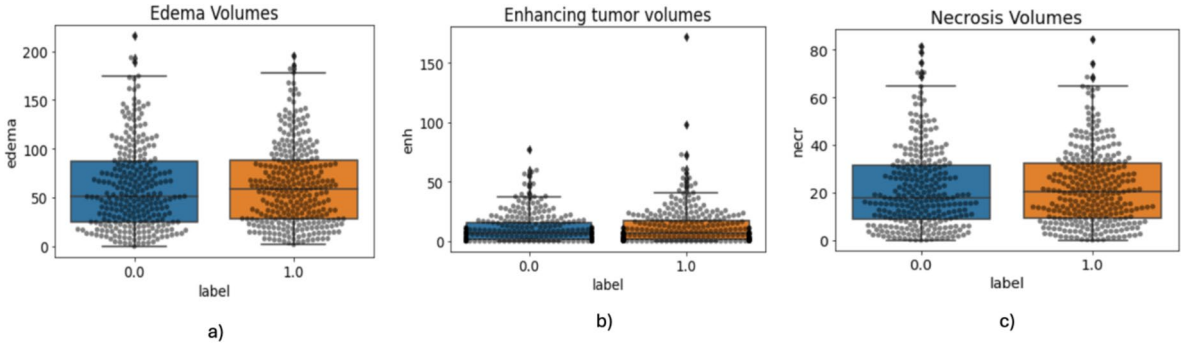


Fig. 6. Network architecture and hyperparameters.



	neg_necr	neg_enh	neg_core	neg_edema	neg_tumor
count	276.000000	276.000000	276.000000	276.000000	276.000000
mean	21.719123	10.629304	32.348428	61.040967	93.389395
std	16.781181	12.594739	26.540493	43.511436	58.021336
min	0.022000	0.000000	0.022000	0.055000	6.880000
25%	8.918250	1.441750	12.055000	24.950250	43.511500
50%	17.709500	6.444500	25.957000	50.798500	84.308000
75%	31.492250	15.890750	47.320750	87.061000	135.715000
max	81.526000	76.834000	121.765000	215.542000	255.108000

d)

	pos_necr	pos_enh	pos_core	pos_edema	pos_tumor
count	301.000000	301.000000	301.000000	301.000000	301.000000
mean	22.439894	12.427478	34.867372	62.804555	97.671927
std	15.912686	16.807534	26.994859	41.297667	54.157264
min	0.106000	0.000000	0.327000	1.859000	6.046000
25%	9.353000	1.857000	12.433000	28.009000	55.781000
50%	20.409000	7.029000	30.227000	58.299000	94.051000
75%	32.208000	17.651000	48.228000	88.222000	134.271000
max	84.287000	171.723000	171.947000	195.271000	263.809000

e)

Fig. 7. Boxplots and superimposed swarm plots of volumes of different habitats in the dataset. (**a**) Edema habitat (**b**) Enhancing tumor habitat (**c**) Necrosis habitat. Descriptive statistics of the volumes of (**d**) the MGMT unmethylated and (**e**) the MGMT methylated tumor habitat. neg: Unmethylated group pos: Methylated group necr: Necrosis enh: Enhancing tumor.

were studied with the radiomics method, and 0.88 accuracy and 0.89 AUC were obtained[17]. However, a total of 87 patients were used in the training cohort of this study, and the test cohort included just 35 samples. While 70% of the training set and 80% of the test set were MGMT methylated, only 7 patients were unmethylated in the test set, which represented a severe class imbalance. Even a dummy model that calls all the samples methylated would have an accuracy of 80% in such a scenario. The very few unmethylated cases in the test set, which reduced statistical power, complicated the interpretation greatly. Our dataset represented a more appropriate scenario, both as being a larger dataset with 577 samples and in a balanced distribution of the two groups.

	Precision	Recall	Accuracy	F1 Score
T1W	0.64 (0.60–68)	0.70 (0.66–0.73)	0.65 (0.61–0.68)	0.67 (0.63–0.70)
T1CE	0.80 (0.76–0.83)	0.72 (0.68–0.75)	0.77 (0.73–0.80)	0.76 (0.72–0.79)
T2W	0.68 (0.64–0.71)	0.80 (0.76–0.83)	0.71 (0.67–0.74)	0.73 (0.69–0.76)
FLAIR	0.83 (0.79–0.86)	0.80 (0.76–0.83)	0.82 (0.78–0.85)	0.82 (0.78–0.85)
FLAIR + T1CE	0.87 (0.84–0.89)	0.90 (0.87–0.92)	0.88 (0.85–0.90)	0.88 (0.85–0.90)
Four sequences	0.88 (0.85–0.90)	0.73 (0.69–0.76)	0.81 (0.77–0.84)	0.80 (0.76–0.83)

Table 1. Performance metrics of the models. CI 95% is given in parenthesis.

