Supplementary Methods and Results

Appendix I.

Pre-Screening Demographic Variables

Methods

Data Analysis

Logistic regression was applied to candidate demographic variables that might predict the three outcomes of interest (i.e., response, partial remission, and full remission). The top two predictors based on the strongest odds ratio value would be retained in the final machine learning analyses. Odds ratio values (OR) close to one indicate no association, whereas values greater than and less than one indicate the predictor is associated with higher and lower odds of the outcome, respectively.

Available demographic variables included gender, race, age, and relationship status. Two of the variables were multi-categorical: race and relationship status. Because there was a small proportion of participants who did not identify as white (16%), the race variable was binary coded to represent white and non-white, as further divisions would lead to categories with very low percentages that could not be analyzed. Because relationship status contained several categories with very low percentages and could not be as easily collapsed into binary groups given the nature of the subcategories, this variable would not be a good candidate predictor for statistical reasons and, therefore, was not considered in the prescreening process.

Results

Results of the logistic regression for response suggested race (OR = 2.56, *p* > 0.05) and gender (OR = 0.50, *p* > 0.05) were the strongest predictors, such that being white and female was associated with higher odds of being a responder. Age was the least predictive variable (OR = 0.99, *p* > 0.05), nearing an OR of 1.

For partial remission, results again suggested race (OR = 2.17, *p* > 0.05) and gender (OR = 0.65, *p* > 0.05) were the strongest predictors, such that being white and female was associated with higher odds of being a partial remitter. Again, age was the least predictive variable (OR = 1.02, *p* > 0.05), nearing an OR of 1.

Finally for full remission, results of the logistic regression were similar to the models for the other two outcomes. Again, race (OR = 1.34, *p* > 0.05) and gender (OR = 0.68, *p* > 0.05) were the strongest predictors, such that being white and female was associated with higher odds of being a full remitter. Age was the least predictive variable (OR = 1.04, *p* > 0.05), nearing an OR of 1.

The collective results from the three logistic regression models for each outcome suggested that none of the predictors were significant, but age was the least successful predictor always nearing an OR of 1. Thus, race and gender were retained as variables for the machine learning models, given that they more related to the outcomes of interest.

Appendix II.

Methods

Data Analysis

To predict the three treatment outcomes of interest, originally several different machine learning algorithms were evaluated and compared to determine the best performing algorithm for the subsequent recurrence feature analyses described in the body of the main paper. In total, four machine learning procedures representing different families of supervised learning algorithms were tested: 1) radial kernel support vector machine (SVM), 2) binary logistic regression, 3) classification and regression tree (CART), and 4) gradient boost machines (GBM).

Although a full description of each of these techniques is beyond the purview of the current paper, each algorithm accomplishes classification invoking different analytic approaches. Radial kernel SVM is a non-linear classification technique that uses hyperplanes to separate the data into different classes. Binary logistic regression is a traditional generalized linear model procedure using the logit link function, permitting prediction of binary outcome variables. CART is a nonparametric method that creates a decision tree based on predictors and their combinations to optimize classification of the outcome variable. GBM is an ensemble technique that optimizes prediction by training weak trees successively with each tree learning and improving on the previous one.

All machine learning models were examined using ten-fold cross-validation, which partitions the sample into ten subsets, of which nine are used in the training process and predictions are made in the remaining subset. This process is repeated for each of the remaining subsets, and results are averaged to produce a single estimate. To appraise classification performance, receiver operator characteristics (ROC) and area under the curve (AUC) metrics were calculated. Also, standard ROC metrics were evaluated such as sensitivity (i.e., proportion of positives correctly identified) and specificity (i.e., proportion of negatives that are correctly identified). Each of these three metrics range between 0 and 1 such that larger values indicate better performance.

For this initial screening step of determining the best machine learning algorithm to employ for the subsequent analyses, all four algorithms were used to predict each of the three outcomes of interest (i.e., response, partial remission, and full remission).

Results

As indicated in Tables S1-S3, the SVM algorithm offered the best performance with respect to AUC for classifying response (AUC = 0.67), partial remission (AUC = 0.79), and full remission (AUC = 0.69). Thus, for SVM will be adopted for the primary analyses conducted as part of the main aims of the study.

Table S1. Classification Performance for Response

|  |  |  |  |
| --- | --- | --- | --- |
| Algorithm | Area Under Curve | Sensitivity | Specificity |
| SVM | 0.67 | 0.77 | 0.37 |
| Logistic Regression | 0.65 | 0.70 | 0.55 |
| CART | 0.52 | 0.73 | 0.40 |
| GBM | 0.65 | 0.74 | 0.47 |

*Notes*: SVM = support vector machine; CART = classification and regression tree; GBM = gradient boost machines.

Table S2. Classification Performance for Partial Remission

|  |  |  |  |
| --- | --- | --- | --- |
| Algorithm | Area Under Curve | Sensitivity | Specificity |
| SVM | 0.79 | 0.69 | 0.69 |
| Logistic Regression | 0.74 | 0.63 | 0.63 |
| CART | 0.61 | 0.73 | 0.56 |
| GBM | 0.78 | 0.72 | 0.74 |

*Notes*: SVM = support vector machine; CART = classification and regression tree; GBM = gradient boost machines.

Table S3. Classification Performance for Full Remission

|  |  |  |  |
| --- | --- | --- | --- |
| Algorithm | Area Under Curve | Sensitivity | Specificity |
| SVM | 0.69 | 0.55 | 0.80 |
| Logistic Regression | 0.66 | 0.55 | 0.70 |
| CART | 0.62 | 0.45 | 0.75 |
| GBM | 0.67 | 0.45 | 0.79 |

*Notes*: SVM = support vector machine; CART = classification and regression tree; GBM = gradient boost machines.