1-2021

Machine Learning Models for 6-Month Survival Prediction after Surgical Resection of Glioblastoma

Jeffrey Gray
Thomas Jefferson University, jeffrey.gray@students.jefferson.edu

Lohit Velagapudi
Thomas Jefferson University, lohit.velagapudi@students.jefferson.edu

Michael Baldassari
Thomas Jefferson University, michael.baldassari@jefferson.edu

Bryan Sadler

David Vuong

SKMC Class of 2023: SI/DH Abstract

Word count: 298

Machine Learning Models for 6-Month Survival Prediction after Surgical Resection of Glioblastoma

Jeffrey Gray, Lohit Velagapudi, Michael Baldassari, Bryan Sadler, David Vuong*

Introduction:

The role of surgical resection for the treatment of glioblastoma multiforme is well established. Survival analysis after resective surgery in the literature comprises mostly of traditional statistical models. Machine learning models offer powerful predictive and analytical capability for varied datasets and offer improved generalizability and scalability. We analyzed survival data of patients with glioblastoma with various machine learning algorithms and compared it to binary logistic regression.

Methods:

We retrospectively identified cases of glioblastoma treated with surgical resection at our institution from 2012-2018. Feature scaling and one-hot encoding was used to better fit the models to the data and used the formula $X' = (X - X_{min})/(X_{max} - X_{min})$. Feature selection was performed using chi-squared analysis (features with $p<0.05$ were kept). Significant variables used in predictive modeling were demographic characteristics, relevant past medical history, imaging, surgical course, post-operative treatment, radiation or chemotherapy, and histology. Outcome measured was 6-month survival. Each model was

run 30 times and the average accuracy was recorded. Split of data was 70% training and 30% testing.

Results:

582 patients fit the inclusion criteria and were used to build these models. 6-month mortality was 43.13%. Accuracy scores (AUC) for models used were 0.670 (logistic regression), 0.704 (Random Forest), 0.585 (Support Vector Machine), 0.560 (Naïve Bayes), 0.650 (XG Boost), 0.585 (Stochastic Gradient Descent Classifier), and 0.740 (Neural Network). 5-fold cross validation was used to ensure generalizability to an independent dataset.

Conclusion:

Machine learning methods for prediction of six-month survival for glioblastoma are promising analytical tools that we show can approach or exceed the accuracy of traditional logistic regression, particularly neural networks and the random forest algorithm. Improved prediction of 6-month survival using machine learning offers increased capabilities for patient education, adjuvant chemotherapy or radiation planning, and post-operative counseling, while maintaining increased adaptability and generalizability compared to regression models.