



HAL
open science

Revolutionizing Brain Cancer Diagnosis: Automated Prediction of MGMT Methylation Status using Histological Images

Manel Mili, Asma Ben Abdallah, Jose Javier Otero, Asma Kerkeni, Mohamed Hedi Bedoui

► **To cite this version:**

Manel Mili, Asma Ben Abdallah, Jose Javier Otero, Asma Kerkeni, Mohamed Hedi Bedoui. Revolutionizing Brain Cancer Diagnosis: Automated Prediction of MGMT Methylation Status using Histological Images. 2023 International Conference on Cyberworlds (CW), Oct 2023, Sousse, Tunisia. pp.149-156, 10.1109/cw58918.2023.00030 . hal-04361460

HAL Id: hal-04361460

<https://hal.science/hal-04361460v1>

Submitted on 26 Dec 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Revolutionizing Brain Cancer Diagnosis: Automated Prediction of MGMT Methylation Status using Histological Images

Manel Mili
Medical Imaging and Technology Lab
Faculty of Sciences of Monastir
Monastir, Tunisia
manel.mili@isimm.u-monastir.tn
0000-0003-3892-8579

Asma Ben Abdallah
Medical Imaging and Technology Lab
ISIMM
Monastir, Tunisia
asma.benabdallah@isimm.u-monastir.tn

Jose Javier Otero
College of Medicine
University of Ohio State
Columbus, USA
jose.otero@osumc.edu

Asma Kerkeni
Medical Imaging and Technology Lab
ISIMM
Monastir, Tunisia
asma.kerkeni@isimm.u-monastir.tn
0000-0002-3134-3802

Mohamed Hedi Bedoui
Medical Imaging and Technology Lab
Faculty of Medicine
Monastir, Tunisia
medhedi.bedoui@fmm.rnu.tn
0000-0003-4846-1722

Abstract—Advancements in deep learning algorithms for medical imaging, combined with the integration of cyberworlds, have shown great promise in providing precise diagnostic results. One area of interest is the application of these advancements in enhancing personalized treatment of gliomas, a particularly challenging type of brain tumor, by providing more reliable information on a clinical biomarker Oxygen 6-methylguanine-DNA methyltransferase (MGMT). To achieve accurate results, a MobileNetV2 model was employed, utilizing transfer learning method and a mechanism spatial attention with correlation was added to further enhance the model's performance. The model was trained using a private dataset of annotated images and evaluated using cross-validation. Results showed high precision and recall in predicting MGMT status, indicating its potential to improve the efficiency of this prediction. The model's ability to predict MGMT promoter methylation status can help clinicians make more informed decisions, which has the potential to improve personalized treatment planning, ultimately leading to better patient outcomes.

Index Terms—Medical Imaging, Gliomas, MGMT Promoter Methylation, Deep Learning, Attention Mechanism.

I. INTRODUCTION

The integration of cyberworlds, along with the advancements in deep learning algorithms for medical imaging, presents a transformative opportunity to enhance personalized treatment. Deep learning algorithms have brought about a significant transformation in the field of medical imaging by enabling highly precise and dependable analysis for various medical tasks [1]–[3]. Cyberworlds, on the other hand, offer virtual environments where clinicians can interact with digital content and access reliable information. This integration allows for more informed decisions in personalized treatment planning,

potentially improving the efficiency of prediction models and ultimately leading to better patient outcomes. One such task is classifying molecular biomarkers, which are essential in predicting the prognosis and treatment response of tumors, according to the 2021 WHO classification [4]; These molecular biomarkers can provide valuable information on the biological characteristics of tumors, such as their genetic makeup, cellular structure, and signaling pathways [5], [6]. Oxygen 6-methylguanine-DNA methyltransferase (MGMT) is a significant prognostic factor in glioma patients [7], [8], a particularly challenging type of brain tumor; One of the major challenges in treating these tumors is that they are known to grow rapidly and invade surrounding brain tissue, making them difficult to completely remove through surgery. Furthermore, the issue of accurately predicting a patient's response to chemotherapy and radiation treatments remains a significant challenge. Clinical biomarker MGMT can assist with this prediction, and its activity is regulated by a process called promoter methylation, which can affect the response to chemotherapy; When the MGMT promoter is methylated, the expression of the MGMT protein is reduced, which can increase the sensitivity of gliomas to chemotherapy with alkylating agents. Conversely, when the MGMT promoter is unmethylated, the expression of MGMT is higher, which can reduce the effectiveness of chemotherapy with alkylating agents. Therefore, determining the MGMT status is crucial for diagnosing and planning treatment for glioma patients. MGMT methylation has been consistently associated with the development of glioblastoma (GBM) and WHO grade III glioma [8]. Additionally, recent studies have revealed an association between MGMT methylation and WHO grade II

glioma [7], [9], [10], highlighting its importance across various glioma grades.

Despite the extensive research on deep learning methods for cancer diagnosis and treatment [11], predicting MGMT promoter methylation using these techniques remains a relatively new area of investigation. To advance this field, our work leverages transfer learning and attention mechanisms to develop an automated supervised model for predicting MGMT promoter methylation from histological images of gliomas. Transfer learning allows us to benefit from pre-trained deep learning models that have learned rich features from large datasets, enabling us to effectively extract relevant information from the histological images. Furthermore, we employ spatial attention mechanism with correlation to focus the model's attention on informative regions within the histological images. Combining transfer learning and attention mechanism, we develop a robust and reliable model that contributes to the advancement of glioma diagnosis and treatment. Utilizing these techniques in our research facilitates the accurate prediction of MGMT promoter methylation status, ultimately leading to improved patient outcomes and enhanced quality of life for individuals with gliomas.

