

Convolutional neural networks for predicting MGMT methylation status in glioblastoma patients

Felipe Sassi¹, Danilo Silva¹

¹Departamento de Engenharia Elétrica e Eletrônica
Universidade Federal de Santa Catarina (UFSC)
Florianópolis – SC – Brazil

`felipesassi@outlook.com, danilo.silva@ufsc.br`

Abstract. *MGMT gene methylation determines the success of chemotherapeutic treatment of brain tumors known as glioblastomas. This paper explores four recent techniques, with the best performances reported in the literature, for classification of the status of the MGMT gene in patients with glioblastomas, aiming to compare the results in a common, public, high quality database. Approaches based on extraction of radiomic features as well as on convolutional neural networks for both 2D/3D classification and semantic segmentation are investigated. None of the tested approaches demonstrated significant classification ability, with ROC-AUC values ranging from 56,44% to 65,02%.*

Resumo. *A metilação do gene MGMT determina o sucesso do tratamento quimioterápico de tumores cerebrais conhecidos como glioblastomas. Este trabalho explora quatro técnicas recentes, com os melhores desempenhos reportados na literatura, para classificação do status do gene MGMT em pacientes com glioblastomas, visando a comparação dos resultados em uma mesma base de dados pública e de elevada qualidade. São investigadas abordagens baseadas em extração de features radiômicas e baseadas em redes neurais convolucionais, tanto para classificação 2D/3D quanto para segmentação semântica. Nenhuma das abordagens testadas demonstrou significativa capacidade de classificação, com valores de ROC-AUC variando de 56,44% a 65,02%.*

1. Introdução

A automatização de diferentes processos médicos via aprendizado de máquina vem ganhando destaque ao longo do tempo, devido a seus resultados equiparáveis aos de especialistas em diversas áreas. Uma das possíveis aplicações é no diagnóstico automatizado de gliomas, tumores cerebrais que afetam células da glia (cuja função é proteger e nutrir neurônios). Esse tipo de neoplasia (massa de tecido anormal) é comumente encontrado sob a forma de glioblastoma multiforme, a qual também é a categoria mais agressiva de tumor cerebral [GLIOBLASTOMA 2018]. Cerca de 45% dos tumores cerebrais que acometem o sistema nervoso central são glioblastomas multiformes, com um tempo médio de sobrevivência de 14,6 meses [Johnson and O'Neill 2012].

A partir do conhecimento da doença e correta localização, seu tratamento é realizado tanto de forma cirúrgica quanto a partir de um tratamento quimioterápico. Atualmente, um dos principais tratamentos é a quimioterapia com agentes alquilantes, como

a Temozolomida (TMZ), o qual tem sua taxa de sucesso atrelada a determinados marcadores genéticos do paciente [Trindade et al. 2012]. Um dos marcadores genéticos de interesse é o gene MGMT (*O6-metilguanina-DNA metiltransferase*) devido a sua relação com o prognóstico dos pacientes. Esse gene tem um efeito reparador nas células tumorais devido à resistência, proporcionada por ele, a agentes alquilantes. O estudo do gene MGMT está relacionado a identificação da metilação (silenciamento) desse gene, o qual provoca alterações que acabam removendo seu efeito restaurador em células tumorais [Gepp 2013]. A presença desse gene metilado está intimamente ligada ao sucesso de quimioterápicos alquilantes como TMZ, sendo comumente usados para complementar os resultados obtidos cirurgicamente [de Almeida et al. 2019] e aumentar o tempo de sobrevida dos pacientes [Trindade et al. 2012].

Atualmente, a principal forma de detecção do *status* da metilação do gene MGMT é via biópsia do tecido cerebral, a qual acarreta riscos para a saúde do paciente, principalmente quando o tumor está localizado em área de difícil acesso. Devido a esses riscos, um significativo esforço está sendo aplicado ao desenvolvimento de técnicas não-invasivas, baseadas na classificação de imagens de ressonância magnética por meio de aprendizado de máquina, para identificação do *status* do gene MGMT [Yogananda et al. 2021] e consequente aprimoramento diagnóstico.

Alguns trabalhos, seja utilizando extração de *features* radiômicas [Chen et al. 2015] ou adaptando modelos de segmentação semântica de tumores cerebrais [Yogananda et al. 2021], tem reportado elevados valores da métrica ROC-AUC, de cerca de 90%. Por outro lado, um trabalho recente [Saeed et al. 2022], utilizando redes convolucionais 2D e 3D para classificação, reportou resultados de ROC-AUC não superiores a 64%. Vale ressaltar que cada um destes trabalhos utilizou uma base de dados diferente.

