Machine Learning Model to Identify Prognostic Factors in Glioblastoma: A SEER-Based Analysis

Batuhan Bakirarar

Ankara University

Emrah Egemen

Pamukkale University School of Medicine: Pamukkale Universitesi Tıp Fakültesi

FATİH YAKAR (✉ yakarneurosurgery@gmail.com)

Pamukkale University School of Medicine: Pamukkale Universitesi Tıp Fakültesi

<https://orcid.org/0000-0001-7414-3766>

Research Article

Keywords: Machine learning, Big Data, Glioblastoma, SEER

Posted Date: February 10th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1327181/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

The aim of this study to create a competing risk model to identify prognostic factors in glioblastoma (GB). The study included 31663 patients diagnosed with GB between 2007 and 2018. The data in the study were taken from the Surveillance, Epidemiology, and End Results (SEER) database. Overall survivals (OS), age, race, gender, primary site, laterality, surgery and tumor size at the time of diagnosis, vital status, and follow-up time (months) were selected for the analyzes. The median OS of the patients was found to be 9.00 ± 0.09 months. In addition, all variables in the table were statistically significant risk factors for survival except gender. Therefore, surgery, age, laterality, primary site, tumor size, race, gender variables were used as independent risk factors, and vital status was used as a dependent variable for ML analysis. Looking at the ML results, Hybrid Model gave the best results according to Accuracy, F-measure, and MCC performance criteria. According to hybrid model, which has the best performance, the diagnosis of alive/dead in 84 and 74 out of 100 patients can be interpreted as correct for 1- and 2-year, respectively. Recognition of the fundamental ideas will allow neurosurgeons to understand BD and help evaluate the extraordinary amount of data within the associated healthcare field.

Introduction

Science and industry have an extraordinary data production in our age. Traditional statistical approaches are not sufficient in the analysis and interpretation of Big Data (BD). Machine learning (ML) and artificial intelligence methods have become essential in the perception of these data [17]. The BD analysis supports the storage, classification, and analysis of patient information in the healthcare field and improves disease identification, treatment evaluation, surgical planning, and outcome prediction [50]. Hidden patterns in large datasets can be revealed by BD analysis [15].

In adults, the most common primary malign brain tumor is glioblastoma (GB) [33]. Surgical resection, adjuvant external beam radiation therapy, plus concurrent and adjuvant temozolomide is the standard management of newly diagnosed high-grade gliomas (HGGs) [43, 44]. The median survival in patients with this protocol was 14.6 months [44], and 5-year survival is 5% despite aggressive therapies [19, 31, 47]. The independent prognostic factors for progression-free survival (PFS) and overall survival (OS) confirm age, preoperative performance status, and tumor size [10]. MGMT promoter methylation was added to these factors in a recent systematic review [14].

This study extracted 31663 patients with histologically confirmed GB from Surveillance, Epidemiology and End Results (SEER) database. This study aims to create a competing risk model to identify prognostic factors in GB.

Materials And Methods

Study Design

The study included 31663 patients diagnosed with GB between 2007 and 2018, and all patient data were analyzed for the study. January 2007 was chosen as the starting point for the study, and December 2018 was selected as the end date of the study. The data in the study were taken from the SEER database. These data, published by the National Cancer Center Institute, are a compilation of databases of 18 SEER cancer registries in the USA. The SEER program is used to summarize data from patients' medical records. It is estimated that more than 95% of all cancer cases are detected and included in this database in areas under surveillance [45]. The duration of follow-up is calculated in months using the date of diagnosis and whichever occurs first, 1) date of death, 2) date last known to be alive, 3) December 2018 (the follow-up cutoff date used in our analysis). Since all patient data were obtained with the permission of SEER without including personal patient information, there is no need to get ethical committee approval from any committee within the scope of this research.

The main hypothesis in the study was OS in years (censored observations), defined from the date of diagnosis to the date of death or, for living patients, the last control date. In addition to survival, other variables selected for the analyzes were age, race, gender, primary site, laterality (unilateral/bilateral), surgery and tumor size at the time of diagnosis, vital status, and follow-up time (months). Surgical methods, radiotherapy, and chemotherapy techniques were not included in the study because of missing data.

In this study, in addition to the classical ML methods, we created a hybrid model consisting of a combination of existing methods. Such hybrid models have been preferred more in recent years, as they are a combination of ML methods and use the most substantial aspects of these methods. For 2-year survival prediction model, we used J48, Multilayer Perceptron and Naïve Bayes to create a hybrid model. For 1-year survival prediction model, we used J48, Multilayer Perceptron and Logistic Regression to create a hybrid model.

Statistical Analysis

SPSS 11.5 and Weka 3.7 programs were used in the analysis of the data. Mean±standard deviation and median (minimum-maximum) were used as descriptors for quantitative variables, and the number of patients (percentage) for qualitative variables. Survival analyzes on qualitative variables were performed using the Kaplan-Meier method, and significant differences between groups were determined using the log-rank test. The statistical significance level was taken as 0.05.

