



Tạp chí Khoa học và Kinh tế Phát triển
Trường Đại học Nam Cần Thơ

Website: jsde.nctu.edu.vn



SỬ DỤNG THUẬT TOÁN BẰNG MÁY DỰA TRÊN RADIOMICS TRONG DỰ ĐOÁN MỨC ĐỘ ĐỘT BIẾN TRONG U Ở BỆNH NHÂN U THẦN KINH ĐỆM PHÂN ĐỘ THẤP

Lưu Hồ Thanh Lâm, Chu Thy Ngân, Trần Thị Oanh, Đỗ Thị Duyên, Nguyễn Lê Quốc Khánh

Ngày nhận bài: 7/7/2023

Ngày duyệt bài: 17/7/2023

ABSTRACT

Lower-grade glioma (LGG) is a type of central nervous system neoplasm that arises from the glial cells. Patients with Lower-grade glioma have a median survival time in the range of 1.5–8 years, depending on the tumor genotypes. In terms of epidemiology, most lower-grade glioma patients are diagnosed at a young age, leading to an early age of death. For an accurate diagnosis and effective treatment, a pathological result from a biopsy sample is required, however, this has a long turnaround time. In this study, using pre-operative magnetic resonance images, we developed a non-invasive model to classify tumor mutational burden (TMB), a prognostic factor of treatment response in lower-grade glioma patients, with an accuracy of 0.7936. To our knowledge, our study represented the best model for the classification of TMB in LGG patients at present.

TÓM TẮT

Glioma là một khối u của Hệ thần kinh trung ương (CNS) phát sinh từ các tế bào thần kinh đệm. U thần kinh đệm phân độ độ thấp (LGG) là u thần kinh đệm ở độ II và III trong bảng phân độ mới của Tổ chức Y tế Thế giới năm 2016. Qua quá trình ức chế điều hoà miễn dịch ngược, liệu pháp miễn dịch đang hứa hẹn là một phương pháp điều trị hiệu quả cho bệnh nhân u thần kinh đệm. Tuy nhiên, phương pháp này không thể áp dụng tốt cho tất cả các loại LGG và mức độ đột biến trong u (TMB) đã được chứng minh là một dấu ấn sinh học tiềm năng cho tính nhạy cảm và tiên lượng

của liệu pháp miễn dịch ở bệnh nhân u thần kinh đệm phân độ thấp. Do đó, dự đoán TMB có lợi cho bệnh nhân ung thư thần kinh đệm. Trong nghiên cứu này, chúng tôi khảo sát mối tương quan của các tính năng radiomics dựa trên MRI (Chụp cộng hưởng từ) và đột biến trong u ở bệnh nhân u thần kinh đệm phân độ thấp bằng cách áp dụng các phương pháp tính toán dùng máy. Sau khi kiểm tra thống nhất giữa dữ liệu xác thực và dữ liệu sẵn có, chúng tôi đã sử dụng sáu loại thuật toán kết hợp các mô hình để tìm các tính năng hiệu quả nhất và so sánh các mô hình để chọn hiệu suất tốt nhất. Sự kết hợp Light Gradient Boosting Machine- thuật toán di truyền đã thành công trong việc chọn chữ ký 11-radiomics cho phân loại TMB. Chúng tôi đã giải quyết vấn đề dữ liệu mất cân bằng bằng cách sử dụng kỹ thuật SVMOTE trong Python và đạt được kết quả ấn tượng. Mô hình LightGBM dẫn đến độ chính xác cao 0,7936 và đạt được sự cân bằng giữa độ nhạy và độ đặc hiệu, với độ nhạy 0,76 và độ đặc hiệu 0,8107. Theo hiểu biết của chúng tôi, nghiên cứu này có số độ chính xác cao nhất trong các mô hình mô hình hiện nay để phân loại TMB ở bệnh nhân LGG.

1. GIỚI THIỆU

Trong nhiều thập kỷ, việc phân loại ung thư thường dựa trên các đặc điểm mô học, thông qua quan sát bằng kính hiển vi, sử dụng thuốc nhuộm hematoxylin và eosin, biểu hiện protein hóa mô miễn dịch và các đặc điểm siêu cấu trúc. Về mặt mô học, u thần kinh đệm là khối u của hệ thần kinh trung ương (CNS) phát sinh từ các tế bào không phải tế bào thần kinh, bao gồm các kiểu hình u tế bào hình sao, u tế bào ít sợi, u thần kinh đệm ít nhánh và u nguyên bào thần kinh đệm [1]. Mong muốn có một chẩn đoán phân biệt hẹp hơn và phân loại chính xác hơn, Tổ chức Y tế Thế giới (WHO) đã sửa đổi phân loại u thần kinh đệm năm 2007 vào năm 2016 bằng cách đưa các thông số phân tử vào sơ đồ [2]. Sơ đồ cập nhật phân loại u thần kinh đệm thành bốn

