Toward Precision Oncology: Integrating Clinical and Mutational Data to Predict Pancreatic Cancer Survival Using Machine Learning

Basas Jaumandreu, Bruna (School: La Salle Manresa)

Pancreatic adenocarcinoma (PAAD) is one of the deadliest cancers, with a five-year survival rate of only 13.3%. This poor prognosis is largely attributed to late-stage diagnosis and limited treatment effectiveness, making it especially important to better understand and predict patient outcomes. Moreover, current predictive models often lack interpretability or do not fully integrate molecular data. Therefore, I developed an interpretable logistic regression model to predict the probability of survival beyond 24 months in PAAD patients using both clinical and mutational data. Following data characterisation and preprocessing, the model was trained with L2 regularization and evaluated using nested cross-validation to optimize hyperparameters and evaluate performance. The model achieved an AUC of ~0.76 and identified clinically relevant patterns, including the association of TP53 mutations with poorer outcomes and the strong protective effect of resection surgery. Also, KRAS and TP53 were highlighted as key molecular predictors, suggesting their importance in understanding survival differences. By performing a comparative analysis across different feature sets, I found that models incorporating mutational variables achieved a higher AUC, demonstrating that molecular features provide complementary prognostic information beyond clinical factors. Overall, this study shows that interpretable machine learning models can accurately predict PAAD survival and support patient risk stratification. This approach could improve clinical decision-making, contribute to better patient outcomes, and lead to more efficient use of healthcare resources.

Awards Won: