

Predição do Estado de Metilação em Pacientes com Glioblastoma Multiforme Utilizando Aprendizado de Máquina

Giulia Silva Fazzi

Orientador: Prof. Dr. Zanoni Dias

Coorientador: Prof. Dr. Hélio Pedrini

18 de Outubro de 2023

Instituto de Computação

Universidade Estadual de Campinas

Introdução

Conceitos

Trabalhos Relacionados

Materiais e Métodos

Plano de Trabalho

Introdução

Conceitos

Trabalhos Relacionados

Materiais e Métodos

Plano de Trabalho

- Inteligência Artificial e câncer.
- Glioblastoma Multiforme (GBM).
- Estado de metilação.
- Exame de ressonância magnética.
- Redes neurais.

Propor uma abordagem fazendo uso de Redes Neurais para a identificação do estado de metilação, a partir de imagens de ressonância magnética, em pacientes com Glioblastomas Multiformes (GBM).

- Proposição de arquitetura de aprendizado de máquina profundo baseada em mecanismos de atenção para predição do estado de metilação em diferentes conjuntos de imagens.
- Proposição de técnicas de aumento de imagens para melhorar o desempenho do método preditivo proposto.

Questões de Pesquisa

1. Os resultados obtidos são comparáveis aos trabalhos correlatos recentemente publicados?
2. Qual o impacto do uso de redes pré-treinadas no resultado final a partir das métricas escolhidas?
3. O tempo de processamento perante uma nova imagem é adequado para aplicações no mundo real?
4. O modelo criado possui explicabilidade e satisfaz especificamente as necessidades desse tipo de modelo na Medicina?
5. Há diferença de desempenho entre redes neurais convolucionais tradicionais e aquelas baseadas em atenção para este problema?

Introdução

Conceitos

Trabalhos Relacionados

Materiais e Métodos

Plano de Trabalho

Redes Neurais Convolucionais

- Implementação de diversas camadas.
- Camadas de convolução, camadas de *pooling* e camadas totalmente conectadas.
- Redes neurais baseadas em mecanismo de atenção.

Redes Neurais Convolucionais

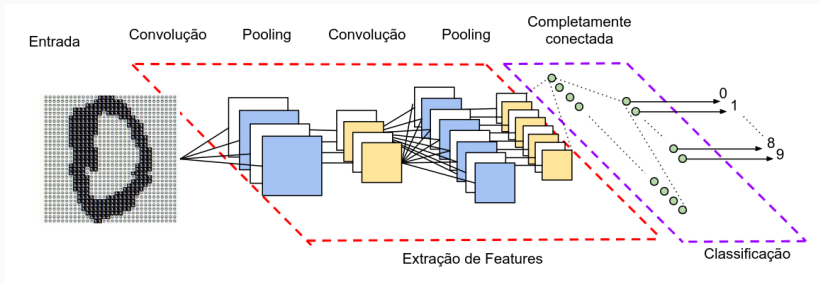


Figura 2: Estrutura em camadas de uma Rede Neural Convolutiva [1].

- Utilização do conhecimento adquirido por uma rede neural previamente treinada.
- Ajustes nos pesos e parâmetros.
- *Fine-Tuning*.
- Redução do tempo de treinamento.
- Aproveitamento de padrões identificados previamente.

- Padronização internacional para imagens médicas.
- Padrões de impressão, armazenamento, transmissão e tratamento.
- Utilizado na radiologia, cardiologia e radioterapia.

- A metilação do DNA é um processo epigenético.
- Relação com a evolução de um câncer.
- Biomarcador para o diagnóstico e prognóstico do câncer.
- Auxiliar na determinação da abordagem terapêutica.

Introdução

Conceitos

Trabalhos Relacionados

Materiais e Métodos

Plano de Trabalho

- Wang (2022) [2]:
 - Uso de CNN.
 - Falta de padronização na aquisição de imagens.
- Dildar et al. (2021) [3]:
 - Melhores resultados com CNN.
- Bai et al. (2021) [4]:
 - Consideração do histórico médico do paciente.
- Lopez et al. (2022) [5]:
 - Uso de Aprendizado de Máquina para prognóstico, diagnóstico, prevenção e terapia.