nhóm: Cấp I đến cấp IV theo mức độ ác tính trong khối u. Trong khi trước đây, cụm từ “u thần kinh đệm cấp thấp” dùng để chỉ u thần kinh đệm cấp I và II, một thuật ngữ mới trong chương trình mới, “u thần kinh đệm phân độ thấp (LGGs)”, bao gồm cấp độ II và III của WHO, đang dần trở nên phổ biến hơn. Phân loại u thần kinh đệm phân độ thấp đóng một vai trò thiết yếu trong tiên lượng của bệnh nhân [3]. Mặc dù phân loại LGG của WHO năm 2016 đã được áp dụng để chẩn đoán phân tử, nhưng các dấu hiệu phân tử đã biết vẫn còn hạn chế để giải thích tiên lượng của LGG. Do đó, việc khám phá thêm về cơ chế di truyền và xác định các dấu ấn sinh học mới để dự đoán tiên lượng của LGG là điều cần thiết để phát triển các phương pháp điều trị chính xác.

Một điều đáng chú ý là giải thưởng Nobel 2018 đã được trao cho James Allison và Tasuku Honjo trong lĩnh vực Sinh lý học hoặc Y học vì đã phát triển phương pháp điều trị bệnh ác tính [4], liệu pháp miễn dịch, sẽ trở thành phương pháp điều trị ung thư được sử dụng rộng rãi. Tuy nhiên, dấu ấn sinh học về đột biến trong u (TMB) đã được chứng minh là một yếu tố tiên đoán đáp ứng với liệu pháp miễn dịch [5]. TMB, được định nghĩa là số lượng đột biến mã hóa soma trên mỗi megabase (đột biến/Mb) của kiểu gen khối u [6], đã trở thành một dấu ấn sinh học có giá trị trên nhiều loại ung thư để dự đoán hiệu quả của phong tỏa điểm kiểm tra miễn dịch (ICB) [5],[7]. Một nghiên cứu trước đây cho thấy TMB có mối tương quan nghịch với tỷ lệ sống sót chung và TMB cao có thể ức chế sự xâm nhập miễn dịch trong LGGs [8]. Tương tự, Wang et al. [9] kết luận rằng TMB có liên quan đến kết quả kém trong u thần kinh đệm lan tỏa và TMB cao kích hoạt cả hoạt động tăng sinh và phản ứng miễn dịch trong u thần kinh đệm.

Trình tự toàn bộ exome (WES) được coi là tiêu chuẩn vàng để đo lường TMB; tuy nhiên, nó hiện không được sử dụng trong môi trường lâm sàng do chi phí cao và thời gian quay vòng dài. Phát triển các mô hình trí tuệ nhân tạo để chẩn đoán chính xác bệnh hay phân loại dấu ấn miễn dịch, trong đó có TMB, giúp tiết kiệm thời gian, giảm chi phí là xu hướng nổi bật trong thế giới hiện đại.

Trong nghiên cứu của chúng tôi, chúng tôi đã phát hiện ra một phương pháp đại diện cho mô hình rất tốt để phân loại TMB cho LGG hiện tại. Nghiên cứu của chúng tôi đưa ra giả thuyết rằng các thuật toán học bằng máy có thể phân

loại TMB cao và thấp bằng cách sử dụng phương pháp chụp cộng hưởng từ (MRI) ở bệnh nhân LGG. Cụ thể, chúng tôi đã đề xuất một mô hình học nhân tạo, có tên là LightGBM, có thể xác định các nhóm TMB bằng cách sử dụng MRI trước phẫu thuật.

Để cải thiện hiệu suất của mô hình, sau đó chúng tôi đã sử dụng thuật toán di truyền để chọn các tính năng radiomics đặc trưng, hiệu quả nhất để xác định TMB cao và thấp, đồng thời đưa chúng vào các thuật toán khác nhau để tìm ra mô hình tốt nhất. Sau đó, chúng tôi đã sử dụng các kỹ thuật giải dữ liệu mất cân bằng khác nhau trên dữ liệu của mình để đạt được sự cân bằng tốt giữa độ nhạy và độ đặc hiệu, cho kết quả chính xác và cân bằng nhất hiện nay. Mặc dù nghiên cứu của chúng tôi có cỡ mẫu nhỏ, nhưng kết quả của chúng tôi cho thấy có thể sử dụng mô hình máy học sử dụng hình ảnh MRI để hỗ trợ các tình huống lâm sàng.