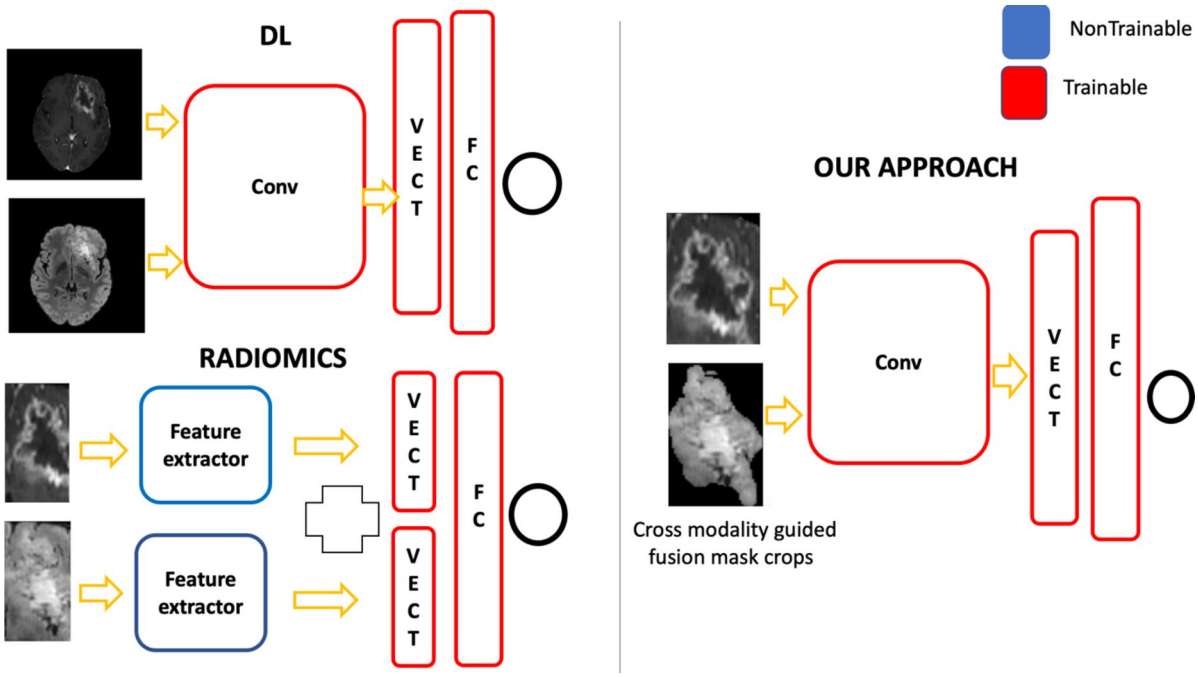


Fig. 8. Key differences between our approach and existing literature. We used region of interest crops, which were obtained by a novel mask fusion strategy, keeping 3D context and multiparametric information of all sequences and trained the network jointly. In standard DL applications, the whole image is given as the input to the models, while in radiomics studies, the usual input is the ROI of the lesion, and the feature extractors are fixed descriptors that are non-trainable.

In the literature, there are various attempts to link imaging phenotypes to the MGMT methylation status in MRI[20–22]. In these studies, the results are conflicting. However, there is a tendency towards specific patterns. The cellularity, vascularity, and proliferation of the MGMT-methylated glial tumors tend to be higher than the unmethylated counterparts. The imaging reflections of these biological behavior differences are reported as follows: MGMT-methylated tumors tend to exhibit less edema, more necrosis, and more enhancement. The enhancement pattern tends to be heterogeneous nodular, which in turn leads to a more heterogeneous texture[20,22]. The enhancement pattern of MGMT unmethylated tumors is typically ring-like with wider peripheral edema. The necrosis tends to be less in MGMT unmethylated tumors. T2-FLAIR mismatch sign, which describes the partial suppression of very high T2 signal of tumor in the FLAIR images, tends to be more prevalent in MGMT unmethylated tumors, which may be attributed to fluid accumulated in the expanded extracellular spaces and/or microcystic changes[21]. When we analyzed the true predictions of our model, we observed more enhancement coupled with less peritumoral edema in MGMT-methylated tumors. (Fig. 11). The enhancement pattern was more heterogeneous nodular, and the overall tumor core size tended to be greater with more heterogeneous texture and less necrosis. Particularly, the necrosis/tumor core ratio is less in the true predictions of the MGMT-methylated group. On the other hand, false positive predictions also reflected this trend. Many MGMT unmethylated tumors that were predicted as methylated by the model exhibited little peritumoral edema and a large tumor core coupled with less necrosis relative to the tumor core (Fig. 11). The typical pattern with the

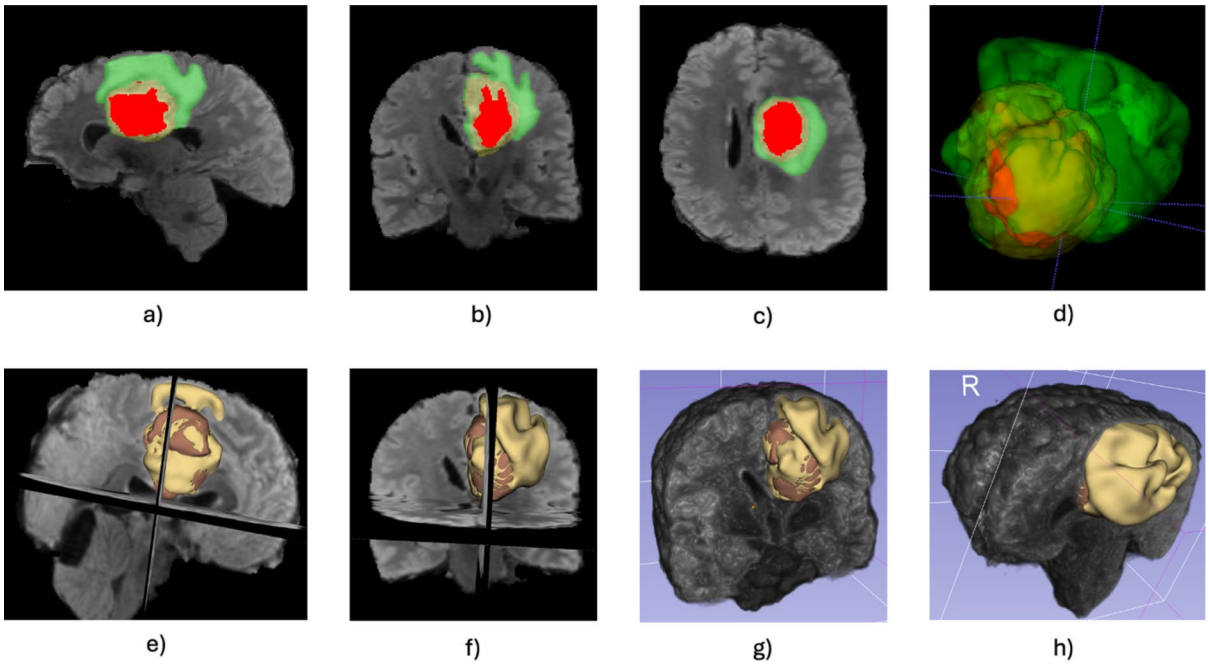


Fig. 9. The topographical relationship of the tumor habitats and the surrounding normal anatomy. (**a**) Sagittal (**b**) Coronal (**c**) Axial slices (**d**) Isolated tumor mesh with different habitats. Tumor mesh superimposed to (**e**) Sagittal oblique (**f**) Coronal oblique projection on multi-planar reconstruction. Tumor mesh superimposed to (**g**) Coronal oblique (**h**) Sagittal oblique volume rendering.

MGMT unmethylated tumors, which were correctly predicted by the model, was a ring-enhancing tumor with more peritumoral edema and more necrosis. The T2/ FLAIR mismatch sign, which was explored by comparing the mean signal ratio of the nonnecrotic tumor volume in T2W and FLAIR sequences between the true positive and true negative groups, was similar in both class predictions. Another observation was the high frequency of hemorrhagic signals in T1W and T2W among mispredictions. Since the segmentation masks did not include a distinct class for hemorrhagic signal, it was not possible to check the relationship between the presence and size of the hemorrhagic signal with false-positive and false-negative predictions. However, those points seem to be an interesting target to explore in future studies. Some ground truth masks also suffered from the hemorrhagic signal. In these samples, the contrast enhancement mask was overestimated or underestimated due to the intrinsic T1W high signal of hemorrhage, which introduced ground truth errors to the dataset (Fig. 12).