Classificação de Estado de Metilação

- Han et al. (2018) [6]:
 - Uso de Rede Neural Convolucional Recorrente (CRNN).
 - 62,0% de acurácia na base de dados TCIA e TCGA-GBM.
- Le et al. (2020) [7]:
 - Mutação no gene IDH1.
 - Uso de XGBoost.
 - 88,7% de acurácia na base de dados BraTS.
- Do et al. (2022) [8]:
 - Algoritmo Genético associado com o XGBoost para seleção de *features*.
 - 92,5% de acurácia na base de dados TCIA e TCGA-GBM.
- Minh et al. (2023) [9]:
 - Incorporação de características radiômicas no treinamento.
 - 80,0% de acurácia na base de dados TCIA e TCGA-GBM.

Introdução

Conceitos

Trabalhos Relacionados

Materiais e Métodos

Plano de Trabalho

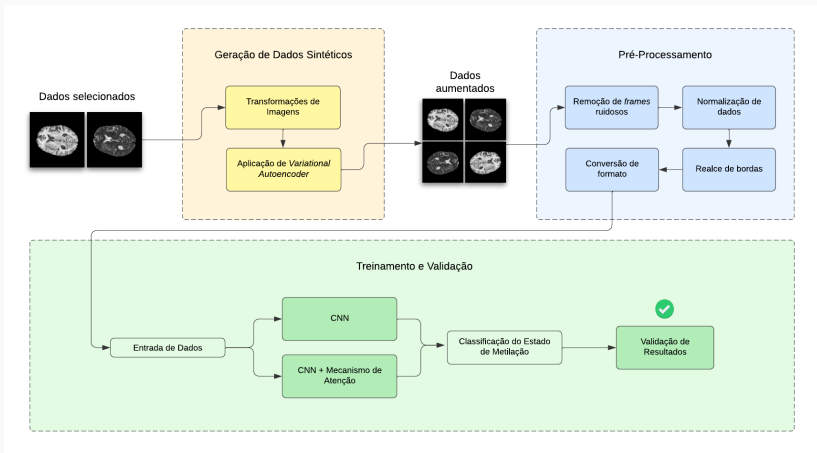


Figura 3: Principais etapas que compõem a metodologia do projeto.

- Seleção do Conjunto de Dados e dos Atributos:
 - Áreas de interesse.
- Separação do Conjunto de Dados:
 - *K-Fold Cross-Validation*.
- Geração de Dados Sintéticos:
 - Transformações de Imagens.
 - Aplicação de *Variational Autoencoder*.

- Pré-Processamento:
 - Etapas de tratamento de imagens.
 - Conversão de formato.
- Treinamento e Validação.
 - CNN e Mecanismo de Atenção.

- *The Cancer Imaging Archive* (TCIA) [10] e *The Cancer Genome Atlas Glioblastoma Multiforme Collection* (TCGA-GBM) [11].
- *Brain Tumor Segmentation* (BraTS) [12, 13].
- Formato DICOM.
- Cerca de 155 imagens por exame, representadas por matrizes de 240×240 pixels.

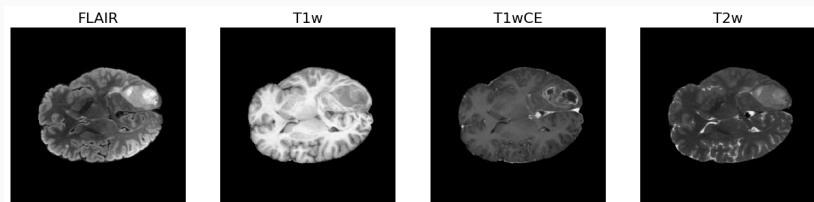


Figura 4: Imagens obtidas com exame de ressonância magnética.