Tal discrepância entre resultados fornece a motivação para o presente trabalho: análise, reimplementação e comparação das abordagens existentes em uma mesma base de dados pública e de elevada qualidade.

2. Trabalhos relacionados

Diferentes abordagens para classificação do *status* do gene MGMT podem ser aplicadas, algumas mais simples e diretas e outras necessitando de etapas robustas tanto de pré-processamento quanto de pós-processamento. Essas abordagens podem ser agrupadas em quatro grupos, descritos a seguir.

2.1. Classificação através de extração de *features*

Em [Le et al. 2020], os autores desenvolveram uma *pipeline* composta por uma etapa de extração de características do tumor (características radiômicas) seguida por um modelo aprendizado de máquina de classificação do *status* do gene MGMT. Esse modelo foi desenvolvido com base em um conjunto de dados composto por 53 pacientes oriundos do *The Cancer Imaging Archive* (TCIA). Neste trabalho os autores reportam uma métrica ROC-AUC de 89,6%

2.2. Classificação *end-to-end* com *deep learning*

Um dos métodos mais naturais de se resolver o problema de classificação do *status* do gene MGMT é a partir de um modelo de redes neurais convolucionais 3D treinado a

partir do conjunto de exames e de rótulos. No trabalho [Saeed et al. 2022] os autores exploram essa técnica a partir de testes com diferentes arquiteturas de redes neurais.

O conjunto de dados utilizado pelos autores é o oriundo da competição BRATS 2021, hospedada na plataforma *Kaggle*. Esse conjunto é composto por imagens de ressonância magnética de 577 pacientes com seus respectivos rótulos (MGMT metilado/MGMT não metilado) anotados por comitês de especialistas [Baid et al. 2021].

Nesse trabalho a abordagem de classificação 3D baseada em uma arquitetura conhecida como *EfficientNet* obteve uma métrica ROC-AUC (%) de 64% no conjunto de teste.

2.3. Deep learning para classificação de fatias

Em [Korfiatis et al. 2017] os autores exploram essa tarefa sob uma ótica de classificação de fatias por meio da utilização de uma arquitetura *ResNet*. Neste trabalho três variações dessa arquitetura são testadas: *ResNet18*, *ResNet34* e *ResNet50*, todas elas treinadas com base nas fatias que compõem os volumes de entrada. Os autores demonstram em seus resultados que redes neurais convolucionais são úteis para diagnóstico do *status* do gene MGMT, sendo a arquitetura *ResNet50* a melhor dentre às três testadas, com uma acurácia de 94,90% (devido ao método utilizado pelos autores, a métrica ROC-AUC não pode ser computada neste trabalho).

Os autores utilizaram 155 exames de ressonância magnética, de pacientes acima de 18 anos, sendo 66 deles com gene MGMT metilado e 89 com gene MGMT não metilado. Todos os volumes continham as sequências T1 e T2 e eram oriundos de um conjunto de dados privado. O treinamento foi realizado a partir da extração das fatias (*slices*) que compõem o volume de entrada. As fatias foram classificadas individualmente com base em três classes: presença de metilação no gene, ausência de metilação no gene ou fatia sem tumor. O terceiro rótulo foi utilizado para auxiliar o treinamento do modelo, reduzindo o ruído dos dados.

2.4. Abordagem via segmentação semântica

Um trabalho que explora essa forma de identificação é o trabalho [Yogananda et al. 2021], onde os autores utilizam modelos de redes neurais para realizar tanto a segmentação do tumor cerebral quanto o diagnóstico da metilação do gene MGMT. Neste trabalho os autores conseguiram obter uma ROC-AUC na detecção do *status* do tumor cerebral de 93,00% e uma métrica *dice* (para segmentação do tumor) de 82,00%. Todos os resultados foram obtidos por meio de uma estratégia de validação cruzada com 3 *folds*. Além da ROC-AUC e da métrica *dice*, os autores também reportaram valores de acurácia de 94,73%, com uma sensibilidade de 96,31% e uma especificidade de 91,66%, resultados promissores para essa tarefa. Foram selecionados 247 pacientes para desenvolvimento do trabalho, com seus dados oriundos do TCIA.

3. Metodologia

3.1. Conjuntos de dados

O conjunto de dados principal do trabalho é o disponibilizado pela competição BRATS 2021. Esse conjunto traz volumes de ressonância magnética de diferentes pacientes, totalmente pré-processados. Além de ser o principal conjunto de dados, ele também é o

maior conjunto de dados públicos disponível. Sua atualização é realizada anualmente, às vezes trazendo novos pacientes e novas tarefas possíveis de serem resolvidas com os dados disponibilizados.