Classification methods of Logistic Regression, Naive Bayes, Multilayer Perceptron, Bagging, and J48 were used in the WEKA program. The data set was evaluated using the 10-fold Cross-Validation test option. Accuracy, F-Measure, Matthews correlation coefficient (MCC), Precision-Recall Curve (PRC Area), and Receiver Operating Characteristic (ROC) Area were used as data mining performance criteria.

Results

General descriptors of the variables in the data set are given in Table 1. According to descriptors, 1.1% of the patients were younger than 19 years old or equal, 7.0% were in the 20-44 age range, 42.3% were in the 45-64 age range, and 49.6% were 65 years old or older. While 88.8% of the patients were White, 5.8% were Black, and 5.3% were from other races. In addition, the male-female ratio was 58.4% / 41.6%. The table shows the primary site, laterality, and surgery information of the patients. Tumor sizes of the patients are also grouped, and the patients' vital status and follow-up periods are given (Table 1).

Table 1
Description of the variables in the data for patients with glioblastoma

Variables		
Age, n (%)	≤19 years	343 (1.1)
	20-44 years	2208 (7.0)
	45-64 years	13403 (42.3)
	≥65 years	15709 (49.6)
Race, n (%)	White	28127 (88.8)
	Black	1849 (5.8)
	Other	1687 (5.3)
Gender, n (%)	Male	18479 (58.4)
	Female	13184 (41.6)
Primary Site, n (%)	Frontal Lobe	10113 (31.9)
	Temporal Lobe	8936 (28.2)
	Parietal Lobe	5490 (17.3)
	Occipital Lobe	1461 (4.6)
	Ventricle	154 (0.5)
	Cerebellum	273 (0.9)
	Brain Stem	201 (0.6)
	Overlapping Lesion of Brain	5696 (19.8)
Laterality, n (%)	Unilateral	31023 (98.0)
	Bilateral	640 (2.0)
Surgery, n (%)	Not Performed	6414 (20.3)
	Performed	25249 (79.7)
Tumor Size, n (%)	Less than 1 cm	170 (0.6)
	Between 1 cm and 2 cm	1291 (4.7)
	Between 2 cm and 3 cm	3329 (12.2)
	Between 3 cm and 4 cm	5117 (18.8)
SD: Standard Deviation, Min.:Minimum, Max: Maximum		

Variables		
	Between 4 cm and 5 cm	7336 (27.0)
	Greater than 5 cm	9976 (36.7)
Follow-up Time (months)	Mean±SD	13.21±17.14
	Median (Min.-Max.)	8.00 (0.00-143.00)
Vital Status, n (%)	Alive	4409 (13.9)
	Dead	27254 (86.1)
SD: Standard Deviation, Min.:Minimum, Max: Maximum		

Table 2 shows the survival analysis results of the patients. The median OS of the patients was found to be 9.00±0.09 months. In addition, all variables in the table were statistically significant risk factors for survival except gender. Median life expectancy was found to be 16.00±0.93 months for those younger than or equal to 19 years of age, 22.00±0.58 months for 20-44 years old, 14.00±0.14 months for 45-64 years old, and 5.00±0.07 months for over 65 years old. When evaluated in terms of race, the median life expectancy was 9.00±0.10 months for the White race, and 10.00±0.39 months and 12.00±0.47 months for the Black and other races, respectively. In the study, the median life expectancy of women was equal to that of men.

Table 2
Kaplan-Meier results (SE: Standard error) of the study

Variables		Survival					
		1 year (%)	3 year (%)	5 year (%)	Survival Time		P value
					Mean±SE	Median±SE	
Overall		40.5	10.2	5.2	17.03±0.17	9.00±0.09	-
Age	≤19 years	56.9	22.8	14.7	33.99±2.75	16.00±0.93	<0.001
	20-44 years	72.7	32.6	20.2	39.50±1.11	22.00±0.58	
	45-64 years	53.5	13.0	6.5	21.03±0.27	14.00±0.14	
	≥65 years	24.4	4.5	1.8	10.09±0.14	5.00±0.07	
Race	White	39.8	9.9	5.1	16.76±0.18	9.00±0.10	<0.001
	Black	42.9	11.9	6.2	18.26±0.71	10.00±0.39	
	Other	48.9	14.6	6.8	19.96±0.76	12.00±0.47	
Gender	Male	40.8	9.8	4.7	16.60±0.21	10.00±0.12	0.544
	Female	42.0	10.8	5.9	17.64±0.28	10.00±0.15	
Primary Site	Frontal Lobe	39.9	11.3	5.9	17.87±0.32	9.00±0.16	<0.001
	Temporal Lobe	45.4	10.6	5.0	17.69±0.30	11.00±0.17	
	Parietal Lobe	40.7	9.7	5.1	17.01±0.40	9.00±0.22	
	Occipital Lobe	43.2	9.9	5.0	16.92±0.70	10.00±0.40	
	Ventricle	34.5	11.7	6.1	18.20±2.74	6.00±1.05	
	Cerebellum	37.8	10.3	5.4	16.52±1.79	6.00±0.78	
	Brain Stem	35.7	10.3	6.7	16.60±2.01	8.00±0.84	
	Overlapping Lesion of Brain	32.4	8.2	4.2	14.06±0.37	6.00±0.20	
Laterality	Unilateral	40.8	10.3	5.2	17.11±0.17	9.00±0.09	<0.001
	Bilateral	26.1	7.9	4.2	12.74±1.03	5.00±0.43	
Tumor Size	Less than 1 cm	50.2	15.3	6.6	19.85±2.35	12.00±0.97	<0.001
	Between 1 cm and 2 cm	48.7	14.8	6.3	19.11±0.83	12.00±0.41	
	Between 2 cm and 3 cm	46.4	12.3	5.4	18.85±0.52	11.00±0.30	

