Artificial intelligence predicts survival outcome of breast carcinomas based on whole-slide histopathology images

C. Bossard[1,2], Y. Salhi[1], J. Paul[1], J. Rynkiewicz[1], F. Molinié[3], S. Salhi[1], J. Chetritt[1,2]

[1] DiaDeep, France - [2] Pathology Department, IHP Group, France, - [3] Loire-Atlantique/Vendée Cancer Registry, France

Background

Breast cancer is the most common cancer among women, with an estimated 2 million new cases and 627,000 deaths globally in 2020. Early diagnosis and appropriate treatment are key to improve patient outcomes and survival rates. Several prognostic biomarkers – clinical, pathological, and molecular – are routinely used to assess the risk of disease progression and the response to targeted therapy, thereby guiding treatment decisions. However, some limitations remain, requiring new tools to better identify patients at high risk of relapse and death compared with those at low risk, to determine the most appropriate therapeutic strategie. In this context, artificial intelligence (AI) applied to digitized histopathological images could serve as a powerful tool to predict prognosis. AI has the potential to detect new morphological features not taken into account by the pathologist. In this study, we aim to develop a deep-learning algorithm to predict the five-year overall survival of patients with breast carcinomas, based solely on the analysis of H&E-stained WSI (Whole slide Image) of tumors.

Methods

We introduce a deep neural network (DNN) specifically designed to calculate a survival risk score for patients with breast carcinomas directly from WSI of H&E-stained sections of the tumor, without requiring any annotations. The model integrates two distinct types of morphological information: one encapsulates cellular-level details, while the other addresses tissue-level features, using the Cox proportional hazard loss function. The model pipeline is presented in Figure 1. The algorithm has been trained and evaluated on the publicly available TCGA-BRCA dataset. Cases lacking a date of diagnosis, a WSI, or those with marker-pen artifacts were not included, resulting in a total of 1,003 patients. Among these cases, there were 141 recorded deaths. We employed a 5-fold stratified framework and enhanced both the assessment of our pipeline and the model's robustness through cross-testing and cross-validation techniques. The concordance index (c-index) was used as a metric to assess the performance of the proposed algorithm. Model outcomes were then integrated into a Cox model including age, pTNM, OR, PR and HER2 to evaluate whether our model could serve as an independent prognostic factor. Additionally, we developed a heatmap on WSI showing Regions Of Interest (ROI) captured by our model to provide the AI-based risk score (Figure 2). An external cohort from the IHP Group of 222 patients was used to further evaluate the performance of our model. Characteristics of the 2 chorts are presented in Table 1.