- Acurácia.
- Sensibilidade.
- Especificidade.
- F_1

$$\text{Acurácia} = \frac{VP + VN}{P + N} \quad (1)$$

$$\text{Acurácia Balanceada} = \frac{1}{2} \times \left(\frac{VP}{P} + \frac{VN}{N} \right) \quad (2)$$

Sensibilidade e Especificidade

$$\text{Sensibilidade} = \frac{VP}{VP + FN} \quad (3)$$

$$\text{Especificidade} = \frac{VN}{VN + FP} \quad (4)$$

$$\text{Precisão} = \frac{VP}{VP + FP} \quad (5)$$

$$\text{Revocação} = \frac{VP}{VP + FN} \quad (6)$$

$$F_1 = 2 \times \frac{\text{Precisão} \times \text{Revocação}}{\text{Precisão} + \text{Revocação}} \quad (7)$$

- Bibliotecas
 - NumPy
 - scikit-learn
 - TensorFlow
 - Pandas
 - Matplotlib

Introdução

Conceitos

Trabalhos Relacionados

Materiais e Métodos

Plano de Trabalho

Atividades

Atividades	1 ^o ano						2 ^o ano					
	1	2	3	4	5	6	1	2	3	4	5	6
Etapa 1 - Preparação												
Estudo bibliográfico	•	•	•	•	•	•	•	•				
Seleção de dados				•	•	•						
Pré-processamento					•	•	•					
Etapa 2 - Rede Neural												
Construção da rede					•	•	•	•				
Realização de testes							•	•	•			
Comparação dos resultados									•	•		
Etapa 3 - Conclusão												
Publicação dos resultados											•	
Escrita da dissertação									•	•	•	•
Defesa da dissertação												•

Tabela 2: Cronograma de atividades dividido em bimestres.

1. Estudo e aprofundamento de conhecimento técnico.
2. Seleção da base de dados a ser utilizada.
3. Pré-processamento e tratamento dos dados.

Atividades

Atividades	1 ^o ano						2 ^o ano					
	1	2	3	4	5	6	1	2	3	4	5	6
Etapa 1 - Preparação												
Estudo bibliográfico	•	•	•	•	•	•	•	•				
Seleção de dados				•	•	•						
Pré-processamento					•	•	•					
Etapa 2 - Rede Neural												
Construção da rede					•	•	•	•				
Realização de testes							•	•	•			
Comparação dos resultados									•	•		
Etapa 3 - Conclusão												
Publicação dos resultados											•	
Escrita da dissertação									•	•	•	•
Defesa da dissertação												•

Tabela 2: Cronograma de atividades dividido em bimestres.

4. Construção e refinamento da rede neural.
5. Realização de testes e análise dos resultados.
6. Comparação dos resultados obtidos.

Atividades

Atividades	1 ^o ano						2 ^o ano					
	1	2	3	4	5	6	1	2	3	4	5	6
Etapa 1 - Preparação												
Estudo bibliográfico	•	•	•	•	•	•	•	•				
Seleção de dados				•	•	•						
Pré-processamento					•	•	•					
Etapa 2 - Rede Neural												
Construção da rede					•	•	•	•				
Realização de testes							•	•	•			
Comparação dos resultados									•	•		
Etapa 3 - Conclusão												
Publicação dos resultados											•	
Escrita da dissertação									•	•	•	•
Defesa da dissertação												•

Tabela 2: Cronograma de atividades dividido em bimestres.

8. Documentação e publicação dos resultados.
9. Escrita da dissertação de mestrado.
10. Apresentação da dissertação de mestrado.

- [1] Ana Caroline Gomes Vargas, Aline Marins Paes Carvalho, and Cristina Nader Vasconcelos.
Um estudo sobre redes neurais convolucionais e sua aplicação em detecção de pedestres.
In Conference on Graphics, Patterns and Images (SIBGRAPI), pages 1–4, 2016.
- [2] Lulu Wang.
Deep learning techniques to diagnose lung cancer.
Cancers, 14(22):5569, 2022.

- [3] Mehwish Dildar, Shumaila Akram, Muhammad Irfan, Hikmat Ullah Khan, Muhammad Ramzan, Abdur Rehman Mahmood, Soliman Ayed Alsaari, Abdul Hakeem M Saeed, Mohammed Olaythah Alraddadi, and Mater Hussen Mahnashi.
Skin cancer detection: A review using deep learning techniques.
International Journal of Environmental Research and Public Health, 18(10):5479, 2021.
- [4] Jun Bai, Russell Posner, Tianyu Wang, Clifford Yang, and Sheida Nabavi.
Applying deep learning in digital breast tomosynthesis for automatic breast cancer detection: A review.
Medical Image Analysis, 71:102049, 2021.