2. KẾT QUẢ

2.1 Đặc điểm bệnh nhân

Dữ liệu của bệnh nhân bao gồm tuổi, giới tính, mô học, cấp độ WHO, và các phân nhóm của phân loại u thần kinh đệm bằng cách tích hợp thông tin di truyền. Theo dịch tễ học của u thần kinh đệm phân độ thấp, hầu hết bệnh nhân của chúng tôi còn trẻ tại thời điểm chẩn đoán, với độ tuổi trung bình lần lượt là 44,06 và 49,67. Sự khác biệt về tuổi giữa dữ liệu đào tạo và xác nhận có ý nghĩa thống kê với $p < 0,05$. Ngoài ra, hầu hết bệnh nhân của chúng tôi có đột biến IDH (84 đột biến so với 21 kiểu hoang dã) và trạng thái không mã hóa 1p_19q (78 không mã hóa so với 27 mã hóa) trong thông tin bộ gen của họ.

2.2 So sánh cơ sở và xây dựng đặc tính radiomics

Trong quy trình lựa chọn tính năng, chúng tôi đã sử dụng các mô hình thuật toán di truyền (GA) kết hợp với loại máy khác nhau để chọn các tính năng hiệu quả nhất, được gọi là tính năng GA, từ 726 tính năng trong giai đoạn trích xuất. Sau đó, các tính năng GA này được đánh giá bởi sáu mô hình máy học khác nhau (hồi quy tuyến tính—LR, vector hỗ trợ—SVM, ngẫu nhiên—RF, phân tích phân biệt tuyến tính—LDA, máy tăng cường độ dốc ánh sáng—LGBM và máy tăng cường độ dốc cực cao—XGB) trong việc dự đoán nhóm TMB. Kết quả lựa chọn tính năng GA và hiệu suất mô hình được đưa ra trong Bảng 2.

Như thể hiện trong Bảng 2, mô hình LGBM đạt hiệu suất tốt nhất với độ nhạy = 0,72, độ đặc hiệu = 0,8893, độ chính xác = 0,8367 và độ chính xác = 0,8218, với 11 tính năng GA. Mô hình LGBM không chỉ đạt độ chính xác cao nhất trong 6 thuật toán máy học dựa trên GA khi dự đoán nhóm TMB mà còn thể hiện sự cân bằng giữa độ nhạy và độ đặc hiệu. Một phiên bản cũ hơn của thuật toán LGBM, thuật toán XGBoost, cũng cho thấy độ chính xác cao là 0,7808 với 7 tính năng GA trong thử nghiệm của chúng tôi. Mô hình RF mang lại hiệu suất tốt thứ hai với độ chính xác 0,8089 với 6 tính năng GA. LR cho kết quả tốt với độ chính xác 0,7936 và độ chính xác 0,836667 của nó bằng với mẫu tốt nhất, LGBM. Hiệu suất kém nhất đến từ hai mô hình

còn lại, SVM và LDA, với độ chính xác lần lượt là 0,7462 và 0,7449; tuy nhiên, đây là mức chấp nhận được đối với thử thách phân loại. Chúng tôi cũng đã thực hiện các thử nghiệm thống kê McNemar để xem sự khác biệt đáng kể giữa mô hình đã chọn của chúng tôi (LightGBM) và các mô hình khác. Kết quả sau đó cho thấy LightGBM tốt hơn đáng kể so với các thuật toán khác về độ nhạy. Hơn nữa, độ chính xác của nó cũng vượt trội so với các mẫu SVM, LDA và XGBoost. Đối với hai chỉ số khác (độ đặc hiệu và độ chính xác), không có bất kỳ sự khác biệt đáng kể nào giữa tất cả các thuật toán. Tuy nhiên, các thử nghiệm thống kê này cho thấy LightGBM là lựa chọn tối ưu cho nhiệm vụ dự đoán này với hiệu suất cao. Vì mô hình LightGBM là mô hình tốt nhất với 11 tính năng tối ưu, được GA chọn, chúng tôi đã coi 11 tính năng GA này là đặc trưng radiomics của dự đoán TMB trong nghiên cứu của chúng tôi.

