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

Joseph Oliver Werenski[1], Sanjeev Rampam, Andrew G Girgis, Bishoy Moussa Galoaa, Marcos R Gonzalez, Santiago Andres Lozano Calderon[2]
[1]Massachusetts General Hospital, [2]Massachusetts General Hospital - Harvard Medical S

INTRODUCTION: Synovial sarcoma is an uncommon soft tissue sarcoma subtype characterized by a relatively young age at diagnosis, peri-articular location, and slow-growing and painful mass. While standard treatment involves surgical excision with negative margins, supplemented by radiation therapy for larger, deeper tumors, almost half of patients present with clinical findings inconsistent with soft tissue sarcoma and undergo unplanned excision without proper imaging studies or a diagnostic biopsy. Prognostication in synovial sarcoma currently relies on several factors, including age at diagnosis, tumor size and grade, margin status, neurovascular invasion, unplanned excision, and use of radiation therapy. Existing synovial sarcoma survival prediction tools rely on threshold rather than individualized clinical values. To overcome this limitation, a machine learning (ML) algorithm has been developed and validated to predict two- and five-year overall survival rates in synovial sarcoma patients, offering a more personalized prognostic tool.

METHODS: A total of 2,747 patients with histologically confirmed extremity synovial sarcoma from the National Cancer Database were included for model development. An additional 152 patients from an institutional database were used for external validation. For model development, tabular patient data was transformed into high-dimensional embeddings using a pre-trained language model, which was then used to classify patient outcomes through a neural network optimized for high-dimensional input, specifically integrating the BERT model with attention mechanisms. This process involved converting patient data into text descriptions, generating embeddings via the pre-trained LLM, and employing a neural network tailored for these embeddings. To address class imbalance, we utilized techniques such as the Synthetic Minority Over-sampling Technique (SMOTE), combining the semantic capabilities of LLMs with the analytical power of neural networks for medical data analysis. Model performance was assessed using the area under the ROC curve (AUC), F1 score, and Brier scores, with primary outcomes of interest being two- and five-year overall survival.

RESULTS: On internal validation, the model demonstrated superior performance compared to standard machine learning algorithms, which typically have an AUC of 0.7 to 0.8. The model achieved an AUC of 0.91 for two-year survival (Precision=0.77, Recall=0.76, and F1=0.80) and an AUC of 0.93 for five-year survival (Precision=0.81, Recall=0.79, and F1=0.75). External validation further confirmed the model's robustness, with excellent discriminative performance indicated by an AUC of 0.93 and 0.92 for two- and five-year survival, respectively.

DISCUSSION AND CONCLUSION: In this study, a ML algorithm was successfully developed and validated to accurately forecast two- and five-year survival in patients with synovial sarcoma. This algorithm enhances clinical decision-making by providing personalized prognostication, thereby improving treatment planning and patient counseling for multidisciplinary care teams. For broader clinical applicability and to confirm global generalizability, further external validation of this algorithm is recommended.

Figure 1. Workflow of artificial intelligence model development and data input structures for generating predictive survival estimates.

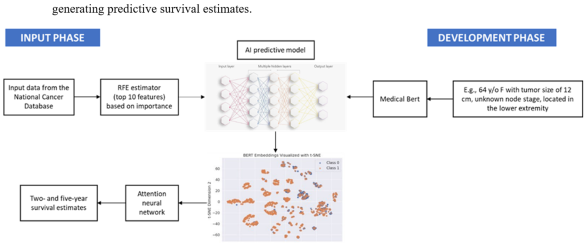

Table 1. Performance metrics of the BERT-based attention neural network models for predicting two- and five-year survival on internal validation (National Cancer Database, n=2,747) and external validation (institutional cohort, n=152).

		Survival prediction			
		Two-year		Five-year	
Validation type		Internal	External	Internal	External
Performance metric	AUC	0.91	0.93	0.93	0.92
	Precision	0.77	0.78	0.81	0.79
	Recall	0.76	0.79	0.79	0.77
	F1 score	0.80	0.81	0.75	0.74