- [5] Xaviera A. López-Cortés, Felipe Matamala, Bernardo Venegas, and César Rivera.
Machine-learning applications in oral cancer: A systematic review.
Applied Sciences, 12(11):5715, 2022.
- [6] Lichy Han and Maulik R. Kamdar.
MRI to MGMT: predicting methylation status in glioblastoma patients using convolutional recurrent neural networks.
Pacific Symposium on Biocomputing, pages 331–342, 2018.
- [7] Nguyen Quoc Khanh Le, Duyen Thi Do, Fang-Ying Chiu, Edward Kien Yee Yapp, Hui-Yuan Yeh, and Cheng-Yu Chen.
XGBoost improves classification of MGMT promoter methylation status in IDH1 wildtype glioblastoma.
Journal of Personalized Medicine, 10(3):128, 2020.

- [8] Duyen Thi Do, Ming-Ren Yang, Luu Ho Thanh Lam, Nguyen Quoc Khanh Le, and Yu-Wei Wu.
Improving mgmt methylation status prediction of glioblastoma through optimizing radiomics features using genetic algorithm-based machine learning approach.
Scientific Reports, 12(1), 2022.
- [9] Tran Nguyen Tuan Minh, Viet Huan Le, and Nguyen Quoc Khanh Le.
Diffusion-tensor imaging and dynamic susceptibility contrast MRIs improve radiomics-based machine learning model of MGMT promoter methylation status in glioblastomas.
Biomedical Signal Processing and Control, 86:105122, 2023.

- [10] Kenneth Clark, Bruce Vendt, Kirk Smith, John Freymann, Justin Kirby, Paul Koppel, Stephen Moore, Stanley Phillips, David Maffitt, Michael Pringle, Lawrence Tarbox, and Fred Prior.
The Cancer Imaging Archive (TCIA): Maintaining and operating a public information repository.
Journal of Digital Imaging, 26(6):1045–1057, 2013.
- [11] Lisa Scarpace, Tom Mikkelsen, Soonmee Cha, Sujaya Rao, Sangeeta Tekchandani, David Gutman, Joel H. Saltz, Bradley J. Erickson, Nancy Pedano, Adam E. Flanders, Jill Barnholtz-Sloan, Quinn Ostrom, Daniel Barboriak, and Laura J. Pierce.
The cancer genome atlas glioblastoma multiforme collection (TCGA-GBM).
The Cancer Imaging Archive, 2016.

- [12] Bjoern H. Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, Roland Wiest, Levente Lanczi, Elizabeth Gerstner, Marc-Andre Weber, Tal Arbel, Brian B. Avants, Nicholas Ayache, Patricia Buendia, D. Louis Collins, Nicolas Cordier, Jason J. Corso, Antonio Criminisi, Tilak Das, Herve Delingette, Cagatay Demiralp, Christopher R. Durst, Michel Dojat, Senan Doyle, Joana Festa, Florence Forbes, Ezequiel Geremia, Ben Glocker, Polina Golland, Xiaotao Guo, Andac Hamamci, Khan M. Iftekharuddin, Raj Jena, Nigel M. John, Ender Konukoglu, Danial Lashkari, Jose Antonio Mariz, Raphael Meier, Sergio Pereira, Doina Precup, Stephen J. Price, Tammy Riklin Raviv, Syed M. S. Reza, Michael Ryan, Duygu Sarikaya, Lawrence Schwartz, Hoo-Chang Shin, Jamie Shotton, Carlos A. Silva, Nuno Sousa, Nagesh K. Subbanna, Gabor Szekely, Thomas J. Taylor, Owen M. Thomas,

Nicholas J. Tustison, Gozde Unal, Flor Vasseur, Max Wintermark, Dong Hye Ye, Liang Zhao, Binsheng Zhao, Darko Zikic, Marcel Prastawa, Mauricio Reyes, and Koen Van Leemput.