	Between 3 cm and 4 cm	42.1	10.3	5.3	17.52±0.42	10.00±0.22	
	Between 4 cm and 5 cm	41.9	9.8	5.0	17.06±0.33	10.00±0.20	
	Greater than 5 cm	36.5	9.6	4.7	16.10±0.30	8.00±0.15	
Surgery	Not Performed	14.4	3.0	1.3	7.16±0.21	3.00±0.05	<0.001
	Performed	47.0	12.1	6.2	19.53±0.20	11.00±0.10	

When survival is evaluated in primary site types, the lowest median survival time is found in the group classified as ventricle, cerebellum, and overlapping brain lesion, followed by the brain stem, parietal, frontal, occipital, and temporal lobes, respectively. Survival statistics for laterality, tumor size, and surgery are also given in Table 2.

Gain Ratio Attribute Evaluation and Information Gain Attribute Evaluation attribute selection methods in WEKA were used. Using these methods, the importance of the variables and the values added to the data set were examined for last 2-year (2017-2018). A total of 8 variables (7 independent variables and one dependent variable) were used from the data set. These variables are surgery, age, laterality, primary site, tumor size, race, gender, and vital status. Percentages of variable importance according to the dependent variable vital status were given in Figure 1A. For 1-year data set, a total of 8 variables (7 independent variables and 1 dependent variable) used. These variables are surgery, age, laterality, primary site, tumor size, race, gender and vital status. Percentages of variable importance according to dependent variable vital status was given in Figure 1B.

The performance criteria of ML Methods for the 2-year survival prediction model are given in Table 3. Looking at the ML results, the Hybrid Model gave the best results according to Accuracy, F-measure, and MCC performance criteria, which are the most accepted criteria in the literature. Considering these three performance criteria, the Hybrid model is followed by J48, Naïve Bayes, Logistic Regression, Bagging, and Multilayer Perceptron, respectively. According to the hybrid model, which has the best performance, the diagnosis of alive/dead in 74 out of 100 patients can be interpreted as correct. As another explanation, when a patient is diagnosed as alive/dead with the hybrid model method, the accuracy rate of this diagnosis is 74.1%.

Table 3
Performance results of Machine Learning methods for 2-year survival

Methods		Performance Criteria				
		Accuracy	F-measure	MCC	PRC Area	ROC Area
Logistic Regression	Alive	0.589	0.613	0.272	0.648	0.681
	Dead	0.682	0.657	0.272	0.688	0.681
	Overall	0.636	0.636	0.272	0.668	0.681
Naive Bayes	Alive	0.591	0.614	0.272	0.648	0.682
	Dead	0.680	0.657	0.272	0.689	0.682
	Overall	0.637	0.636	0.272	0.669	0.682
Multilayer Perceptron	Alive	0.648	0.618	0.218	0.622	0.653
	Dead	0.570	0.598	0.218	0.660	0.653
	Overall	0.608	0.608	0.218	0.641	0.653
Bagging	Alive	0.601	0.611	0.250	0.639	0.668
	Dead	0.649	0.639	0.250	0.676	0.668
	Overall	0.626	0.625	0.250	0.658	0.668
J48	Alive	0.568	0.607	0.279	0.629	0.664
	Dead	0.708	0.668	0.279	0.647	0.664
	Overall	0.640	0.638	0.279	0.638	0.664
Hybrid Model	Alive	0.698	0.725	0.481	0.714	0.764
	Dead	0.781	0.755	0.481	0.793	0.764
	Overall	0.741	0.740	0.481	0.754	0.764
MCC: Matthews correlation coefficient, PRC: Precision Recall Curve, ROC: Receiver Operating Characteristic						

The performance criteria of ML methods for the 1-year survival prediction model are given in Table 4. Looking at the ML results, the Hybrid Model gave best results according to Accuracy, F-measure and MCC performance criteria, which are the most accepted performance criteria in the literature. Considering these three performance criteria, the Hybrid model is followed by J48, Naïve Bayes, Logistic Regression, Bagging and Multilayer Perceptron, respectively. According to the hybrid model which has the best performance, the diagnosis of alive/dead in 85 out of 100 patients can be interpreted as correct. As another explanation, when a patient is diagnosed as alive/dead with the hybrid model method, the accuracy rate of this diagnosis is 84.9%.