Como pré-processamento, as imagens brutas (RMI) são submetidas a diferentes etapas pelos organizadores da competição: conversão dos arquivos DICOM para o formato NIfTI, co-registro para o *template* SRI24, *resample* para uma resolução de 1mm³ e remoção de crânio. Esse pré-processamento ajuda a reduzir a quantidade de ruído presente nas imagens e são passos comuns em todas as competições. Após essa etapa os volumes gerados estão padronizados e com um tamanho de 240X240X155.

Na última atualização (2021), as tarefas de segmentação semântica de tumores cerebrais e de classificação do *status* do gene MGMT dos pacientes com tumores cerebrais foram selecionadas para exploração. Em ambos os casos, o conjunto de dados trouxe rótulos anotados por comitês de especialistas, ajudando a garantir a qualidade do conjunto e a utilidade dos estudos desenvolvidos. Para a tarefa de classificação da metilação 577 exames foram disponibilizados, bem como seus rótulos (MGMT metilado/não metilado). Desses, 52% são da classe MGMT metilado e 48% são da classe MGMT não metilado, mostrando um bom nível de balanceamento do conjunto de dados.

3.2. Divisão do conjunto de dados

Para construção dos experimentos o conjunto de dados BRATS 2021 foi dividido em treino e teste (validação *hold-out*). Para treinamento, 75% do conjunto de dados foi amostrado e o restante dos dados foram utilizados como conjunto de testes. Uma amostragem estratificada foi utilizada na geração dos conjuntos para manter a distribuição original das classes. Em termos numéricos, 433 pacientes foram selecionados para treino enquanto 144 pacientes foram selecionadas para teste.

Os conjuntos de dados são divididos com base nos identificadores únicos dos pacientes. Essa abordagem é necessária para evitar que informações sobre pacientes utilizados no treinamento também sejam utilizadas para teste. Esse cenário pode ocorrer nas abordagens baseadas em fatias e em segmentação semântica, devido à relação 1:N entre pacientes e fatias/*patches*.

Todos os experimentos serão baseados no conjunto de treino e todas as comparações entre abordagens serão realizadas a partir do conjunto de teste, garantindo a capacidade de comparação entre os diferentes experimentos.

3.3. Abordagem baseada em extração de *features*

A abordagem baseada em extração de *features* utiliza a biblioteca *PyRadiomics* para construir um conjunto de características a partir dos volumes dos tumores. Para garantir que os resultados não sejam afetados pela qualidade das segmentações, somente as máscaras originais dos tumores foram utilizadas.

A partir do conjunto de características extraído, diferentes modelos serão testados: regressão logística, modelos baseados em *random forests* e *gradient boosting*. A determinação do melhor modelo será realizada por validação cruzada baseada no método *repeated random sub-sampling validation* (validação cruzada de Monte Carlo) [Dubitzky et al. 2007], técnica especialmente útil para determinação de métricas robustas em cenários de conjuntos de dados pequenos.

3.4. Abordagem 3D (*end-to-end*)

Na abordagem 3D os testes focaram em duas arquiteturas: *ResNet* e *EfficientNet*. Originalmente essas arquiteturas foram desenvolvidas para lidar com imagens, mas com devidas adaptações (como utilização de camadas convolucionais 3D), conseguem também lidar com volumes 3D.

Para treinamento e seleção de hiperparâmetros o conjunto de treino original será dividido novamente para construção de dois novos, sendo um utilizado para o treinamento e outro utilizado para validação em etapas de otimização de hiperparâmetros. O novo conjunto de treino é composto por 75% dos pacientes presentes no conjunto de treino original. O restante dos pacientes constitui o conjunto de validação. Essa divisão será utilizada em todas as abordagens envolvendo redes neurais convolucionais devido ao tempo de treinamento desses modelos.

3.5. Abordagem 2D (baseada em fatias)

Para a classificação ao nível de fatias, os testes serão realizados a partir de modelos baseados na arquitetura *ResNet* (assim como em [Korfiatis et al. 2017]) e na arquitetura *EfficientNet*. Diferentemente das abordagens anteriores, a análise do problema sob a ótica 2D faz com que o processo de classificação necessite de uma reformulação: a classificação deverá classificar cada fatia em três classes, sem tumor, com tumor MGMT metilado e com tumor e MGMT não metilado (necessitando do conhecimento da segmentação do tumor para correta construção).