When we restrict our analysis to DL studies, we encounter a few successful DL studies. In one study[7], only T2 sequences were used, and the entire image was given to the model for classification. In addition, a separate test set was not used, and the results were evaluated over the images with the cross-validation method. Furthermore, the authors employed 3D U-Net and prepared two mask labels. All the pixels in the MGMT methylated tumor region were labeled as one class, and all the pixels in the MGMT unmethylated tumor region were labeled as another class. Using this dense pixel-wise classification approach and determining the final label of the image by majority voting increased their accuracy. However, the classification of each pixel as MGMT methylated or unmethylated should be questioned since the overall interaction of all the regional pixels can barely represent the exact class, not each individual pixel. These biases might lead to overly optimistic results. In this study, which was conducted with 247 patients, a sensitivity of 0.96 and a specificity of 0.91 were achieved[7]. In another study, authors used an ROI approach with four critical sequences similar to ours. They conducted their research with 259 patients[8]. Furthermore, they used segmentation maps of each sequence and down-sampled the resulting crops to 32×32. Then, they combined corresponding images of each sequence channel-wise and constructed $32 \times 32 \times 4$ size input. Regarding the large size of GBM, which generally spans 1/3 to 2/3 of a hemisphere, the 32×32 size matrix seems too small to retain important details in the image. Moreover, channel-wise concatenating the individual images of different sequences may lead to losing important information even after the first convolution. 3D convolution filters will compress the channel information into 2D. Therefore, the high signal of enhancing tumor in the T1CE channel and of edema in the FLAIR channel will dominate over other channels. Furthermore, the analysis was conducted slice-wise, resulting in the loss of 3D context. They reached 0.83 mean accuracy in their 5-fold CV experiments. Another DL study was a single-center study with 155 patients[9]. The authors used just T2W images with ResNet50 and demonstrated an accuracy of 94.9% on a test set. However, they employed slice-wise classification with an additional normal class, and 78% of the test set were normal slices.

In another study, the researchers studied the prediction of the MGMT methylation status of the glial tumors in brain MRI extensively[23]. Actually, their extensive work was based on permutations of different model architectures coupled with different MRI sequences as input. They employed three CNN models (Res_Net, Dense_Net, and Efficient_Net) with three different architectures coupled with either pretraining or direct training.

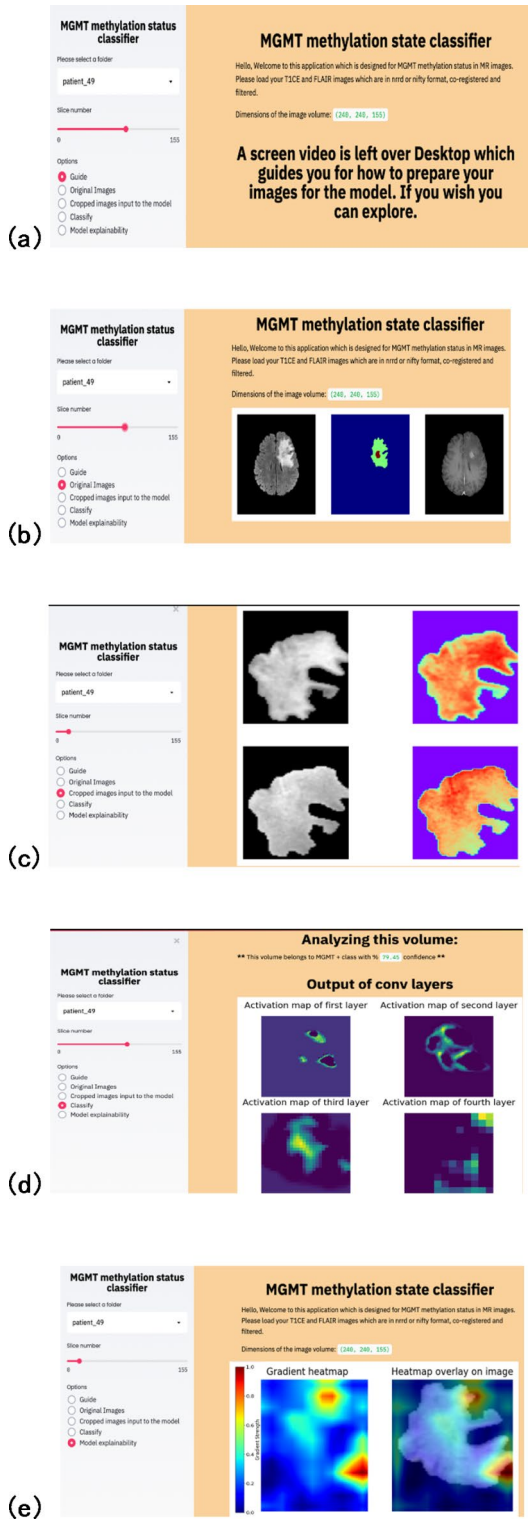


Fig. 10. **(a-e)** We developed a Streamlit-based application for experimental use. The application includes 5 modules. **(a)** First module helps the user prepare the images for the model. **(b)** The second module helps the user freely scroll around T1CE and FLAIR images and the image masks to grasp the overall extent of the pathology and morphological features. **(c)** The third module helps the user freely scroll around the image crops obtained by a novel fusion mask, which will be given into the model as input. To highlight subtle differences, color-coded versions of the images are accompanied by the grayscale crops. **(d)** The fourth module classifies the image volume along with its confidence score and displays feature maps of each of the four convolutional layers. **(e)** The last module overlays GradCam maps, which highlight the pixel importance based on gradient strength on input images. It may help the doctor to check the subregions that the model thought were more important for the classification task. This module may highlight the subtle differences in the image leading to correct diagnosis and may train the physician to look for those clues.

