Development and External Validation of Machine Learning Algorithms for Survival Prediction in Undifferentiated Pleomorphic Sarcoma

Tom Maarten De Groot, Marcos R Gonzalez, Angad Bedi, Joseph Hasbrouck Schwab[1], Job N Doornberg[2], Santiago Andres Lozano Calderon[3]
[1]MGH Dept. of Ortho Surg., [2]University Medical Centre Groningen, [3]Massachusetts General Hospital - Harvard Medical S

INTRODUCTION:

Undifferentiated pleomorphic sarcoma (UPS) is a common and aggressive subtype of soft tissue sarcoma (STS) associated with high potential for recurrence and metastasis. Currently, the American Joint Commission on Cancer (AJCC) serves as the gold standard for prognostic assessment and is based on tumor size, regional lymph node involvement, and metastatic disease. However, AJCC staging guidelines rely on arbitrary threshold values and do not take into account the histology of the STS. Machine learning (ML) algorithms offer a potential solution to predict overall survival of UPS patients, providing individualized clinical insights on prognostic factors. This study aimed to develop and validate a ML algorithm to predict two- and five-year overall survival of UPS patients.

METHODS:

We included patients from the Surveillance, Epidemiology, and End Results (SEER) database for our training and internal validation set. To be included, patients were required to **1**) have a histologic diagnosis of UPS, **2**) be located in the trunk or extremities, and **3**) have complete follow-up data. We externally validated the models with patients retrieved from our institutional STS database. Five algorithms were evaluated: XGBoost, Random Forest, support vector machine, neural network, and penalized logistic regression. Using the training dataset, an exhaustive grid search was conducted to identify the optimal hyperparameters which provided the highest accuracy in a 10-fold internal cross-validation of each model. Model performance was assessed using metrics such as the area under the curve (AUC), calibration (Figure 2), Brier score, null model Brier score, and decision curve analysis.

RESULTS:

A total of 3,596 from the SEER database were included. For external validation, 127 patients treated at our institution were included. Tumor size, age, metastases at diagnosis, regional lymph node involvement at diagnosis, and sex were factors associated with overall survival (Figure 3). On internal validation, the best performing algorithm for two-year survival was the neural network, with an AUC of 0.79, a slope of 1.01 and an intercept of -0.01 (Table 1). For 5-year survival, the support vector machine performed best, with an AUC of 0.79, a slope of 1.06, and an intercept of -0.05 (Table 1). Brier scores did not differ between predictive modalities for all timepoints, except for 2-year survival, where the Brier-score of the neural network and XGBoost algorithm were 0.13, compared to 0.17 for the other algorithms. All Brier scores were lower than their respective null-model Brier scores for every survival timepoint.

On external validation, the neural network model for 2-year survival showed excellent discriminative performance with an AUC of 0.93, a slope of 1.06, and an intercept of -0.07 (Table 2). For 5-year survival, the support vector machine algorithm showed good performance in terms of discrimination with an AUC of 0.83 (Figure 1), a slope of 0.96, and an intercept of -0.11. Brier scores were lower than their respective null-model Brier scores for both the 2-year timepoint (0.16 versus 0.29) and the 5-year survival timepoint (0.14 – 0.24).

DISCUSSION AND CONCLUSION:

Our study successfully developed and validated a machine learning algorithm that accurately predicted 2-year and 5-year survival in patients with UPS. The accuracy of our model was higher than that of the *Sarculator*, an online prediction model widely used in the orthopaedic oncology community. To confirm worldwide generalizability, further external validation of this algorithm is encouraged.