Os dados são obtidos a partir da extração de “fatias” que compõe o volume de entrada. A partir dos dados utilizados, um volume de dimensões 240X240X155 gerará um conjunto de 155 imagens de dimensão 240X240. A última dimensão foi utilizada para “fatiamento” para possibilitar a obtenção de imagens de dimensões iguais.

Como o treinamento é efetuado por fatia e a classificação deve ser feita ao nível de paciente, as previsões individuais precisam ser fundidas para geração de um *score* de saída. Para geração desse *score* um pós-processamento será aplicado. Essa etapa inicialmente irá remover as fatias classificadas como sem tumor (fatias com o maior *score* de saída referente a essa classificação) para depois gerar dois *scores* a partir de dois tipos de votação:

- *hard voting*: a geração do *score* de saída do paciente é feita com base na proporção de fatias classificadas como MGMT metilado em relação às classificadas como MGMT não metilado;
- *soft voting*: nessa abordagem a média dos *scores* para metilação e não metilação para cada paciente é obtida para posterior re-normalização (soma das médias deve ser 1).

3.6. Abordagem via segmentação semântica

Seguindo a linha apresentada no trabalho original [Yogananda et al. 2021], uma arquitetura utilizada para segmentação semântica será utilizada para classificação individual de cada *voxel* pertencente ao tumor. Essa abordagem gera dois volumes de saída, um correspondendo ao gene não metilado e outro correspondendo ao gene metilado.

A arquitetura utilizada será a U-Net [Ronneberger et al. 2015], a qual será submetida a variações na quantidade de filtros utilizada em cada camada, bem

como a quantidade de etapas utilizadas nos blocos de *encoding* e *decoding*, visando a maximização da métrica ROC-AUC no conjunto de validação. A camada final será alterada para geração de dois volumes de saída.

Todos os treinamentos dessa abordagem serão baseados em *patches* de 32X32X32, com 75% de *overlap* e com no mínimo 1 *voxel* de tumor. Cada um desses *patches* será processado pelo modelo e um *score* será gerado (valor entre 0 e 1) após normalização por uma função sigmoide. A função custo utilizada é a entropia cruzada binária. Todos os hiperparâmetros utilizados pelo modelo, bem como a necessidade de *data augmentation* serão ajustados conforme a necessidade, visando o melhor resultado de métrica ROC-AUC possível no conjunto de validação.

Para geração da máscara as previsões para cada um dos *patches* gerados serão comparadas *voxel a voxel* de forma se obter o maior *score* de saída para cada um dos *voxels*. Após isso, uma etapa de binarização será realizada: *scores* maiores que 0,5 serão transformados em 1 e *scores* menores em 0, resultando em uma máscara binária que pode ser utilizada para remoção de estruturas cerebrais fora da região do tumor.

Após as previsões, um pós-processamento será realizado com o intuito de fundir as previsões e gerar um *label* para o paciente (para posterior análise via métrica ROC-AUC). A fusão, assim como na abordagem anterior, será realizada com base em um esquema *soft voting* e também com base em um esquema *hard voting*. Devido à capacidade de geração da máscara do tumor, esses esquemas de votação serão realizados de duas maneiras: sem a utilização da máscara do tumor e com a utilização da máscara do tumor. A utilização da máscara permitirá a remoção de *voxels* fora do tumor para cômputo do *score* de saída de cada paciente. Os esquemas de votação *hard* e *soft* são definidos da seguinte forma:

- *hard voting* inicialmente somam-se todas as previsões para cada saída, e após isso se determina o *status* do *patch* a partir da comparação desses valores. Com isso, a geração do *score* de saída do paciente é feita com base na proporção de *patches* classificados como MGMT metilado em relação às classificadas como MGMT não metilado;
- *soft voting*: nessa abordagem a soma (considerando todos os *patches*) dos *scores* para metilação e não metilação de cada paciente é obtida para uma posterior re-normalização (soma dos valores deve ser 1).

4. Resultados

4.1. Classificação a partir de *features* radiômicas

Os três classificadores foram implementados usando a biblioteca *sklearn* [Pedregosa et al. 2011]. Hiperparâmetros relacionados à regularização foram utilizados para evitar o *overfitting*, presente principalmente nos classificadores baseados em *random forests* e em *gradient boosting*.