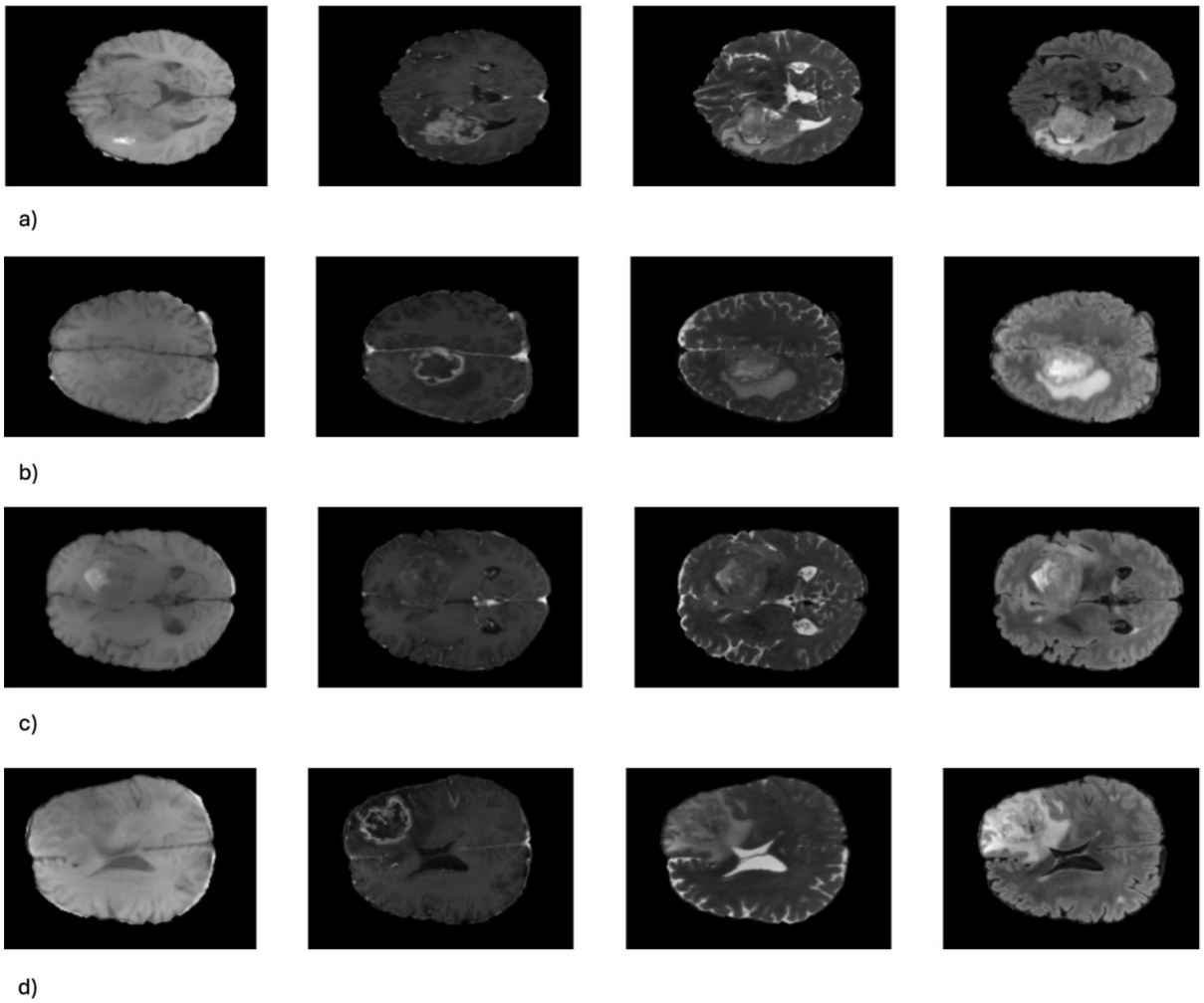


Fig. 11. Sample predictions of the model. (**a**) True positive, MGMT methylated tumors (**b**) True negative, MGMT unmethylated tumors (**c**) False positive predictions that are actually unmethylated (**d**) False negative predictions that are actually methylated.

Additionally, they employed transformer models. They permuted these models with mainly single sequence input and reported the performance metrics. The performance metrics were poor, and the authors concluded that the MGMT methylation status of glial tumors could not be predicted using the BRATS 2021 dataset and the state-of-the-art models. They also mentioned that, with optimum datasets and novel methodologies, the MGMT methylation status could still be predicted.