Our work is structured as follows: (i) Firstly, the current literature on the use of deep learning algorithms for predicting MGMT status in gliomas is reviewed. (ii) Secondly, the histological images are preprocessed to ensure consistency and optimize the performance of the deep learning model. This involves a series of steps, including resizing, histogram equalization, contrast stretching, and CDF mapping, to ensure that the images are of uniform size and quality, and that the model is robust to variations in the images. (iii) Thirdly, a deep learning model is developed using transfer learning and attention mechanisms, and trained on the dataset. (iv) Finally, the model's performance is evaluated using various metrics and compared with existing methods for predicting MGMT status.

II. LITERATURE REVIEW

The rapid development of artificial intelligence approaches including machine learning and deep learning has favored the automatic extraction of important tumor markers, from medical imaging, for the clinical treatment planning and post-treatment monitoring for patients with gliomas [12]. Based on these methods, researchers have shown a keen interest in predicting the expression of biomarkers, specifically the methylation status of the MGMT. In their study [13], Korfiatis et al. employed a deep convolutional neural network (DCNN) to predict the methylation status of the MGMT gene for patients with glioblastoma multiforme (GBM). To improve the flow of information through the network and address the issue of vanishing gradients, they used a residual architecture known as ResNet50. The dataset used in the study was collected from Mayo Clinic between January 1, 2007, and December 31, 2015, and included a total of 545 patients. Out of these patients, only 155 MRI examinations were used for the analysis, with 66 tumors being methylated and 89

being unmethylated. The ResNet50 model achieved impressive results, with a test set f1-score of 97%, precision of 97%, and recall of 97%. However, it is worth noting that a major limitation of this approach is the significant amount of time required for training the model, which can range from 3 to 10 hours. This may make it challenging to apply this model as a transfer learning approach for predicting clinical biomarkers in other datasets, especially those with limited computational resources. Therefore, further research is needed to optimize the training process and reduce the computational burden of the model to make it more practical for widespread clinical use. In [14], Han et al. aimed to predict the methylation status of MGMT in glioblastoma patients by utilizing convolutional recurrent neural networks (CRNN). Used MRI imaging dataset were collected from The Cancer Imaging Archive (TCIA) combined with methylation data from The Cancer Genome Atlas (TCGA), a total of 159 patients. The CRNN achieved a modest test set accuracy of 62%, with a precision of 67% and recall of 67%. Chang et al. conducted a study, as described in their research paper [15], to classify genetic variations of diffuse gliomas using convolutional neural network (CNN) with residual connections. The researchers retrospectively obtained MRI imaging data and molecular information from The Cancer Imaging Archives for 259 patients diagnosed with either low- or high-grade gliomas. The CNN was trained independently to classify isocitrate dehydrogenase 1 (IDH1) mutation status, short arm of chromosome 1 (1p) and long arm of chromosome 19 (1p/19q) codeletion, and MGMT promoter methylation status. The results of the study showed that the CNN was able to accurately classify genetic mutations in gliomas with a high level of accuracy, achieving 94% for IDH1, 92% for 1p/19q, and 83% for MGMT. In [16], Chen et al. explore the feasibility of integrating tumor segmentation and MGMT methylation status prediction of GBM patients into a deep learning pipeline. They used MRI imaging dataset collected from TCIA in combination with clinical and molecular data from TCGA, which included a total of 106 GBM patients. For tumor segmentation, they adapted one state-of-the-art model [17] from the BraTS 2018 challenge. Meanwhile, for the classification of MGMT methylation status, they designed a 4-layer CNN. Furthermore, they cascaded the classification model with the tumor segmentation model. The experimental results showed promising performance in accurate glioma segmentation, achieving a Dice score of 89.7%, as well as accurate MGMT status prediction, with an accuracy of 82.7%, recall of 85.2%, precision of 82.1%, and f1-score of 83.6%. One of the major limitations of this paper is that the model was exclusively tested on patients diagnosed with GBM. This lack of diversity in the patient population under evaluation precludes us from determining the model's effectiveness in classifying molecular biomarkers for other grades of glioma. Therefore, it is essential to conduct further research to evaluate the model's performance on a broader range of cancer grades in order to ascertain its true potential and applicability in clinical settings. In [18], Yogananda et al. presented a deep learning network, MGMT-net, for determining the methylation