Table 4
Performance results of Machine Learning methods for 1-year survival

Methods		Performance Criteria				
		Accuracy	F-measure	MCC	PRC Area	ROC Area
Logistic Regression	Alive	0.927	0.816	0.297	0.814	0.704
	Dead	0.295	0.409	0.297	0.548	0.704
	Overall	0.719	0.682	0.297	0.726	0.704
Naive Bayes	Alive	0.918	0.814	0.297	0.815	0.704
	Dead	0.312	0.422	0.297	0.543	0.704
	Overall	0.718	0.685	0.297	0.725	0.704
Multilayer Perceptron	Alive	0.877	0.796	0.257	0.776	0.665
	Dead	0.340	0.427	0.257	0.506	0.665
	Overall	0.700	0.675	0.257	0.687	0.665
Bagging	Alive	0.914	0.812	0.292	0.810	0.704
	Dead	0.313	0.421	0.292	0.540	0.704
	Overall	0.716	0.683	0.292	0.721	0.704
J48	Alive	0.938	0.818	0.301	0.722	0.609
	Dead	0.281	0.399	0.301	0.468	0.609
	Overall	0.721	0.680	0.301	0.638	0.609
Hybrid Model	Alive	0.941	0.893	0.647	0.958	0.856
	Dead	0.661	0.742	0.647	0.698	0.856
	Overall	0.849	0.843	0.647	0.872	0.856
MCC: Matthews correlation coefficient, PRC: Precision Recall Curve, ROC: Receiver Operating Characteristic						

Discussion

Many studies [5, 6, 9, 11, 13, 22, 23, 26, 28, 30, 36, 40–42, 48, 54] investigate prognosis and survival in GBs using the SEER database. The main difference of our study is that it processes data created following the last two World Health Organisation (WHO) classifications and creates a high-performance model that predicts 1- and 2-year survival using ML.

The overall median survival of our study was 9.00 ± 0.09 months. It is quite a short time compared to the literature, but the main reason is that 49.6% of the patient group in our study was 65 years and older. Less than 20% of elderly GB patients survive up to 1 year, with median survival between 5 and 9 months [21, 38]. Survival may differ according to race and ethnicity in patients diagnosed with GB [3]. The incidence of GB was higher in the White population than others in our study, and it is consistent with previous publications [29, 32, 34, 44]. Survival in the White race was lower than in the other races, as in the analysis by Qstrom et al. [34]. Although some publications are stating that survival is higher in the female gender [34, 44, 48], no significant relationship was found between gender and survival in our study.

There is no consensus on whether tumor location is a prognostic factor. In a recent study [12], GBs' survival in the central core (basal ganglia, corpus callosum) and left temporal lobe pole was less than six months. The survival of the dorsomedial right temporal lobe GBs was more than 24 months. In our study, the temporal lobe tumors' survival was the highest, but no comparison was made in the right or left hemispheres. The prognosis of ventricular [4, 25, 51], brainstem [24], and bilateral hemispheric [8] HGGs are poor, and the results of our study are similar. Although some authors state that cerebellar GBs are worse, comparable, or better than supratentorial ones [1, 2, 6, 18, 27], cerebellar GBs had significantly improved lower survival in our study.

Liu et al. [26] stated that tumor size over 5.4 cm in the SEER database between 2007 and 2016 in patients over 65 years of age is an independent risk factor for GB-related deaths. The larger the FLAIR-T2 hyperintensity volume correlates with, the worse OS and PFS prediction [35]. In our study, the survival of tumors larger than 5 cm was the shortest.

Despite the existence of different treatment modalities, the management of GBs remains a challenge [20]. Although there is no consensus on the limits of surgery in the literature [16, 20] when the maximal surgical resection of abnormal tissue (including FLAIR signal) is safe, it optimizes the patient survival [52]. In our study, the survival of patients who underwent surgical resection was significantly higher.

Various survival predicting models created with the ML method has been published [7, 37, 39, 46, 49, 53], and a recent systematic review reported that the accuracy of these studies was in the range of 0.66–0.98 [46]. The success of our model to predict 1- and 2-year survival was 0.849 and 0.741, respectively.

There are some limitations to this study. There are many subclassifications for each variable when creating data stored in online databases. The authors who process the data can combine or narrow these subsets to the extent they choose for the years they will evaluate. For this reason, different results can be obtained using the same database. The clusters we created in our study are a similar limitation.

Conclusions

Age, race, gender, tumor site/laterality/size, and surgical resection are independent survival risk factors in the analysis performed on 31633 patients between 2007-2018 in the SEER database. The model created by ML was 84.9% and 74.1% successful in predicting 1- and 2-year survival in GB patients, respectively.