Na regressão logística a constante de regularização foi alterada com o intuito de se reduzir o *overfitting*. Nos modelos de *random forests* e *gradient boosting* parâmetros como a profundidade, número de estimadores e total de amostras por nó foram alterados, com o intuito de se reduzir a complexidade dos modelos (e também reduzir o *overfitting*).

Outra técnica que se mostrou útil foi a remoção de *features* correlacionadas. Para cada grupo de características, com correlação maior que 80%, somente a *feature* de maior

variância foi mantida. A partir dessa seleção o conjunto de *features* foi reduzido para 35. Essa abordagem melhorou o resultado de ambas abordagens, principalmente reduzindo o sobreajuste nos dados de treinamento.

A tabela 1 sintetiza tanto o valor médio de cada métrica quanto o desvio-padrão obtido (representado pelo valor entre parênteses).

Tabela 1. Resultados obtidos na abordagem baseada em extração de *features*.

	Treino	Validação	Teste
Regressão logística	68,42 (1,32)	58,17 (4,10)	65,02 (2,10)
Gradient boosting	78,24 (3,50)	54,08 (4,38)	56,93 (4,57)
Random forests	76,53 (1,38)	55,43 (4,70)	60,00 (4,69)

4.2. Classificação 3D *end-to-end*

Essa abordagem visa classificar cada paciente individualmente, a partir de seu volume de entrada. Para essa abordagem os rótulos de cada paciente (MGMT metilado/MGMT não metilado) foram utilizados para treinamento. Diferentemente da solução anterior, essa abordagem não necessita do conhecimento da segmentação do tumor, tornando sua aplicação mais direta.

Diferentes arquiteturas foram testadas, sendo uma arquitetura *ResNet* a selecionada. Essa arquitetura, chamada *ResNet10* é uma versão menor (de menor complexidade) das arquiteturas *ResNet*. A redução de complexidade do modelo foi necessária devido à facilidade de *overfitting* encontrada durante o treinamento de modelos com maior número de parâmetros, além de instabilidade (variação muito grande) na métrica de validação ao longo do treinamento.

Mesmo com a alteração de diferentes hiperparâmetros (e diminuição da variação das métricas) e com a redução de complexidade do modelo o *overfitting* continuava presente durante o treinamento. Com o intuito de se aumentar a variabilidade do conjunto de treinamento, diferentes técnicas de *data augmentation* foram testadas, como inserção de ruído aleatório, rotações aleatórias, etc (sempre aplicadas no volume completo de entrada). Com os testes observou-se que a utilização de um *zoom* aplicado aleatoriamente no volume de entrada conseguia não só reduzir significativamente o *overfitting* como suavizar a evolução das métricas ao longo do treinamento.

A arquitetura *ResNet10* foi treinada a partir do otimizador o *Adam*, com uma taxa de aprendizado de 0,001. A função custo selecionada para treinamento foi a entropia cruzada binária.

A Tabela 2 apresenta os resultados em termos da métrica ROC-AUC (%) para essa abordagem. Nota-se a proximidade entre os resultados obtidos no conjunto de validação e no conjunto de teste.

Tabela 2. Resultados em termos de métrica ROC-AUC (%) obtidos na abordagem *end-to-end* 3D.

	Treino	Validação	Teste
<i>End-to-end</i>	71,60	62,26	58,46

4.3. Classificação 2D baseada em fatias

A terceira abordagem testada visa classificar cada uma das fatias que compõe o volume de entrada (ressonância magnética do paciente). A partir dessa classificação individual o *status* do paciente pode ser inferido através de uma etapa de pós-processamento.

A classificação é realizada para determinar se a fatia não tem tumor ou, caso tiver um tumor, classificá-lo em MGMT metilado ou MGMT não metilado, tornando a abordagem um problema de classificação multiclasse. As fatias de entrada são oriundas de um volume de tamanho 240x240x155 o qual gera fatias de dimensão 240X240 *pixels*.

Antes do treinamento uma etapa de seleção de fatias foi realizada. Para uma fatia ser incluída no treinamento (e na inferência) ela deve ter ao menos 5028 *pixels* não nulos (70% das fatias presentes no conjunto de treinamento satisfazem essa condição). Isso foi útil para reduzir a quantidade de ruído inserido por fatias com pouca ou nenhuma área cerebral presente.

Mesmo com a remoção de fatias com pouca ou nenhuma área cerebral, o processo de treinamento não apresentava sinais de aprendizado ao longo das épocas. Para auxiliar ainda mais o modelo, informações da máscara do tumor foram utilizadas para filtrar as fatias com poucos *pixels* de tumor. Fatias com menos de 50 *pixels* de tumor (em média a quantidade de *pixels* com tumor é de 683) são removidas durante a etapa de treinamento e durante a etapa de inferência. Caso a fatia seja sem tumor, ela é incluída sem passar por esse filtro.