status of the MGMT promoter. The MGMT-net architecture consists of a 3D-dense-UNet. The algorithm was trained and tested using MRI imaging data from TCGA and TCIA, a total of 247 patients diagnosed with either low- or high-grade gliomas. The MGMT-net achieved an accuracy of 94.73% with a sensitivity and specificity of 96.31% and 91.66%, respectively. Despite achieving promising results, the authors highlight two significant limitations of the approach. Firstly, the variability in acquisition parameters and imaging vendor platforms may impede the accuracy of the results. Secondly, the largest connected component step used to minimize false-positives in the current classification approach may not be suitable for multi-focal tumors. Chen et al. [19] sought to predict MGMT methylation status using residual network (ResNet18) and MRI imaging data obtained from the Affiliated Drum Tower Hospital of Nanjing University Medical School between 2018 and 2020. The study involved a total of 111 patients diagnosed with low- or high-grade gliomas. To input data into the model, the authors opted to extract radiomics features from a selected region of interest (ROI) of MRI images, rather than using the entire MRI images. This technique has the benefit of eliminating non-discriminatory features. The model successfully predicted MGMT promoter methylation status in patients with diffuse gliomas with a high level of accuracy and f1-score, achieving a score of 91% and of 90%, respectively. Although CNNs and their variants have shown success in the discussed context [13], [15]–[19], they have solely relied on radiomics. Medical images can be affected by various factors, such as imaging parameters, patient motion, and image artifacts, which can limit their reliability [18]. Furthermore, it has been confirmed in [20], [21] that current deep learning models are unable to accurately predict MGMT Promoter state using only patient MRI scans. In light of the restrictions and constraints present, our research endeavors will be directed towards exploring the field of pathomics with the objective of enhancing the accuracy and effectiveness of predicting genetic mutations in gliomas; Unlike medical images, pathomics involves analyzing physical tissue samples, which may offer greater consistency and reproducibility. Therefore, we aim to explore the potential of pathomics in enhancing the accuracy and reliability of CNN-based models for medical image analysis and diagnosis.

III. MATERIALS AND METHODS

A. Dataset and Preprocessing

1) *Dataset*: The dataset utilized in this study was procured from the Ohio State Wexner Medical Center ¹, encompassing a total of 1075 biopsy images of brain tumors that varied in size and magnification levels (40x, 100x) from 107 patients. The tumors were of two distinct types, namely Astrocytoma and Oligodendroglioma, with grades ranging from II to III, as depicted in Table I.

¹<https://wexnermedical.osu.edu>

TABLE I: Number of patients per tumor and grade.

| Tumor | Grade II | Grade III |
|-------------------|----------|-----------|
| Astrocytoma | 40 | 12 |
| Oligodendroglioma | 32 | 23 |

Among the 107 histological images scrutinized, 87 patients exhibited tumors with methylation, while the remaining 20 patients had tumors without methylation as plotted in Fig. 1.

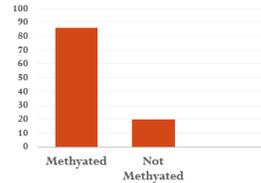


Fig. 1: Distribution of tumors with and without methylation in the dataset. 87 patients exhibited tumors with methylation, while the remaining 20 patients had tumors without methylation.

To avert overfitting to the majority class, we counterbalanced the dataset by handpicking 20 patients with tumor methylation and 20 patients with no tumor methylation for our training, validation, and testing sets. It is worth noting that the images of size 100x were considered in this study as they were more numerous than the 40x images.

2) *Training, Validating, and Testing Datasets*: To create training, validating, and testing dataset for this study, the original images were first divided into multiple patches with a fixed size of 512x512. This was done to reduce the computational burden and to ensure that the images could fit into the available memory. We applied two rules for the border of the images: if a patch’s size was less than half of 512, we discarded the image. If a patch’s size was more than half of 512, we enlarged them by selecting additional neighboring patches to create a final image size of 512x512 (see Fig. 2).

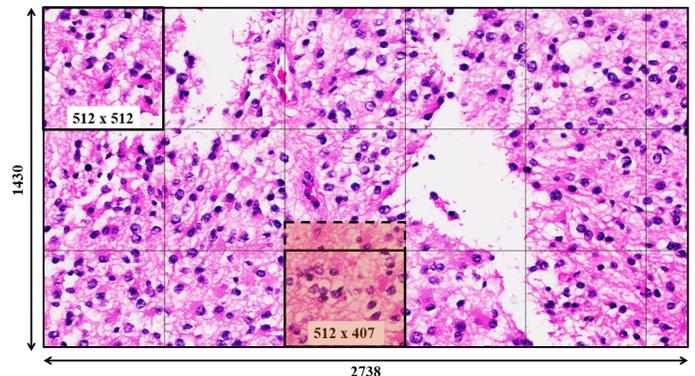


Fig. 2: Original images were divided into patches of 512x512. Patches smaller than half the size of 512 were discarded, while patches larger than half the size were enlarged by incorporating neighboring patches.

It is important to note that all the resulted patches will have the same label target as the original image. This resulted in a total of:

- Training
 - Total images: 1757
 - Methylated: 898
 - Not methylated: 859
- Validation
 - Total images: 883
 - Methylated: 449
 - Not methylated: 434
- Testing
 - Total images: 411

3) *Images Preprocessing*: Images preprocessing and enhancement were performed using OpenCV and NumPy libraries. The preprocessing step applied to the images involved converting them from the RGB color space to the Lab color space. This conversion provided multiple benefits. One advantage was that it separated color information into three distinct channels, simplifying color-based images analysis. Additionally, the Lab color space could represent a broader spectrum of colors than the RGB color space, including highly saturated colors and colors with extreme brightness or darkness. After the conversion of the images into the Lab color space, the next step involved splitting the image into its three channels: Lightness (L), a (green-red), and b (blue-yellow). Following this, histogram equalization was applied to the L channel. This step was important in enhancing the contrast of the image and improving the overall quality of the image. By applying histogram equalization, the distribution of pixel values in the L channel was adjusted to improve the visibility of the image’s features. After histogram equalization was applied to the L channel, the three channels were merged back together to form the equalized Lab image. This image was then converted back to the RGB color space. To further enhance the image quality, the intensity values of the image were stretched using the contrast stretching method. This involved computing the histograms of the blue, green, and red channels of the image and calculating the cumulative distribution function (CDF) of each channel. The code then masked all pixels with a value of zero in the CDF and scaled the remaining pixel values to the range of (0, 255). The resulting values were filled with zero where the mask was applied. The masked and scaled CDFs were then used to create new blue, green, and red channels by mapping the pixel values of the original image to their corresponding CDF values. Finally, these channels were merged back together to form a new image with improved contrast and visibility of features. Fig. 3 shows an example of an image that underwent the preprocessing steps described above. The image on the left is the original image, while the image on the right is the processed image.