Recognition of the fundamental ideas will allow neurosurgeons to understand BD and help assimilate and evaluate the extraordinary amount of data within the associated healthcare field.

Declarations

Acknowledgements

Funding Not applicable

Conflicts of interest The authors have no relevant financial or non-financial interests to disclose

Availability of data and material The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Code availability Not applicable

Ethics approval Not applicable

Consent to participate Not applicable

Consent for publication Not applicable

Authors' contributions B.B. conceived and designed the analysis, collected the data, performed the analysis. B.B and F.Y. wrote the paper. E.E. revised manuscript critically for important intellectual content

References

1. Adams H, Chaichana KL, Avendano J, Liu B, Raza SM, Quinones-Hinojosa A (2013) Adult cerebellar glioblastoma: understanding survival and prognostic factors using a population-based database from 1973 to 2009. *World Neurosurg* 80:e237–e243. <https://doi.org/10.1016/j.wneu.2013.02.010>
2. Babu R, Sharma R, Karikari IO, Owens TR, Friedman AH, Adamson C (2013) Outcome and prognostic factors in adult cerebellar glioblastoma. *J Clin Neurosci* 20:1117–1121. <https://doi.org/10.1016/j.jocn.2012.12.006>
3. Barnholtz-Sloan JS, Maldonado JL, Williams VL, Curry WT, Rodkey EA, Barker FG 2, Sloan AE (2007) Racial/ethnic differences in survival among elderly patients with a primary glioblastoma. *J Neurooncol* 85(2):171–180. <https://doi.org/10.1007/s11060-007-9405-4>
4. Ben Nsir A, Gdoura Y, Thai QA, Zhani Kassar A, Hattab N, Jemel H (2016) Intraventricular Glioblastomas. *World Neurosurg* 88:126–131. <https://doi.org/10.1016/j.wneu.2015.12.079>
5. Bohn A, Braley A, de la Rodriguez P, Zevallos JC, Barengo NC (2018) The association between race and survival in glioblastoma patients in the US: A retrospective cohort study. *PLoS ONE* 13(6):e0198581. <https://doi.org/10.1371/journal.pone.0198581>

6. Chandra A, Lopez-Rivera V, Dono A, Brandel MG, Lewis C, O'Connor KP, Sheth SA, Ballester LY, Aghi MK, Esquenazi Y (2021) Comparative Analysis of Survival Outcomes and Prognostic Factors of Supratentorial versus Cerebellar Glioblastoma in the Elderly: Does Location Really Matter? *World Neurosurg* 146:e755–e767. <https://doi.org/10.1016/j.wneu.2020.11.003>
7. Chang K, Zhang B, Guo X, Zong M, Rahman R, Sanchez D, Winder N, Reardon DA, Zhao B, Wen PY, Huang RY (2016) Multimodal imaging patterns predict survival in recurrent glioblastoma patients treated with bevacizumab. *Neurooncology* 18:1680–1687. <https://doi.org/10.1093/neuonc/now086>
8. Dayani F, Young JS, Bonte A, Chang EF, Theodosopoulos P, McDermott MW, Berger MS, Aghi MK (2018) Safety and outcomes of resection of butterfly glioblastoma. *Neurosurg Focus* 44(6):E4. <https://doi.org/10.3171/2018.3.FOCUS1857>
9. Doyle J, Khalafallah AM, Yang W, Sun Y, Bettgowda C, Mukherjee D (2019) Association between extent of resection on survival in adult brainstem high-grade glioma patients. *J Neurooncol* 145(3):479–486. <https://doi.org/10.1007/s11060-019-03313-w>
10. Filippini G, Falcone C, Boiardi A, Broggi G, Bruzzone MG, Caldiroli D, Farina R, Farinotti M, Fariselli L, Finocchiaro G, Giombini S, Pollo B, Savoirdo M, Solero CL, Valsecchi MG (2008) ; Brain Cancer Register of the Fondazione IRCCS (Istituto Ricovero e Cura a Carattere Scientifico) Istituto Neurologico Carlo Besta Prognostic factors for survival in 676 consecutive patients with newly diagnosed primary glioblastoma. *Neuro Oncol* 10:79–87. <https://doi.org/10.1215/15228517-2007-038>
11. Forjaz G, Barnholtz-Sloan JS, Kruchko C, Siegel R, Negoita S, Ostrom QT, Dickie L, Ruhl J, Van Dyke A, Patil N, Cioffi G, Miller KD, Waite K, Mariotto AB (2020) An updated histology recode for the analysis of primary malignant and nonmalignant brain and other central nervous system tumors in the Surveillance, Epidemiology, and End Results Program. *Neurooncol Adv* 3(1):vdad175. <https://doi.org/10.1093/noajnl/vdad175>
12. Fyllingen EH, Bø LE, Reinertsen I, Jakola AS, Sagberg LM, Berntsen EM, Salvesen Ø, Solheim O (2021) Survival of glioblastoma in relation to tumor location: a statistical tumor atlas of a population-based cohort. *Acta Neurochir (Wien)* 163(7):1895–1905. <https://doi.org/10.1007/s00701-021-04802-6>
13. Goldman DA, Reiner AS, Diamond EL, DeAngelis LM, Tabar V, Panageas KS (2020) Lack of survival advantage among re-resected elderly glioblastoma patients: a SEER-Medicare study. *Neurooncol Adv* 3(1):vdad159. <https://doi.org/10.1093/noajnl/vdad159>
14. González Bonet LG, Piqueras-Sánchez C, Roselló-Sastre E, Broseta-Torres R, de Las Peñas R (2021) Long-term survival of glioblastoma A systematic analysis of literature about a case. *Neurocirugia (Astur: Engl Ed)* 18. <https://doi.org/10.1016/j.neucie.2021.11.001>. S2529-8496(21)00049-6
15. Hashem IAT, Yaqoob I, Anuar NB, Mokhtar S, Gani A, Ullah Khan S (2015) The rise of 'Big Data' on cloud computing: review and open research issues. *Inf Syst* 47:98–115. <https://doi.org/10.1016/j.is.2014.07.006>
16. Hess KR (1999) Extent of resection as a prognostic variable in the treatment of gliomas. *J Neurooncol* 42:227–231. <https://doi.org/10.1023/a:1006118018770>