The dataset footprint and the modifications based on that footprint might alter the performance of the deep learning models more than the used architectures and the training schemes. Actually, one of the most successful segmentation approaches, no new U-Net (nnUnet) was based on this idea[24]. The modifications based on the dataset specifications were more important than the selected model. To date, the input to the models was either a whole slice or a crop of the slice guided by the segmentation masks. The inputs in the ROI approach were still contaminated by the surrounding anatomy. Additionally, the crops were based on the segmentation maps, which only considered the signal change in the corresponding sequence. For instance, FLAIR images were cropped by using the masks which consider the signal change in FLAIR image only. However, some habitats might be imperceivable in some sequences. This insensitivity of a particular sequence to all habitats does not change the fact that there is pathology in that region. To the best of our knowledge, no study used the guidance of a mask to crop a region in a particular sequence, which was obtained using the information of another sequence. Furthermore, to the best of our knowledge, no deep learning study used the masked-out tumor region, which purely consisted of pathology harboring pixels instead of ROI crops based on the maximum diameters of the segmentation mask.

The built Streamlit application was also an important output of our study. We developed an easy-to-use pipeline for this application. The user can export the image volume to the desktop and drag and drop the exported volume to the application to use the features of the application freely. Deployment in a Docker container locally requires no additional cost to the organization and no additional training for the radiologists. The simple user interface with one button for each function makes the usage straightforward immediately after deployment.