B. Methods

1) *Transfer Learning*: The effectiveness of deep learning models largely depends on the availability of extensive labelled

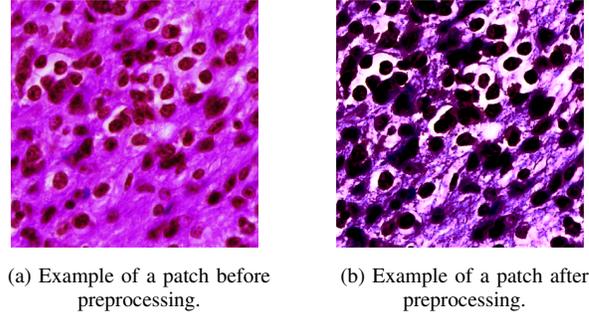


Fig. 3: Resulted image after the application of the preprocessing phase: the processed image has improved contrast and better visibility of the features present in the image. This result demonstrates the effectiveness of the preprocessing steps in enhancing the quality of the images.

datasets for training. Unfortunately, our dataset fell short in this regard, posing a significant challenge to our research. To tackle this issue, we initially resorted to data augmentation techniques to artificially increase the size of our dataset [22]. However, the sensitive nature of histological images proved to be a bottleneck, as this approach failed to yield significant improvements in our results.

In light of this, we decided to pivot to a transfer learning approach [23]. In transfer learning, a pre-trained model’s weights and architecture are used as a starting point and are then fine-tuned on a smaller dataset for a new task. This approach allows the new model to leverage the pre-existing knowledge and improve its performance on the new task with less data and computational resources. Our proposed approach involves transferring knowledge from the origin domain ImageNet (D_O) to our target domain (D_T) of histological images. Specifically, we aim to train a target classifier h_T that can predict the methylation status of the gene MGMT in two categories, namely methylated (Class 1) and not methylated (Class 0), for a given patch image X_{T_i} from the target domain D_T , as shown in Eq. 1:

$$Y_{T_i} = h_t(X_{T_i}) \quad (1)$$

We employ transfer learning as a feature extractor by freezing all layers in the base model (section III-B2) except for the last one. This is achieved by setting the trainable attribute to False for all layers except the last one. Additionally, we define a regularizer using L2 regularization with a coefficient of 0.001 to add a penalty term to the loss function, encouraging the model to learn simpler weight patterns. To further enhance the model’s performance, we add a new layer on top of the base model’s output. This layer is a DepthwiseConv2D layer with a kernel size of (3, 3), strides of (2, 2), padding set to ‘same’, and activation function set to ‘relu’. The layer performs depthwise separable convolution, which helps reduce the number of parameters and improve computational efficiency. A whole description of the model is detailed in Fig. 5.

2) *MobileNetV2 as Transfer Learning Model*: We chose the MobileNet architecture [24] as our transfer learning model. MobileNet is an architecture comprised of convolutional layers with depthwise separable convolutions. Depthwise separable convolutions involve two separate operations: depthwise convolution and pointwise convolution. The depthwise convolution applies a single filter to each input channel, while the pointwise convolution combines the output using a 1x1 filter. This approach reduces parameters and computations while maintaining accuracy. MobileNet also includes a global average pooling layer, which reduces feature map dimensions, and a fully connected layer for classification. These design choices minimize parameters and computations, making MobileNet efficient for mobile and embedded devices. Specifically, we fine-tuned MobileNetv2 on our dataset, which is an improved version of MobileNet that includes linear bottlenecks, inverted residuals, and shortcut connections. These additions improve the accuracy and efficiency of the model (see Fig. 5).

3) *Spatial Attention Mechanism with Correlation*: Spatial attention is a highly effective approach in deep learning models that enhances the representation of crucial spatial regions in an image. This technique enables the model to selectively concentrate on specific regions of an image by assigning greater importance to the features in those areas. As a result, the model can more accurately capture the most significant information and make more precise predictions. By incorporating the correlation mechanism, we can enhance the effectiveness of the spatial attention technique and improve the overall efficiency of the model. This approach involves capturing the relationships between different patch regions by computing correlation coefficients and using these correlations to guide the attention mechanism. The correlation coefficients are calculated between pairs of feature maps, and capture the degree of similarity between different spatial regions. Let's suppose we have two feature maps, X and Y , the correlation coefficient between X and Y is calculated using Eq. 2:

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} \quad (2)$$

where $cov(X,Y)$ is the covariance between X and Y , and σ_X and σ_Y are the standard deviations of X and Y , respectively. By computing these correlations, the model can identify which regions of the patch are most related to each other, and which regions are most relevant to the task at hand. The attention mechanism then uses these correlations to weight the feature maps, giving more weight to regions that are more relevant to the task using Eq.3.

$$A_i = \sum_{j=1}^n \alpha_{i,j} F_j \quad (3)$$

where A_i is the attention map for patch i , F_j is the feature map for patch j , and $\alpha_{i,j}$ is the attention weight between patches i and j , computed as:

$$\alpha_{i,j} = \frac{\exp(\rho_{i,j})}{\sum_{k=1}^n \exp(\rho_{i,k})} \quad (4)$$

where $\rho_{X,Y}$ is the correlation coefficient between patches i and j , and n is the total number of patches.