17. Hinton GE, Osindero S, Teh YW (2006) A fast-learning algorithm for deep belief nets. *Neural Comput* 18(7):1527–1554. <https://doi.org/10.1162/neco.2006.18.7.1527>
18. Jeswani S, Nuno M, Folkerts V, Mukherjee D, Black KL, Patil CG (2013) Comparison of survival between cerebellar and supratentorial glioblastoma patients: surveillance, epidemiology, and end results (SEER) analysis. *Neurosurgery* 73:240–246. <https://doi.org/10.1227/01.neu.0000430288.85680.37>
19. Johnson DR, O'Neill BP (2012) Glioblastoma survival in the United States before and during the temozolomide era. *J Neurooncol* 107:359–364. <https://doi.org/10.1007/s11060-011-0749-4>
20. Lacroix M, Abi-Said D, Fourney DR, Gokaslan ZL, Shi W, DeMonte F, Lang FF, McCutcheon IE, Hassenbusch SJ, Holland E, Hess K, Michael C, Miller D, Sawaya R (2001) A multivariate analysis of 416 patients with glioblastoma multiforme: prognosis, extent of resection, and survival. *J Neurosurg* 95(2):190–198. <https://doi.org/10.3171/jns.2001.95.2.0190>
21. Laperriere N, Weller M, Stupp R, Perry JR, Brandes AA, Wick W (2013) Optimal management of elderly patients with glioblastoma. *Cancer Treat Rev* 39(4):350–357. <https://doi.org/10.1016/j.ctrv.2012.05.008>
22. Li H, He Y, Huang L, Luo H, Zhu X (2020) The Nomogram Model Predicting Overall Survival and Guiding Clinical Decision in Patients with Glioblastoma Based on the SEER Database. *Front Oncol* 10:1051. <https://doi.org/10.3389/fonc.2020.01051>
23. Lin J, Bytnar JA, Theeler BJ, McGlynn KA, Shriver CD, Zhu K (2020) Survival among patients with glioma in the US Military Health System: A comparison with patients in the Surveillance, Epidemiology, and End Results program. *Cancer* 126(13):3053–3060. <https://doi.org/10.1002/cncr.32884>
24. Liu H, Qin X, Zhao L, Zhao G, Wang Y (2021) Epidemiology and Survival of Patients with Brainstem Gliomas: A Population-Based Study Using the SEER Database. *Front Oncol* 11:692097. <https://doi.org/10.3389/fonc.2021.692097>
25. Liu S, Wang Y, Fan X, Ma J, Qiu X, Jiang T (2017) Association of MRI-classified subventricular regions with survival outcomes in patients with anaplastic glioma. *Clin Radiol* 72(5):426. <https://doi.org/10.1016/j.crad.2016.11.013>. e1-426.e6
26. Liu ZY, Feng SS, Zhang YH, Zhang LY, Xu SC, Li J, Cao H, Huang J, Fan F, Cheng L, Jiang JY, Cheng Q, Liu ZX (2021) Competing risk model to determine the prognostic factors and treatment strategies for elderly patients with glioblastoma. *Sci Rep* 11(1):9321. <https://doi.org/10.1038/s41598-021-88820-5>
27. Levine SA, McKeever PE, Greenberg HS (1987) Primary cerebellar glioblastoma multiforme. *J Neurooncol* 5:231–236. <https://doi.org/10.1007/BF00151226>
28. Lopez-Rivera V, Dono A, Lewis CT, Chandra A, Abdelkhaleq R, Sheth SA, Ballester LY, Esquenazi Y (2021) Extent of resection and survival outcomes of geriatric patients with glioblastoma: Is there benefit from aggressive surgery? *Clin Neurol Neurosurg* 202:106474. <https://doi.org/10.1016/j.clineuro.2021.106474>