The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS).

IEEE Transactions on Medical Imaging, 34(10):1993–2024, 2015.

- [13] Ujjwal Baid, Satyam Ghodasara, Suyash Mohan, Michel Bilello, Evan Calabrese, Errol Colak, Keyvan Farahani, Jayashree Kalpathy-Cramer, Felipe C. Kitamura, Sarthak Pati, Luciano M. Prevedello, Jeffrey D. Rudie, Chiharu Sako, Russell T. Shinohara, Timothy Bergquist, Rong Chai, James Eddy, Julia Elliott, Walter Reade, Thomas Schaffter, Thomas Yu, Jiaxin Zheng, Ahmed W. Moawad, Luiz Otavio Coelho, Olivia McDonnell, Elka Miller, Fanny E. Moron, Mark C. Oswood, Robert Y. Shih, Loizos Siakallis,

Referências

Yulia Bronstein, James R. Mason, Anthony F. Miller, Gagandeep Choudhary, Aanchal Agarwal, Cristina H. Besada, Jamal J. Derakhshan, Mariana C. Diogo, Daniel D. Do-Dai, Luciano Farage, John L. Go, Mohiuddin Hadi, Virginia B. Hill, Michael Iv, David Joyner, Christie Lincoln, Eyal Lotan, Asako Miyakoshi, Mariana Sanchez-Montano, Jaya Nath, Xuan V. Nguyen, Manal Nicolas-Jilwan, Johanna Ortiz Jimenez, Kerem Ozturk, Bojan D. Petrovic, Chintan Shah, Lubdha M. Shah, Manas Sharma, Onur Simsek, Achint K. Singh, Salil Soman, Volodymyr Statsevych, Brent D. Weinberg, Robert J. Young, Ichiro Ikuta, Amit K. Agarwal, Sword C. Cambron, Richard Silbergleit, Alexandru Dusoi, Alida A. Postma, Laurent Letourneau-Guillon, Gloria J. Guzman Perez-Carrillo, Atin Saha, Neetu Soni, Greg Zaharchuk, Vahe M. Zohrabian, Yingming Chen, Milos M. Cekic, Akm Rahman, Juan E. Small, Varun Sethi, Christos Davatzikos, John Mongan, Christopher

Hess, Soonmee Cha, Javier Villanueva-Meyer, John B. Freymann, Justin S. Kirby, Benedikt Wiestler, Priscila Crivellaro, Rivka R. Colen, Aikaterini Kotrotsou, Daniel Marcus, Mikhail Milchenko, Arash Nazeri, Hassan Fathallah-Shaykh, Roland Wiest, Andras Jakab, Marc-Andre Weber, Abhishek Mahajan, Bjoern Menze, Adam E. Flanders, and Spyridon Bakas.

The RSNA-ASNR-MICCAI BraTS 2021 benchmark on brain tumor segmentation and radiogenomic classification.

arXiv preprint arXiv:2107.02314, 2021.

- [14] Szymon Grochans, Anna Maria Cybulska, Donata Simińska, Jan Korbecki, Klaudyna Kojder, Dariusz Chlubek, and Irena Baranowska-Bosiacka.

Epidemiology of glioblastoma multiforme—literature review.

Cancers, 14(10), 2022.

- [15] Claudio Luchini, Antonio Pea, and Aldo Scarpa.
Artificial intelligence in oncology: current applications and future perspectives.
British Journal of Cancer, 126(1):4–9, 2021.
- [16] Eyal Lotan, Rajan Jain, Narges Razavian, Girish M. Fatterpekar, and Yvonne W. Lui.
State of the art: Machine learning applications in glioma imaging.
American Roentgen Ray Society, 212(1):26–37, 2019.

- [17] Spyridon Bakas, Hamed Akbari, Aristeidis Sotiras, Michel Bilello, Martin Rozycki, Justin S. Kirby, John B. Freymann, Keyvan Farahani, and Christos Davatzikos.
Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features.
Scientific Data, 4(1), 2017.
- [18] Brenda Auffinger, Drew Spencer, Peter Pytel, Atique U Ahmed, and Maciej S Lesniak.
The role of glioma stem cells in chemotherapy resistance and glioblastoma multiforme recurrence.
Expert Review of Neurotherapeutics, 15(7):741–752, 2015.