This allows the model to focus on the most important regions, while ignoring irrelevant ones. To further enhance the interpretability and performance of our model, we add Gradient-weighted Class Activation Mapping (Grad-CAM) method [25]. Grad-CAM is a type of class activation mapping that highlights the regions of an image that the model is focusing on to make its prediction. It works by computing the gradient of the output class score with respect to the feature maps of the last convolutional layer in the model. This gradient is then used to weight the feature maps, producing a heatmap that shows the regions of the image that are most important for the model's prediction as depicted in Fig. 4.

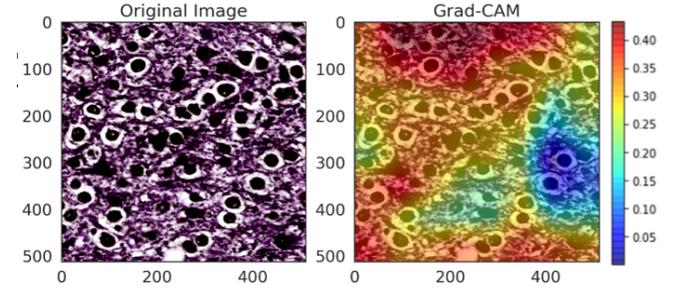


Fig. 4: By using Grad-CAM, we were able to highlight the regions of the input image that the model is focusing on to make its prediction.

By using Grad-CAM to identify the most relevant regions of the input image and then using spatial attention mechanism to selectively weight the feature maps based on their importance, the model can focus more on the important regions of the image and ignore irrelevant regions. It should be emphasized that during the training phase, the model may encounter patches that have been labeled as positive targets but in reality correspond to negative targets or the opposite. To prevent the model from learning to predict false positive targets or true negative targets for these patches, a weight of zero is assigned to them during training. This effectively ignores these patches and prevents the model from being biased towards predicting positive targets or negative targets for them. Additionally, by using Grad-CAM to produce a heatmap that highlights the regions of the image that are most important for the model's prediction, we can gain insight into how the model is making its predictions.

4) *Classification Block*: Several dense layers with dropout regularization are added before the final output layer. Each dense layer applies a rectified linear unit (ReLU) activation function and uses the specified regularizer (L2 regularization with a coefficient of 0.001) to prevent overfitting. The dropout layers with a rate of 0.2 are inserted after each dense layer to randomly set a fraction of input units to 0 at each update during training, which helps to further reduce overfitting. The dense layers gradually decrease the number of units from 1024 to 75, capturing increasingly abstract representations of