29. Noone AM, Lund JL, Mariotto A, Cronin K, McNeel T, Deapen D, Warren JL (2016) Comparison of SEER treatment data with Medicare claims. *Med Care* 54(9):e55–e64.
<https://doi.org/10.1097/MLR.000000000000073>
30. Ostrom QT, Cote DJ, Ascha M, Kruchko C, Barnholtz-Sloan JS (2018) Adult Glioma Incidence and Survival by Race or Ethnicity in the United States From 2000 to 2014. *JAMA Oncol* 4(9):1254–1262.
<https://doi.org/10.1001/jamaoncol.2018.1789>
31. Ostrom QT, Gittleman H, Farah P, Ondracek A, Chen Y, Wolinsky Y, Stroup NE, Kruchko C, Barnholtz-Sloan JS (2013) CBTRUS statistical report: primary brain and central nervous system tumors diagnosed in the United States in 2006-2010. *Neuro Oncol* 15(suppl 2):ii1–ii56.
<https://doi.org/10.1093/neuonc/not151>
32. Ostrom QT, Gittleman H, Liao P, Vecchione-Koval T, Wolinsky Y, Kruchko C, Barnholtz-Sloan JS (2017) CBTRUS Statistical Report: Primary brain and other central nervous system tumors diagnosed in the United States in 2010-2014. *Neuro Oncol* 19(suppl5):v1–v88.
<https://doi.org/10.1093/neuonc/nox158>
33. Ostrom QT, Cioffi G, Gittleman H, Patil N, Waite K, Kruchko C, Barnholtz-Sloan JS (2019) CBTRUS statistical report: primary brain and other central nervous system tumors diagnosed in the United States in 2012–2016. *Neuro-oncology* 21(Supplement_5. v1–v100.
<https://doi.org/10.1093/neuonc/noz150>
34. Ostrom QT, Rubin JB, Lathia JD, Berens ME, Barnholtz-Sloan JS (2018) Females have the survival advantage in glioblastoma. *Neuro Oncol* 20(4):576–577. <https://doi.org/10.1093/neuonc/noy002>
35. Palpan Flores A, Vivancos Sanchez C, Roda JM, Cerdán S, Barrios AJ, Utrilla C, Royo A, Gandía González ML (2020) Assessment of Pre-operative Measurements of Tumor Size by MRI Methods as Survival Predictors in Wild Type IDH Glioblastoma. *Front Oncol* 10:1662.
<https://doi.org/10.3389/fonc.2020.01662>
36. Patel NP, Lyon KA, Huang JH (2019) The effect of race on the prognosis of the glioblastoma patient: a brief review. *Neurol Res* 41(11):967–971. <https://doi.org/10.1080/01616412.2019.1638018>
37. Peeken JC, Goldberg T, Pyka T, Bernhofer M, Wiestler B, Kessel KA, Tafti PD, Nüsslin F, Braun AE, Zimmer C, Rost B, Combs SE (2019) Combining multimodal imaging and treatment features improves machine learning-based prognostic assessment in patients with glioblastoma multiforme. *Cancer Med* 8:128–136. <https://doi.org/10.1002/cam4.1908>
38. Roa W, Kepka L, Kumar N, Sinaika V, Matiello J, Lomidze D, Hentati D, Guedes de Castro D, Dytus-Cebulok K, Drodge S, Ghosh S, Jeremić B, Rosenblatt E, Fidarova E (2015) International Atomic Energy Agency Randomized Phase III Study of Radiation Therapy in Elderly and Frail Patients with Newly Diagnosed Glioblastoma Multiforme. *J Clin Oncol* 33(35):4145–4150.
<https://doi.org/10.1200/JCO.2015.62.6606>
39. Sanghani P, Ang BT, King NKK, Ren H (2018) Overall survival prediction in glioblastoma multiforme patients from volumetric, shape and texture features using machine learning. *Surg Oncol* 27:709–714. <https://doi.org/10.1016/j.suronc.2018.09.002>