O conjunto de treino gerado após extração e seleção de fatias é composto por 33 970 fatias, sendo 43% sem tumor, 30% com MGMT metilado e 27% com MGMT não metilado. O conjunto de teste é composto por 15 104 fatias com a mesma distribuição de classes.

Três técnicas de *data augmentation* foram combinadas. Assim como no caso anterior, o uso de *data augmentation* reduziu o *overfitting* (que surgia principalmente com taxas de aprendizado menor que 0,001) e a variação das métricas ao longo do treinamento. As transformações utilizadas foram *zoom* aleatório, rotações aleatórias e ruído *gaussiano*.

A partir dos experimentos, a arquitetura selecionada para essa abordagem foi a ResNet34, treinada por 13 épocas a partir do otimizador *Adam*, com uma taxa de aprendizado de 0,001 e uma regularização L2 de 0,0001. A entropia cruzada foi utilizada como função custo. Outras arquiteturas/variações foram utilizadas, mas não apresentaram resultados significativos.

A Tabela 3 sintetiza os resultados obtidos para essa abordagem. Nota-se a proximidade entre as métricas calculadas com diferentes formas de votação.

Tabela 3. Resultados em termos de métrica ROC-AUC (%) obtidos na abordagem baseada em fatias.

	Treino	Validação	Teste
Votação <i>soft</i>	78,60	60,28	-
Votação <i>hard</i>	77,52	60,68	60,55

A técnica de votação *hard* conseguiu obter a melhor métrica no conjunto de

validação. Essa abordagem, aplicada ao conjunto de teste, gera um resultado de ROC-AUC (%) de 60,55%.

4.4. Classificação a partir de segmentação semântica

A quarta e última abordagem utiliza um modelo de segmentação semântica para geração de máscaras para o tumor presente no paciente. A partir das segmentações geradas, a classificação do paciente pode ser determinada a partir de uma etapa de votação, a qual gera um *score* de saída para cada paciente.

Uma arquitetura baseada em U-Net foi utilizada para treinamento a partir de *patches* de tamanho 32X32X32, com 24 *voxels* de *overlap*, extraídos da RMI original. Tanto *patches* do cérebro quanto *patches* da máscara foram extraídos. Assim como em [Yogananda et al. 2021] uma etapa de seleção, baseada na quantidade de *voxels* do tumor foi utilizada para reduzir o ruído no conjunto de dados. Os *patches* utilizados para treinamento devem ter ao menos 1 *voxel* de tumor, devido a isso a quantidade de *patches* varia conforme o paciente selecionado.

O conjunto de dados utilizado para treinamento foi constituído por 538 185 *patches*. Desses, 52% estão com *status* positivo para a metilação do gene MGMT enquanto que 48% estão com *status* negativo. O conjunto de teste é composto por 177 822 *patches* com 54% deles pertencentes à classe MGMT metilado.

Logo nos primeiros testes a necessidade de *data augmentation* se mostrou necessária devido à facilidade de *overfitting* em menos do que 5 épocas de treinamento. Além das técnicas utilizadas anteriormente, outras foram utilizadas para aumentar a variabilidade do conjunto de dados: rotação e translações aleatórias, *blur*, e *flip* horizontal e vertical. Essas transformações também são aplicadas às segmentações utilizadas para garantir o *match* entre os dados utilizados para treino.

O treinamento da arquitetura U-Net foi realizado por 28 épocas, com um tamanho de *batch* de 256. O otimizador selecionado foi o *Adam* com uma taxa de aprendizado de 0,0001. A função custo utilizada foi entropia cruzada.

Os resultados obtidos podem ser observados na Tabela 4.

Tabela 4. Resultados em termos de ROC-AUC (%) obtidos na abordagem baseadas em segmentação semântica.

	Treino	Validação	Teste
<i>Soft voting sem máscara</i>	86,01	62,25	-
<i>Hard voting sem máscara</i>	73,15	59,01	-
<i>Soft voting com máscara</i>	86,35	63,42	56,44
<i>Hard voting com máscara</i>	75,95	59,65	-

O uso de *soft voting*, com a máscara do tumor, conseguiu obter melhores resultados no conjunto de validação. No conjunto de teste, essa abordagem obteve uma métrica ROC-AUC de 56,44 %.