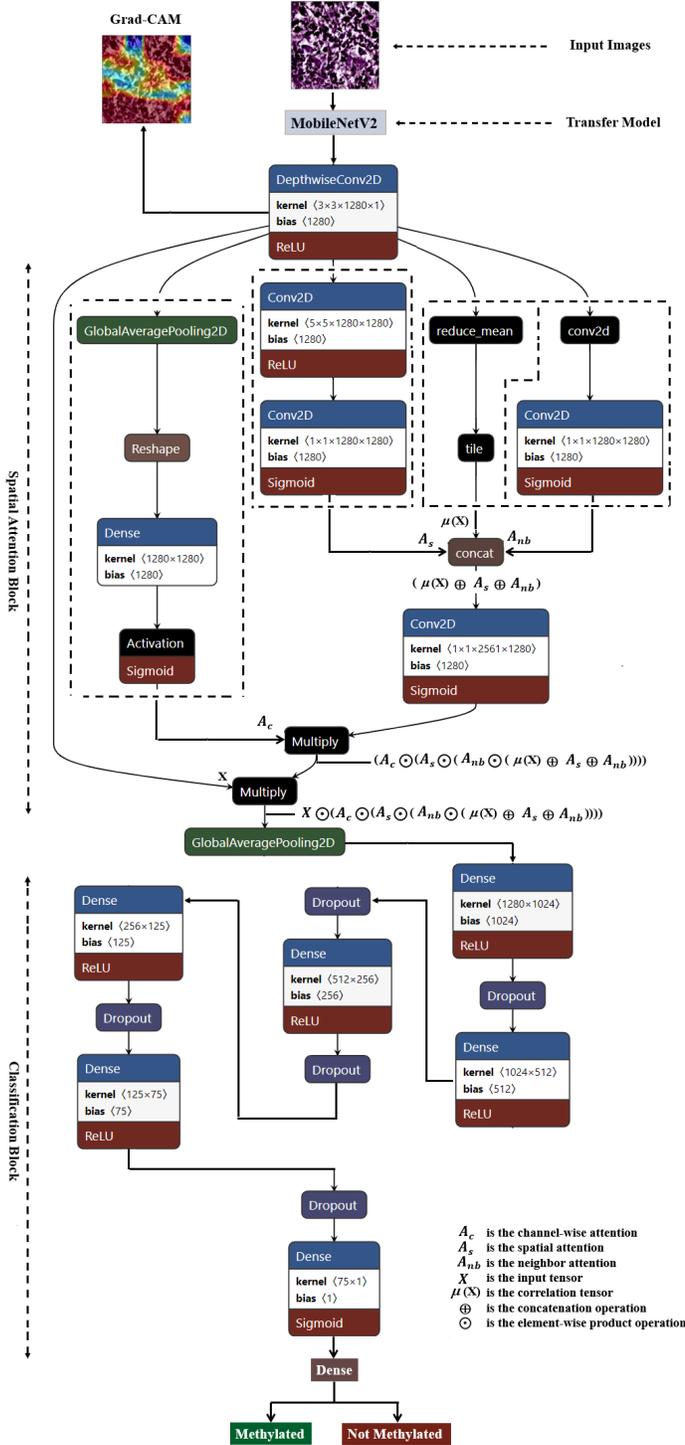


Fig. 5: Visual depiction of the proposed model. Our proposed model comprises of three main blocks that work together to produce the final predicted class probability. Firstly, we leveraged transfer learning by selecting and implementing the popular deep learning architecture, MobileNetV2. This block enabled fast and effective feature extraction from input images.

Fig. 5: Secondly, we added a spatial attention mechanism with correlation that helped the model to focus on critical regions of the input images while ignoring irrelevant regions. Finally, to produce a predicted class probability, we incorporated a classification block that takes the feature maps generated from the transfer learning and spatial attention blocks and outputs the final probability distribution over the classes.

the input features. The ReLU activation function introduces non-linearity to the model, allowing it to learn complex patterns and relationships in the data. The regularizer penalizes the model's weights, encouraging them to be smaller and preventing excessive complexity. Finally, the output layer is added with a single unit and a sigmoid activation function. This layer produces a single output value that represents the predicted class probability, ranging from 0 to 1, indicating the likelihood of belonging to a specific class (see Fig. 5).

5) *Model Training and Assessment*: To compare the performance of our transfer model with spatial attention, we implemented ResNet50 model as alternative approach. ResNet50 is a well-established pre-trained model that has been widely used for image classification tasks. We trained each model for 25 epochs using Adam optimizer with a learning rate of 0.0001. Early-Stopping callback is used to monitor validation loss and terminate training if it does not improve for five consecutive epochs. To prevent overfitting and ensure a fair distribution of the training set, k-fold cross-validation with k=3 was conducted. This approach involved dividing the dataset into three equal parts, with each part serving as a validation set while the remaining two parts were used for training. As a result, each fold generated three different models for the considered model, allowing us to assess the model's performance on different subsets of the data. This method helped to ensure that the model was not biased towards any particular subset of the data, resulting in a more accurate and reliable performance evaluation. The model with the highest performance metrics is then selected as the final model to be tested on our test set.

IV. RESULTS

A. Evaluation Metrics

In this work, we employed conventional measures to assess the effectiveness of our transfer learning approach, including precision (Eq. 5), recall (Eq. 6), and f1-score (Eq. 7) metrics.

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

$$F1 - score = \frac{2TP}{2TP + FP + FN} \quad (7)$$

where TP are the true positives, FP are false positives, and FN are false negatives.

B. Experiments and Results

1) *Experiments*: Two experiments were conducted:

- Experiment $n^{\circ}1$: Comparing MobileNetV2 versus ResNet50: In this experiment, we want to compare the performance of two popular convolutional neural network architectures, MobileNetV2 and ResNet50, on a common image classification task. Both models were tested on a separate test dataset of 411 images.
- Experiment $n^{\circ}2$: Adding an Attention Mechanism with Correlation to MobileNetV2: In our second experiment, we want to investigate the impact of adding an attention mechanism with correlation to MobileNetV2. To add this mechanism, we modify the MobileNetV2 architecture (see Fig. 5). We then train the modified model on the same dataset as before, and evaluate its efficiency on the testing set. We compare the results of the original MobileNetV2 model with and without the attention mechanism to see if there was a significant difference in performance.

2) *Results*:

- Experiment $n^{\circ}1$: Here is the structured presentation of the results from Experiment $n^{\circ}1$, as shown in Tab. II:

TABLE II: Comparing MobileNetV2 and ResNet50

| | Recall | Precision | F1-score | Training Time (Second (s)) |
|-------------|---------------|---------------|---------------|----------------------------|
| MobileNetV2 | 74.30% | 92.78% | 82.18% | 480s |
| ResNet50 | 70.75% | 79.85% | 74.56% | 720s |

- Experiment $n^{\circ}2$: Here is the structured presentation of the results from Experiment $n^{\circ}2$, as shown in Tab. III:

TABLE III: Adding an Attention Mechanism with Correlation to MobileNetV2

| | Recall | Precision | F1-score |
|--|---------------|---------------|---------------|
| MobileNetV2 | 74.30% | 92.78% | 82.18% |
| MobileNetV2 + Spatial attention mechanism with correlation | 87.30% | 87.44% | 87.08% |

V. DISCUSSION

The results of Experiment $n^{\circ}1$ demonstrate that MobileNetV2 is a superior model for image classification tasks compared to ResNet50. The precision, recall, and f1-score metrics all showed that MobileNetV2 outperformed ResNet50. Additionally, MobileNetV2's use of depthwise separable convolutions allowed it to achieve these high results while using fewer resources, making it a more efficient and cost-effective option. In Experiment $n^{\circ}2$, we find that adding a spatial attention mechanism with correlation to MobileNetV2 further improved its performance on image classification tasks. By assigning different weights to different parts of the image, the attention mechanism enhances the model's ability to capture subtle patterns and important details related to MGMT promoter methylation. This attention-based approach improves the accuracy

and specificity of our model, enabling more precise predictions and a better understanding of the underlying molecular characteristics of gliomas. Additionally, the attention mechanism did not significantly increase the training time or computational cost of the model, making it a feasible option for real-world applications. Overall, these findings demonstrate that adding a spatial attention mechanism to MobileNetV2 can significantly enhance its performance on image classification tasks.