40. Senders JT, Staples P, Mehrtash A, Cote DJ, Taphoorn MJB, Reardon DA, Gormley WB, Smith TR, Broekman ML, Arnaout O (2020) An Online Calculator for the Prediction of Survival in Glioblastoma Patients Using Classical Statistics and Machine Learning. *Neurosurgery* 86(2):E184–E192. <https://doi.org/10.1093/neuros/nyz403>
41. Shu C, Yan X, Zhang X, Wang Q, Cao S, Wang J (2019) Tumor-induced mortality in adult primary supratentorial glioblastoma multiforme with different age subgroups. *Future Oncol* 15(10):1105–1114. <https://doi.org/10.2217/fon-2018-0719>
42. Soon WC, Goacher E, Solanki S, Hayes J, Kapetanstrataki M, Picton S, Chumas PD, Mathew RK (2021) The role of sex genotype in paediatric CNS tumour incidence and survival. *Childs Nerv Syst* 37(7):2177–2186. <https://doi.org/10.1007/s00381-021-05165-0>
43. Stupp R, Hegi ME, Mason WP, van den Bent MJ, Taphoorn MJ, Janzer RC, Ludwin SK, Allgeier A, Fisher B, Belanger K, Hau P, Brandes AA, Gijtenbeek J, Marosi C, Vecht CJ, Mokhtari K, Wesseling P, Villa S, Eisenhauer E, Gorlia T, Weller M, Lacombe D, Cairncross JG, Mirimanoff RO, European Organisation for Research and Treatment of Cancer Brain Tumour and Radiation Oncology Groups; National Cancer Institute of Canada Clinical Trials Group (2009) Effects of radiotherapy with concomitant and adjuvant temozolomide versus radiotherapy alone on survival in glioblastoma in a randomized phase III study: 5-year analysis of the EORTC-NCIC trial. *Lancet Oncol* 10(5):459–466. [https://doi.org/10.1016/S1470-2045\(09\)70025-7](https://doi.org/10.1016/S1470-2045(09)70025-7)
44. Stupp R, Mason WP, van den Bent MJ, Weller M, Fisher B, Taphoorn MJ, Belanger K, Brandes AA, Marosi C, Bogdahn U, Curschmann J, Janzer RC, Ludwin SK, Gorlia T, Allgeier A, Lacombe D, Cairncross JG, Eisenhauer E, Mirimanoff RO, Clinical Trials Group (2005) ; European Organisation for Research and Treatment of Cancer Brain Tumor and Radiotherapy Groups; National Cancer Institute of Canada Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *N Engl J Med* 352(10):987-996. <https://doi.org/10.1056/NEJMoa043330>
45. Surveillance Research Program, National Cancer Institute SEER*Stat software (seer.cancer.gov/seerstat) version 8.3.6.1
46. Tewarie IA, Senders JT, Kremer S, Devi S, Gormley WB, Arnaout O, Smith TR, Broekman MLD (2021) Survival prediction of glioblastoma patients-are we there yet? A systematic review of prognostic modeling for glioblastoma and its clinical potential. *Neurosurg Rev* 44(4):2047–2057. <https://doi.org/10.1007/s10143-020-01430-z>
47. Thakkar JP, Dolecek TA, Horbinski C, Ostrom QT, Lightner DD, Barnholtz-Sloan JS, Villano JL (2014) Epidemiologic and molecular prognostic review of glioblastoma. *Cancer Epidemiol Biomarkers Prev*. <https://doi.org/10.1158/1055-9965.EPI-14-0275>. 23:1985-1996
48. Tian M, Ma W, Chen Y, Yu Y, Zhu D, Shi J, Zhang Y (2018) Impact of gender on the survival of patients with glioblastoma. *Biosci Rep* 38(6):BSR20180752. <https://doi.org/10.1042/BSR20180752>
49. Upadhaya T, Morvan Y, Stindel E, Le Reste PJ, Hatt M (2015) A framework for multimodal imaging-based prognostic model building: preliminary study on multimodal MRI in glioblastoma multiforme. *IRBM* 36:345–350. <https://doi.org/10.1016/j.irbm.2015.08.001>

50. White SE (2014) A review of Big Data in healthcare: challenges and opportunities. Open Access Bioinf 6:13–18. <https://doi.org/10.2147/OAB.S50519>

51. Yang W, Xu T, Garzon-Muvdi T, Jiang C, Huang J, Chaichana KL (2018) Survival of Ventricular and Periventricular High-Grade Gliomas: A Surveillance, Epidemiology, and End Results Program-Based Study. World Neurosurg 111:e323–e334. <https://doi.org/10.1016/j.wneu.2017.12.052>

52. Youngblood MW, Stupp R, Sonabend AM (2021) Role of Resection in Glioblastoma Management. Neurosurg Clin N Am 32(1):9–22. <https://doi.org/10.1016/j.nec.2020.08.002>

53. Zacharaki EI, Morita N, Bhatt P, O'Rourke DM, Melhem ER, Davatzikos C (2012) Survival analysis of patients with high-grade gliomas based on data mining of imaging variables. AJNR Am J Neuroradiol 33:1065–1071. <https://doi.org/10.3174/ajnr.A2939>

54. Zhou X, Niu X, Mao Q, Liu Y (2020) Clinical Significance of Various Classification Standards of Age Groups in Predicting Survival of Patients with Glioblastoma. Med Sci Monit 26:e920627. <https://doi.org/10.12659/MSM.920627>

Figures

Figure 1

Variable importance according to vital status variable