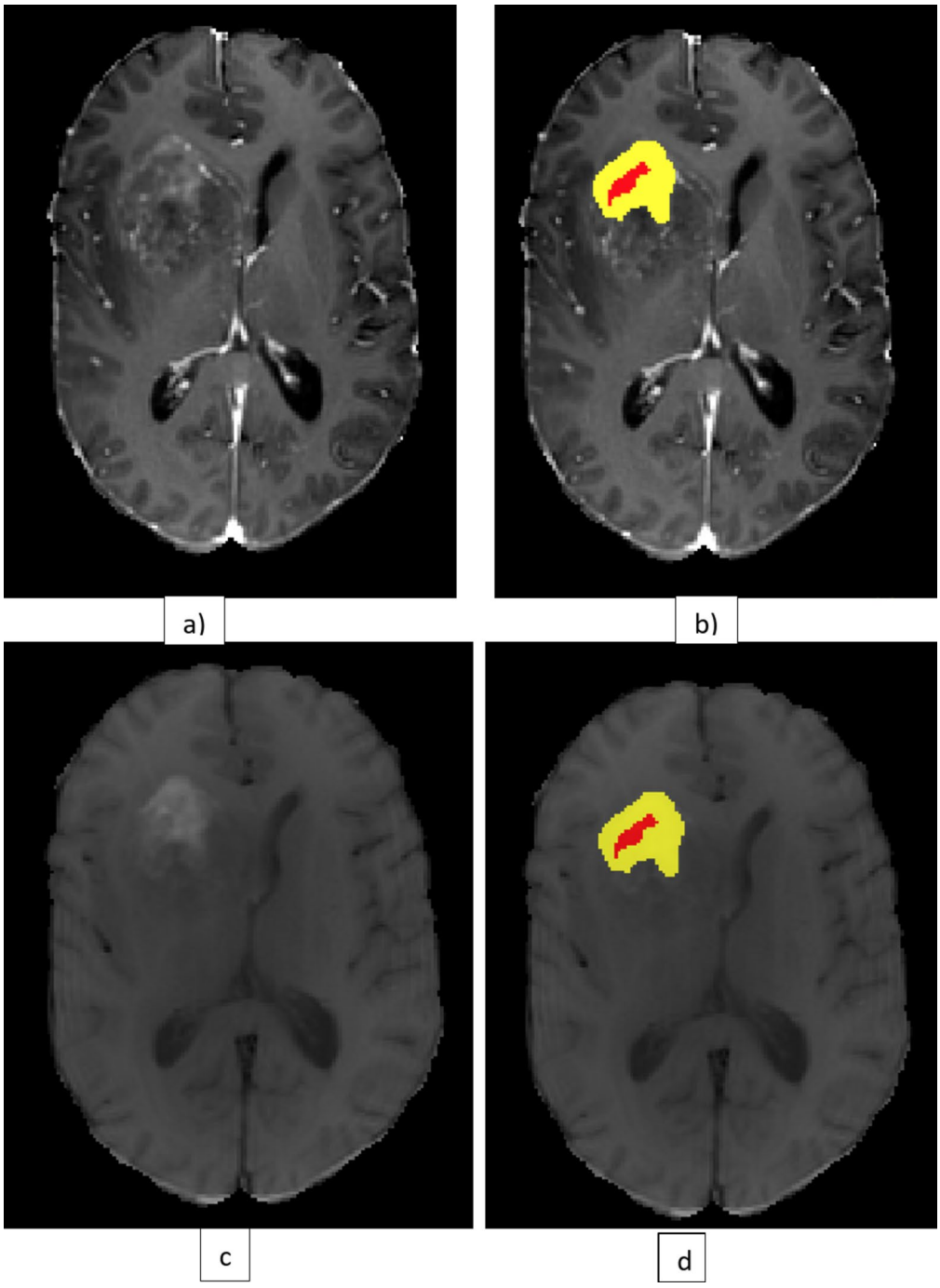


Fig. 12. Effect of hemorrhage on the ground truth masks and the predictions. (**a**) T1CE (**b**) T1CE with superimposed mask (**c**) T1W d) T1W with superimposed mask. The hyperintense signal of hemorrhage interfered with the annotation process and led to habitat uncertainties in the masks. The exact enhancing and necrosis regions are unclear. The mask seems to underestimate both habitats.

For sure, scaling this application to more organizations requires different deployment strategies and different security precautions. However, for the experimental use, deployment in the local environment of the user, which is protected by the user password, and which is also connected to PACS (Picture archiving and communication system), RIS (Radiology information system), and HIS (Hospital information system) systems of the organization provide an easier transition. The application is not a complete medical device. The main objective is to collect more information by checking its performance on the new images and molecular analysis in the environments of the local organizations. We believe this strategy should be adopted for all promising new models published in the literature, particularly for the ones applied to rare diseases. Many of the built models remain experimental forever due to the lack of connection between the development environment and the potential users. An easy-

to-use pipeline that involves just locally exporting the image and putting it into the application might take under a few minutes. This approach has two benefits. First, the radiologists can become more familiar with the AI tools and the outputs of these tools that were built by their team at no additional cost. Second, the local prospective validation of the models may help in selecting the more successful candidates to further work on.

The primary assumption of the machine learning models is the similarity of the training and the test datasets. Therefore, dataset collection should be carried out extensively and carefully considering the distribution of the problem in the real world. The size, diversity, and constitution of the dataset have utmost importance. Although convolution and max pooling operations, along with input and batch normalization, introduce some robustness to the models, the output is still conditioned on the pixel distribution of the input. Different MRI machines, different acquisition protocols, and different user preferences affect the real pixel distribution of the images. In order to combat the overfitting of the model to the training data, a test check on external data should be carried out along with internal testing. BRATS 2021 dataset was collected considering this requirement[25]. The multi-institutional nature of this dataset satisfies the external data requirement for the models trained with local data.

There are some limitations of our study. One of them was the lack of important demographic and clinical information such as age, gender, and clinical features of the patients since it was carried out on a public data set. Therefore, it is not possible to make a clear statement about whether these factors affect our comparison. On the other hand, due to the use of a preprocessed image set, all images had a common size of 240*240*155, and the voxel dimensions were fixed as 1*1*1* mm. Therefore, it has not been possible to carry out some quality control tests, such as showing how the accuracy changes by pooling the images according to their quality and resolution. The pixel distribution of an image acquired with a medium resolution protocol at 1.5 T field strength and an image obtained with higher resolution at 3 T field strength can be quite different from each other. On the other hand, in our study, although the information of all sequences was used, only the regions that had a signal change in at least one sequence were used. However, the displacement and compression patterns of the surrounding structures can also provide additional information about the nature of the tumor. Another limitation was the number of samples in the dataset. Although dataset size is one of the largest for this problem, 577 patients may still not be enough for building robust DL applications.

Conclusıon

In conclusion, we achieved to determine the MGMT methylation status of GBM in MR images with high performance by using one of the largest datasets in the literature, multiple-sequences guided masks, 3D context kept, noise-reduced tensor patches, and deep learning. We believe the different biological compositions driven by the underlying genomic profile of the glial tumors exhibit different imaging phenotypes. However, the imaging phenotypes of the glial tumors are not solely determined by MGMT methylation status. This may explain the controversial outcomes of research studies. Larger datasets that are stratified according to more detailed, comprehensive genetic labels, including other important genomic signatures such as IDH, EGFR, 1p19q codeletion, CDKN2A/B, TERT, and ATRX may provide better opportunities to build more comprehensive and robust classifiers.

Data availability

The anonymised data collected are available as open data via the Kaggle platform repository: https://www.rsna.org/rsnai/ai-image-challenge/brain-tumor-ai-challenge-2021. Example from: https://www.kaggle.com/competitions/rsna-miccai-brain-tumor-radiogenomic-classification/data.