Incorporating other types of data, such as par-clinical data or MRI images, may enhance the accuracy and generalizability of our model. By considering a wider range of data sources, our model may be better equipped to capture the complex and multifaceted nature of gliomas, leading to more precise and reliable predictions. Therefore, future studies should aim to incorporate a more diverse range of data sources to improve the robustness and generalizability of our model. Despite these limitations, our findings provide valuable insights into the potential of histological image analysis and attention-based approaches for molecular biomarkers classification.

VI. CONCLUSION

As technology continues to advance and more data becomes available, the use of deep learning algorithms in medical imaging is likely to become even more widespread, offering new hope for those with gliomas and other difficult-to-treat medical conditions. By leveraging the MobileNetV2 as our transfer learning model and incorporating an attention mechanism, we were able to overcome the challenge of limited labeled data and obtain highly satisfactory outcomes while simultaneously maintaining a compact model size and computational requirements. As a future perspective, we aim to incorporate a more diverse range of data sources to enhance the robustness and generalizability of our model. Additionally, we aim to explore the progression of glioma grades from grade II to grade III and subsequently to grade IV. By studying the evolution of glioma grades, we can gain insights into the molecular mechanisms underlying glioma progression and develop more effective diagnostic and treatment strategies. Additionally, this can contribute to the development of personalized medicine approaches tailored to individual patients based on their specific glioma grade and MGMT promoter methylation status.

DECLARATION OF COMPUTING INTEREST

The authors state that they have no competing financial interests that could have influenced this paper.

REFERENCES

- [1] Zhu, G., Jiang, B., Tong, L., Xie, Y., Zaharchuk, G., and Wintermark, M., "Applications of deep learning to neuro-imaging techniques," *Frontiers in neurology*, 10, 869, 2019. <https://doi.org/10.1016/j.media.2022.102444>
- [2] Chen, X., Wang, X., Zhang, K., Fung, K.-M., Thai, T.C., Moore, K., Mannel, R.S., Liu, H., Zheng, B., and Qiu, Y., "Recent advances and clinical applications of deep learning in medical image analysis", *Medical Image Analysis* 79, 102444, 2022. <https://doi.org/10.1016/j.media.2022.102444>
- [3] Echle, A., Rindtorff, N. T., Brinker, T. J., Luedde, T., Pearson, A. T., and Kather, J. N., "Deep learning in cancer pathology: a new generation of clinical biomarkers". *British journal of cancer*, 124(4), 686-696, 2021. <https://doi.org/10.1038/s41416-020-01122-x>