3. KẾT LUẬN

Chúng tôi đã nghiên cứu các đặc trưng radiomics dựa trên MRI trong mối quan hệ với gánh nặng đột biến khối u ở u thần kinh đệm phân độ thấp, mô hình của chúng tôi là mô hình đầu tiên áp dụng phương pháp tính toán bằng máy. Bằng cách bắt chước mô hình của sự phát triển, máy LightGBM-GA đã thu được 11 đặc tính radiomics để phân loại TMB. Mô hình LightGBM của chúng tôi đã đạt được kết quả đầy hứa hẹn với độ chính xác cao là 0,7936 và sự cân bằng giữa độ nhạy và độ đặc hiệu (trung ứng là 0,76 và 0,8107).

Table 1. Comparative performance among different GA-based machine learning algorithms in predicting the TMB group

Algorithm	GA Features	Sensitivity	Specificity	Precision	Accuracy	Running Time (s)
LR	13	0.64 ± 0.265	0.9 ± 0.079	0.8367 ± 0.162	0.7936 ± 0.132	0.159057
SVM	10	0.56 ± 0.15	0.8714 ± 0.177	0.7733 ± 0.228	0.7462 ± 0.133	0.050138
RF	6	0.64 ± 0.15	0.9179 ± 0.064	0.8833 ± 0.108	0.8089 ± 0.041	0.817011
LDA	4	0.56 ± 0.16	0.8714 ± 0.131	0.77 ± 0.131	0.7449 ± 0.056	0.125044
LGBM	11	0.72 ± 0.204	0.8893 ± 0.131	0.8367 ± 0.131	0.8218 ± 0.1	0.094271
XGB	7	0.6 ± 0.204	0.9 ± 0.009	0.8 ± 0.106	0.7808 ± 0.08	1.63018

GA: genetic algorithm, logistic regression (LR), random forest (RF), support vector machine (SVM), linear discriminant analysis (LDA), light gradient boosting machine (LGBM), extreme gradient boosting (XGB).

TÀI LIỆU THAM KHẢO

[1] Louis, D.N. (2007). *WHO Classification of Tumours of the Central Nervous System*; WHO Regional Office Europe: Copenhagen, Denmark. [Google Scholar].

[2] Louis, D.N., Perry, A., Reifenberger, G., Von Deimling, A., Figarella-Branger, D., Cavenee, W.K., Ohgaki, H., Wiestler, O.D., Kleihues, P., & Ellison, D.W. (2016). The 2016 World Health Organization classification of tumors of the central nervous system: A summary. *Acta Neuropathol.* 131, 803–820. [Google Scholar] [CrossRef] [PubMed][Green Version].

[3] The Cancer Genome Atlas Research Network. (2015). Comprehensive, integrative genomic analysis of diffuse lower-grade gliomas. *N. Engl. J. Med.* 372, 2481–2498. [Google Scholar] [CrossRef] [PubMed][Green Version].

[4] Ballas, Z.K. (2018). The 2018 Nobel Prize in Physiology or Medicine: An exemplar of bench to bedside in immunology. *J. Allergy Clin. Immunol.* 142, 1752–1753. [Google Scholar] [CrossRef] [PubMed][Green Version].

[5] Chan, T.A., Yarchoan, M., Jaffee, E., Swanton, C., Quezada, S.A., Stenzinger, A., & Peters, S. (2019). Development of tumor mutation burden as an immunotherapy biomarker: Utility for the oncology clinic. *Ann. Oncol.* 30, 44–56. [Google Scholar] [CrossRef].

[6] Chalmers, Z.R., Connelly, C.F., Fabrizio, D., Gay, L., Ali, S.M., Ennis, R., Schrock, A., Campbell, B., Shlien, A., & Chmielecki, J. (2017). Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational

- burden. *Genome Med.* 9, 34. [Google Scholar] [CrossRef][Green Version].
- [7] Braun, D.A., Burke, K.P., & Van Allen, E.M. (2016). Genomic approaches to understanding response and resistance to immunotherapy. *Clin. Cancer Res.* 22, 5642–5650. [Google Scholar] [CrossRef][Green Version].
- [8] Yin, W., Jiang, X., Tan, J., Xin, Z., Zhou, Q., Zhan, C., Fu, X., Wu, Z., Guo, Y., & Jiang, Z. (2020). Development and Validation of a Tumor Mutation Burden–Related Immune Prognostic Model for Lower-Grade Glioma. *Front. Oncol.* 10, 1409. [Google Scholar] [CrossRef].
- [9] Wang, L., Ge, J., Lan, Y., Shi, Y., Luo, Y., Tan, Y., Liang, M., Deng, S., Zhang, X., & Wang, W. (2020). Tumor mutational burden is associated with poor outcomes in diffuse glioma. *BMC Cancer.* 20, 213. [Google Scholar] [CrossRef].