Received: 7 July 2024; Accepted: 22 January 2025
Published online: 25 January 2025

References

1. WHO, Statistics on Brain Cancer. July 12. (2020). http://www.who.int/cancer/en
2. Michael, W. et al. MGMT promoter methylation in malignant gliomas: ready for personalized medicine? *Nat. Reviews Neurol.* **6** (1), 39–51 (2010).
3. Baid et al. The RSNA-ASNR-MICCAI BraTS 2021 Benchmark on Brain Tumor Segmentation and Radiogenomic classification. arXiv:2107.02314 [cs.CV].
4. https://turkradyolojiseminerleri.org/content/files/sayilar/11/buyuk/20-36.pdf
5. Van Kempen, E. J. et al. Accuracy of machine learning algorithms for the Classification of Molecular Features of Gliomas on MRI: a systematic literature review and Meta-analysis. *Cancers* **13**, 2606. https://doi.org/10.3390/cancers13112606 (2021).
6. Han, L. & Kamdar, M. R. MRI to MGMT: Predicting methylation status in glioblastoma patients using convolutional recurrent neural networks. *Pac. Symp. Biocomput.* **23**, 331–342 (2018).
7. Yogananda, C. G. B. et al. MRI-Based deep-learning method for determining glioma MGMT promoter methylation Status AJNR. *Am. J. Neuroradiol.* **42**:845–852 May 2021.
8. Chang, P. et al. Deep-learning convolutional neural networks accurately classify genetic mutations in gliomas. *Am. J. Neuroradiol.* **39** (7), 1201–1207 (2018).
9. Panagiotis, K. et al. Residual deep convolutional neural network predicts MGMT methylation status. *J. Digit. Imaging.* **30** (5), 622–628 (2017).
10. Sasaki, T. et al. Radiomics and MGMT promoter methylation for prognostication of newly diagnosed glioblastoma. *Sci. Rep.* **9**, 14435 (2019).
11. Menze, B. et al. The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE Trans. Med. Imaging.* **34** (10), 1993–2022 (2014).
12. Medikal Görüntü İşlemede Derin Öğrenme Ayşe Gül Eker, Nevcihan Duru Acta Infologica, Volume 5, Number 2, (2021).
13. Chollet, F. Python ile Derin Öğrenme, Buzdağı Yayınevi, (2019).
14. https://keras.io/examples/vision
15. Larraitz Egañ et al et al. Methylation of MGMT promoter does not predict response to temozolomide in patients with glioblastoma in Donostia hospital. *Sci. Rep.* **10** (1), 1–11 (2020).

16. Liu, D. et al. Imaging Genomics in Glioblastoma: Combining Molecular and Imaging Signatures.Frontiers in Oncology July 2021Vol 11: 699265.
17. Jiang, C. et al. Fusion Radiomics features from conventional MRI predict MGMT promoter methylation European. *J. Radiol. Volume*. **121**, 108714 (2019).
18. Rathore, S. et al. Non-invasive determination of the O6-methylguanine-DNA-methyltransferase (MGMT) promoter methylation status in glioblastoma (GBM) using magnetic resonance imaging (MRI). *J. Clin. Oncol. Conf.* **36**, 2051 (2018).
19. Shboul, Z. A. et al. Prediction of molecular mutations in diffuse low-Grade Gliomas using MR Imaging features. *Sci. Rep.* **10**, 3711 (2020).
20. Suh, C. H., Kim, H. S., Jung, S. C., Choi, C. G. & Kim, S. J. Clinically relevant Imaging features for *MGMT* promoter methylation in multiple glioblastoma studies: a systematic review and Meta-analysis. *AJNR Am. J. Neuroradiol.* **8**, 1439–1445 (2018).
21. Eoli, M. et al. Methylation of O6-methylguanine DNA methyltransferase and loss of heterozygosity on 19q and/or 17p are overlapping features of secondary glioblastomas with prolonged survival. *Clin. Cancer Res.* **13**, 2606 (2007). – 13.
22. Yamashita, S. et al. T2-Fluid-attenuated Inversion Recovery Mismatch sign in Lower Grade Gliomas: correlation with pathological and molecular findings. *Brain Tumor Pathol.* **39** (2), 88–98 (2022).
23. Saeed, N., Ridzuan, M., Alasmawi, H., Sobirov, I. & Yaqub, M. MGMT promoter methylation status prediction using MRI scans? An extensive experimental evaluation of deep learning models. *Med. Image Anal.* **90**, 102989 (2023).
24. Isensee, F. et al. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat. Methods*. **18**, 2021203–2021211 (2021).
25. Bakas, S. et al. Advancing the Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. *Nat. Sci. Data*. **4**, 170117 (2017).

Author contributions

Author Contributions Statement"I.O.K. and C.K. designed the study concept, C.K. collected data and prepared for the analysis, I.O.K and C.K write the AI code and inference, I.O.K. wrote the main manuscript text and C.K. prepared figures. All authors reviewed the manuscript."

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to İ.Ö.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.