- [4] Louis, D. N., Perry, A., Wesseling, P., Brat, D. J., Cree, I. A., Figarella-Branger, D., and Ellison, D. W., "The 2021 WHO classification of tumors of the central nervous system: a summary. *Neuro-oncology*", 23(8), 1231-1251, 2021. <https://doi.org/10.1093/neuonc/noab106>
- [5] Craig H., Keith L L., Priscilla B., Jason T H., Monica V., Susan C., Jan B., Timothy C., Robert B J., Caterina G., Roger S., L B. N., Patrick Y W., Kenneth J A., Rimas V L., Evanthia G., Charles G E., Daniel J B., and Jann N S., "The medical necessity of advanced molecular testing in the diagnosis and treatment of brain tumor patients", *Neuro-Oncology*, Volume 21, Issue 12, Pages 1498–1508, 2019. <https://academic.oup.com/neuro-oncology/article/21/12/1498/5528614>
- [6] Alice M., and Audrey R., "Le diagnostic histo-moléculaire des tumeurs gliales et glioneurales", *Annales de Pathologie*, Volume 41, Issue 2, Pages 137-153, 2021. <https://www.sciencedirect.com/science/article/abs/pii/S0242649820302844?via%3Dihub>
- [7] Haque, W., Teh, C., Butler, E. B., and Teh, B. S., "Prognostic and predictive impact of MGMT promoter methylation status in high risk grade II glioma", *Journal of Neuro-Oncology*, 157(1), 137-146, 2022. <https://link.springer.com/10.1007/s11060-022-03955-3>
- [8] Haque, W., Thong, E., Andrabi, S., Verma, V., Butler, E. B., and Teh, B. S., "Prognostic and predictive impact of MGMT promoter methylation in grade 3 gliomas", *Journal of Clinical Neuroscience*, 85, 115-121, 2021. <https://doi.org/10.1016/j.jocn.2020.12.028>
- [9] Karschnia, P., Teske, N., Dorostkar, M. M., Siller, S., Weller, J., Baehring, J. M., ... and Thon, N., "Extent and prognostic value of MGMT promoter methylation in glioma WHO grade II". *Scientific Reports*, 10(1), 1-9, (2020). <https://doi.org/10.1038/s41598-020-76312-x>
- [10] Ochsenein, A.F., Schubert, A.D., Vassella, E., and Luigi M., "Quantitative analysis of O6-methylguanine DNA methyltransferase (MGMT) promoter methylation in patients with low-grade gliomas", *J Neurooncol* 103, 343–351, 2011. <https://doi.org/10.1007/s11060-010-0395-2>
- [11] Zlochower, A., Chow, D. S., Chang, P., Khatri, D., Boockvar, J. A., and Filippi, C. G., "Deep learning AI applications in the imaging of glioma", *Topics in Magnetic Resonance Imaging*, 29(2), 115-00, 2020. <https://journals.lww.com/10.1097/RMR.0000000000000237>
- [12] Tran, K. A., Kondrashova, O., Bradley, A., Williams, E. D., Pearson, J. V., and Waddell, N., "Deep learning in cancer diagnosis, prognosis and treatment selection", *Genome Medicine*, 13(1), 1-17, 2021. <https://genomemedicine.biomedcentral.com/articles/10.1186/s13073-021-00968-x>
- [13] Korfiatis, P., Kline, T. L., Lachance, D. H., Parney, I. F., Buckner, J. C., and Erickson, B. J., "Residual deep convolutional neural network predicts MGMT methylation status", *Journal of digital imaging*, 30, 622-628, 2017. <http://link.springer.com/10.1007/s10278-017-0009-z>
- [14] Han, L., and Kamdar, M. R., "MRI to MGMT: predicting methylation status in glioblastoma patients using convolutional recurrent neural networks", In *Pacific Symposium on Biocomputing 2018: Proceedings of the Pacific Symposium* (pp. 331-342), 2018. https://www.worldscientific.com/doi/abs/10.1142/9789813235533_0031
- [15] Chang, P., Grinband, J., Weinberg, B. D., Bardis, M., Khy, M., Cadena, G., ... and Chow, D., "Deep-learning convolutional neural networks accurately classify genetic mutations in gliomas", *American Journal of Neuroradiology*, 39(7), 1201-1207, 2018. <http://www.ajnr.org/lookup/doi/10.3174/ajnr.A5667>
- [16] Chen, X., Zeng, M., Tong, Y., Zhang, T., Fu, Y., Li, H., ... and Jiang, X., "Automatic prediction of MGMT status in glioblastoma via deep learning-based MR image analysis", *BioMed research international*, 2020. <https://www.hindawi.com/journals/bmri/2020/9258649/>
- [17] Myronenko, A., "3D MRI brain tumor segmentation using autoencoder regularization", In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th International Workshop, BrainLes 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers, Part II 4* (pp. 311-320). Springer International Publishing, 2019.
- [18] Yogananda, C. G. B., Shah, B. R., Nalawade, S. S., Murugesan, G. K., Yu, F. F., Pinho, M. C., ... and Maldjian, J. A., "MRI-based deep-learning method for determining glioma MGMT promoter methylation status", *American Journal of Neuroradiology*, 42(5), 845-852, 2021. <http://www.ajnr.org/lookup/doi/10.3174/ajnr.A7029>
- [19] Chen, S., Xu, Y., Ye, M., Li, Y., Sun, Y., Liang, J., ... and Zhang, B., "Predicting MGMT Promoter Methylation in Diffuse Gliomas Using Deep Learning with Radiomics", *Journal of Clinical Medicine*, 11(12), 3445, 2022. <https://www.mdpi.com/2077-0383/11/12/3445>
- [20] Saeed, N., Hardan, S., Abutalip, K., and Yaqub, M., "Is it possible to predict MGMT promoter methylation from brain tumor MRI scans using deep learning models?", In *International Conference on Medical Imaging with Deep Learning* (pp. 1005-1018). PMLR, 2022, December. <http://arxiv.org/abs/2201.06086>
- [21] Egana, L., Auzmendi-Iriarte, J., Andermatten, J., Villanua, J., Ruiz, I., Elua-Pinin, A., ... and Samprón, N. (2020). Methylation of MGMT promoter does not predict response to temozolomide in patients with glioblastoma in Donostia Hospital. *Scientific Reports*, 10(1), 18445, 2020. <https://www.nature.com/articles/s41598-020-75477-9>
- [22] Kanwal, N., Pérez-Bueno, F., Schmidt, A., Egan, K., and Molina, R., "The Devil is in the Details: Whole Slide Image Acquisition and Processing for Artifacts Detection, Color Variation, and Data Augmentation: A Review". *IEEE Access*, 10, 58821-58844, 2022. <https://ieeexplore.ieee.org/abstract/document/9777677>
- [23] Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., ... and He, Q., "A comprehensive survey on transfer learning", *Proceedings of the IEEE*, 109(1), 43-76, 2020. [doi:10.1109/JPROC.2020.3004555](https://doi.org/10.1109/JPROC.2020.3004555)
- [24] Wang, W., Li, Y., Zou, T., Wang, X., You, J., and Luo, Y., "A novel image classification approach via dense-MobileNet models", *Mobile Information Systems*, 2020. <https://doi.org/10.1155/2020/7602384>
- [25] Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D., "Grad-cam: Visual explanations from deep networks via gradient-based localization", In *Proceedings of the IEEE international conference on computer vision* (pp. 618-626), 2017. https://openaccess.thecvf.com/content_ICCV_2017/papers/Selvaraju_Grad-CAM_Visual_Explanations_ICCV_2017_paper.pdf