4.5. Comparação de abordagens e discussão

Todas as abordagens resultaram em métricas ROC-AUC em torno de 60%, diferentemente dos resultados obtidos em alguns dos trabalhos apresentados anteriormente, porém em conformidade com os resultados obtidos pelos participantes da competição BRATS 2021 e com [Saeed et al. 2022].

A Tabela 5 apresenta os resultados obtidos em cada uma das abordagens. Os valores em negrito indicam que o valor foi o maior obtido. Pode-se notar a proximidade entre as métricas obtidas ao longo das diferentes abordagens, sendo a principal diferença observada nas métricas de treino.

Tabela 5. Comparação em termos de ROC-AUC (%) dos resultados obtidos ao longo dos experimentos.

	Treino	Validação	Teste
Extração de <i>features</i> - Regressão Logística	68,42	58,17	65,02
Extração de <i>features</i> - <i>Gradient boosting</i>	78,24	54,08	56,83
Extração de <i>features</i> - <i>Random forests</i>	76,53	55,43	60,00
<i>End-to-end</i>	71,60	62,26	58,46
Fatias - <i>Hard voting</i>	77,52	60,68	60,55
Segmentação semântica - <i>Soft voting</i> - Com máscara	86,35	63,42	56,44

A Figura 1 apresenta a comparação das curvas ROC para cada uma das abordagens presentes na Tabela 5. Essa comparação permite uma análise do desempenho dos classificadores em diferentes pontos de operação.

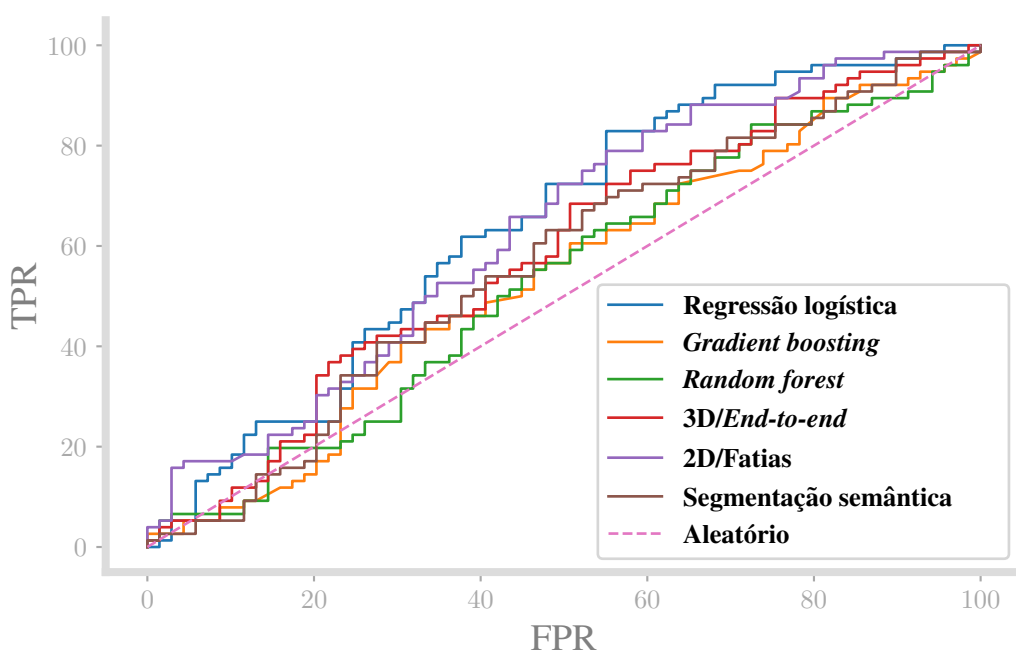


Figura 1. Curvas ROC para o conjunto de teste.

Algumas abordagens apresentadas na Tabela 5 apresentaram valores de métrica de treino superiores aos obtidos no conjunto de validação. Tentativas de redução dessa diferença, como regularização e *dropout* mais agressivos e mais *data augmentation* foram testadas mas não impactaram nas métricas de validação obtidas, somente reduzindo as métricas de treino.

O melhor resultado obtido foi o referente à abordagem via extração de *features*, utilizando um modelo de regressão logística para classificação dos pacientes. Um ponto de atenção é o desvio-padrão obtido nessa métrica (2,10), o qual ajuda a justificar esse desempenho superior.