- [19] Marinus T. Vlaardingerbroek and Jacques A. den Boer.
MRI and its hardware.
In *Magnetic Resonance Imaging*, page 1–43. Springer Berlin Heidelberg, 1996.
- [20] Oleg S. Pianykh.
What is DICOM?
In *Digital Imaging and Communications in Medicine (DICOM)*, page 3–5. Springer Berlin Heidelberg, 2008.
- [21] Sergio Alonso, Yuichi Dai, Kentaro Yamashita, Shina Horiuchi, Tomoko Dai, Akihiro Matsunaga, Rosa Sánchez-Muñoz, Cristina Bilbao-Sieyro, Juan Carlos Díaz-Chico, Andrei V. Chernov, Alex Y. Strongin, and Manuel Perucho.

Methylation of MGMT and ADAMTS14 in normal colon mucosa: biomarkers of a field defect for cancerization preferentially targeting elder african-americans.

Oncotarget, 6(5):3420–3431, 2015.

[22] Sinno Jialin Pan and Qiang Yang.

A survey on transfer learning.

IEEE Transactions on Knowledge and Data Engineering, 22(10):1345–1359, 2010.

[23] Gerd Pfeifer.

Defining driver dna methylation changes in human cancer.

International Journal of Molecular Sciences, 19(4):1166, 2018.

- [24] Chunlong Zhang, Hongyan Zhao, Jie Li, Hongbo Liu, Fang Wang, Yanjun Wei, Jianzhong Su, Dongwei Zhang, Tiefu Liu, and Yan Zhang.
The identification of specific methylation patterns across different cancers.
PLoS ONE, 10(3):e0120361, 2015.
- [25] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner.
Gradient-based learning applied to document recognition.
Proceedings of the IEEE, 86(11):2278–2324, 1998.
- [26] Saad Albawi, Tareq Abed Mohammed, and Saad Al-Zawi.
Understanding of a convolutional neural network.
In *International Conference on Engineering and Technology (ICET)*, pages 1–6, 2017.

- [27] Phil Kim.
Convolutional neural network.
In MATLAB Deep Learning: With Machine Learning, Neural Networks and Artificial Intelligence, pages 121–147. Apress, 2017.
- [28] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun.
Deep residual learning for image recognition.
arXiv preprint arXiv:1512.03385, 2015.
- [29] Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q. Weinberger.
Densely connected convolutional networks.
arXiv preprint arXiv:1608.06993, 2016.

- [30] Zihang Dai, Hanxiao Liu, Quoc V. Le, and Mingxing Tan.
Coatnet: Marrying convolution and attention for all data sizes.
arXiv preprint arXiv:2106.04803, 2021.
- [31] Xiangrui Li, Paul S. Morgan, John Ashburner, Jolinda Smith, and Christopher Rorden.
The first step for neuroimaging data analysis: Dicom to nifti conversion.
Journal of Neuroscience Methods, 264:47–56, 2016.

- [32] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby.

An image is worth 16x16 words: Transformers for image recognition at scale.

arXiv preprint arXiv:2010.11929, 2020.

- [33] Lucas Beyer, Pavel Izmailov, Alexander Kolesnikov, Mathilde Caron, Simon Kornblith, Xiaohua Zhai, Matthias Minderer, Michael Tschannen, Ibrahim Alabdulmohsin, and Filip Pavetic.

FlexiViT: One model for all patch sizes.

arXiv preprint arXiv:2212.08013, 2022.

- [34] Diederik P Kingma and Max Welling.
Auto-encoding variational bayes.
arXiv preprint arXiv:1312.6114, 2013.

Predição do Estado de Metilação em Pacientes com Glioblastoma Multiforme Utilizando Aprendizado de Máquina

Giulia Silva Fazzi

Orientador: Prof. Dr. Zanoni Dias

Coorientador: Prof. Dr. Hélio Pedrini

18 de Outubro de 2023

Instituto de Computação

Universidade Estadual de Campinas