Vale mencionar que tanto a abordagem baseada em extração de *features* quanto a abordagem baseada em fatias assumiram uma segmentação perfeita tanto para treinamento quanto para inferência, tornando as métricas apresentadas possivelmente mais otimistas que as reais. Na prática, será necessário utilizar um modelo de segmentação para essas tarefas, o qual introduz erro na máscara gerada, resultando possivelmente em métricas inferiores às observadas no trabalho.

5. Conclusão

Este trabalho investigou a classificação do *status* do gene MGMT a partir de imagens de ressonância magnética usando aprendizado de máquina, com base em quatro abordagens de estado-da-arte, visando a identificação da melhor abordagem a partir da comparação de resultados em uma base de dados comum, pública e de elevada qualidade.

Os resultados obtidos variaram entre 57,11% e 65,02%, esses resultados ficam em linha com os apresentados em [Saeed et al. 2022] e com os resultados obtidos na competição BRATS 2021, mostrando a dificuldade de se obter bons diagnósticos automatizados da metilação do gene MGMT somente com a utilização desse tipo de dado. Uma hipótese para o baixo desempenho dos métodos testados é a falta de informação útil para predição na RMI. Pode ser interessante investigar se a combinação de informações genéticas e dados do paciente, com RMIs, pode gerar modelos com melhor desempenho na tarefa de classificação.

Apesar dos resultados pessimistas, a comparação de diferente abordagens em um conjunto de dados comum e com elevado número de pacientes serviu para realizar uma avaliação justa das diferentes técnicas mapeadas, eliminando as diferenças de resultados oriundas da diferença entre as bases de dados.

Referências

- Baid, U., Ghodasara, S., Mohan, S., Bilello, M., Calabrese, E., Colak, E., Farahani, K., Kalpathy-Cramer, J., Kitamura, F. C., Pati, S., et al. (2021). The rsna-asnr-miccai brats 2021 benchmark on brain tumor segmentation and radiogenomic classification. *arXiv preprint arXiv:2107.02314*.
- Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K., et al. (2015). Xgboost: extreme gradient boosting. *R package version 0.4-2*, 1(4):1–4.
- de Almeida, M. L. M., de Aguiar, P. H. P., de Gois, K., de Sousa Gehrke, F., and Fonseca, F. (2019). The importance of mgmt promoter methylation status for glioblasto-

- mas prognosis: meta-analysis. *JBNC-JORNAL BRASILEIRO DE NEUROCIRURGIA*, 29(4):595–611.
- Dubitzky, W., Granzow, M., and Berrar, D. P. (2007). *Fundamentals of data mining in genomics and proteomics*. Springer Science & Business Media.
- Gepp, R. d. A. (2013). Análise prognóstica comparativa entre a perfusão cerebral por ressonância magnética e análise da metilação do gene promotor de mgmt nos gliomas malignos. Dissertação de mestrado, Universidade de Brasília.
- GLIOBLASTOMA, N. D. (2018). Glioblastoma multiforme recém-diagnosticado: diagnóstico e tratamento cirúrgico inicial. *Acta Médica*, 39(2):164.
- Johnson, D. R. and O’Neill, B. P. (2012). Glioblastoma survival in the united states before and during the temozolomide era. *Journal of neuro-oncology*, 107(2):359–364.
- Korfiatis, P., Kline, T. L., Lachance, D. H., Parney, I. F., Buckner, J. C., and Erickson, B. J. (2017). Residual deep convolutional neural network predicts mgmt methylation status. *Journal of digital imaging*, 30(5):622–628.
- Le, N. Q. K., Do, D. T., Chiu, F.-Y., Yapp, E. K. Y., Yeh, H.-Y., and Chen, C.-Y. (2020). Xgboost improves classification of mgmt promoter methylation status in idh1 wildtype glioblastoma. *Journal of personalized medicine*, 10(3):128.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.
- Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer.
- Saeed, N., Hardan, S. E., Abutalip, K., and Yaqub, M. (2022). Is it possible to predict MGMT promoter methylation from brain tumor MRI scans using deep learning models? In *Medical Imaging with Deep Learning*.
- Trindade, V., Picarelli, H., Figueiredo, E. G., and Teixeira, M. J. (2012). Gliomas: marcadores tumorais e prognóstico. *Arquivos Brasileiros de Neurocirurgia: Brazilian Neurosurgery*, 31(02):91–94.
- Yogananda, C., Shah, B. R., Nalawade, S., Murugesan, G., Yu, F., Pinho, M., Wagner, B., Mickey, B., Patel, T. R., Fei, B., et al. (2021). Mri-based deep-learning method for determining glioma mgmt promoter methylation status. *American Journal of Neuro-radiology*, 42(5):845–852.