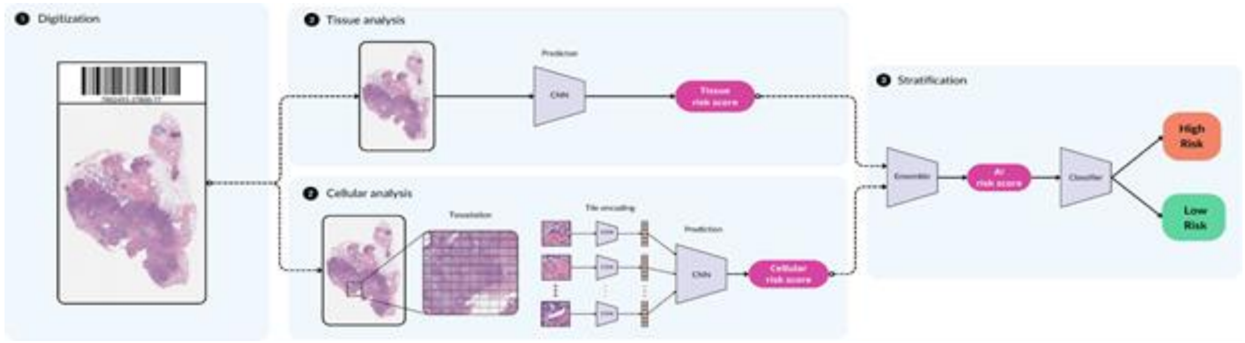

Figure 1 - Model pipeline

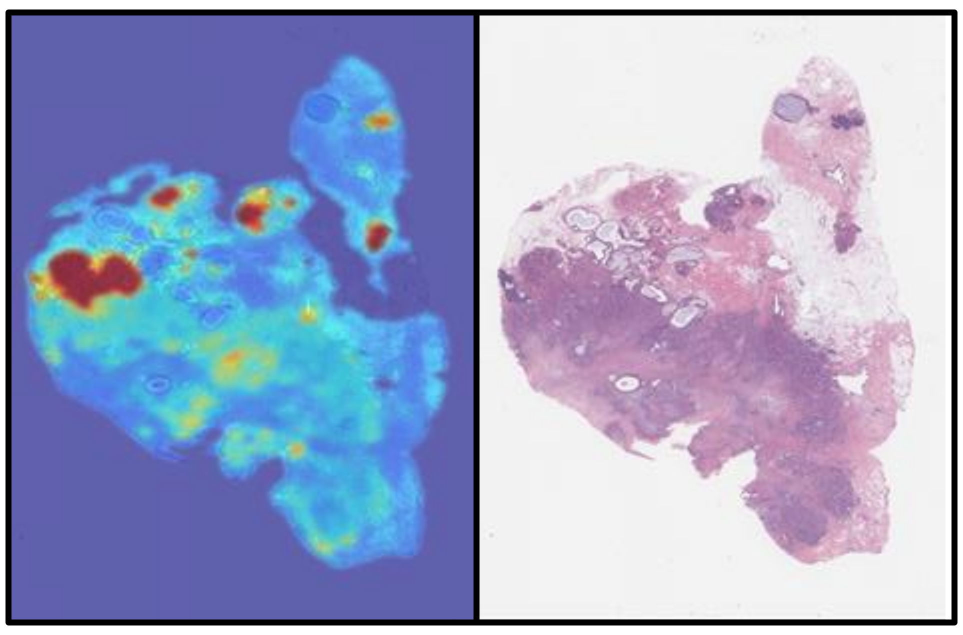

Figure 2 - Heatmap representing the most informative regions of the WSI (red)

Conclusion

Interestingly, our algorithm can automatically extract specific morphological features from HE WSI, to predict an AI-based risk score associated with a significant prognostic influence in terms of overall survival for breast cancer patients. This emerging and disruptive prognostic approach represents a new concept in the field of precision oncology and personalized medicine. Further studies based on other independent cohorts are required to validate the performance of this algorithm, and confirm its superiority over the current prognostic markers, as well as to discover new morphological biomarkers. Indeed, deciphering and understanding the prognostic ROI on tumor tissues will add somes important informations in terms of interpretability of our models, and thus confidence.

Dataset		TCGA 1003 patients	IHP-BR 222 patients
Deaths, n		141 (14.05%)	17 (7.6%)
Age at diagnosis, mean, [range]		58.09 [26-90]	58.66 [27-89]
Stage, n (%)			
	I	170 (16.94%)	83 (37.38%)
	II	570 (56.82%)	63 (28.37%)
	III	222 (22.13%)	19 (8.5%)
	IV	19 (1.89%)	0 (0%)
	X	11 (1.09%)	0 (0%)
	N/A	11 (1.09%)	57 (25.67%)
ER, n (%)			
	Positive	751 (74.87%)	176 (79.28%)
	Negative	223 (22.23%)	45 (20.27%)
	N/A	29 (2.89%)	1 (0.45%)
RP, n (%)			
	Positive	654 (65.20%)	159 (71.62%)
	Negative	317 (31.60%)	62 (27.92%)
	N/A	32 (3.19%)	1 (0.4%)
HER2, n (%)			
	Positive	151 (15.05%)	2 (0.9%)
	Equivocal	172 (17.14%)	0 (0%)
	Negative	525 (52.34%)	218 (98.20%)
	N/A	155 (15.45%)	2 (0.9%)

Table 1 – Clinicopathological characteristics of the 2 cohorts of patients

Results

The SmartProg model demonstrated an average c-index of 0.682 when assessed on the testing set. The Cox model, including only clinicopathological features (age, ER/PR, HER2, and pTNM stage), produced an average c-index of 0.78. This value increased to 0.79 with the inclusion of our AI-based risk score. These results indicate that the proposed model not only matches but also surpasses existing models in predicting survival, thanks to the robust cross-testing and cross-validation techniques. Additionally, the Cox model suggests that our AI-based risk score can be used as an independent prognostic factor for predicting overall survival ($p < 0.005$), see Table 2. Furthermore, we were able to significantly discriminate between two groups of patients in terms of survival outcomes, depending on their AI-based high or low risk. A Kaplan-Meier survival curve combining results from each of the five testing sets illustrates the stratification of the population into two distinct groups: the high-risk group with a median OS of 8.5 years, and the low-risk group (not reached median OS) (log-rank test p-value of 9.02e-10) (Figure 3). Finally, an example of a heatmap from a WSI is depicted in Figure 2, showing regions of interest (ROI) in the tumor tissue with important prognostic value for our model. This approach aims to provide an onset of interpretability. The evaluation of the model on the external dataset yielded performance indices of 0.74 and 0.71, with and without the AI-risk score, respectively.

	TCGA		
	HR	95% CI	p-value
Age	1.04	[1.02-1.05]	<0.005*
Stage	1.73	[1.47-2.04]	<0.005*
ER	0.91	[0.53-1.58]	0.74
PR	0.8	[0.48-1.33]	0.39
HER2	1.07	[0.74-1.54]	0.72
SmartProg-BR	1.69	[1.35-1.35]	<0.005*

Table 2 - Multivariate Cox analysis

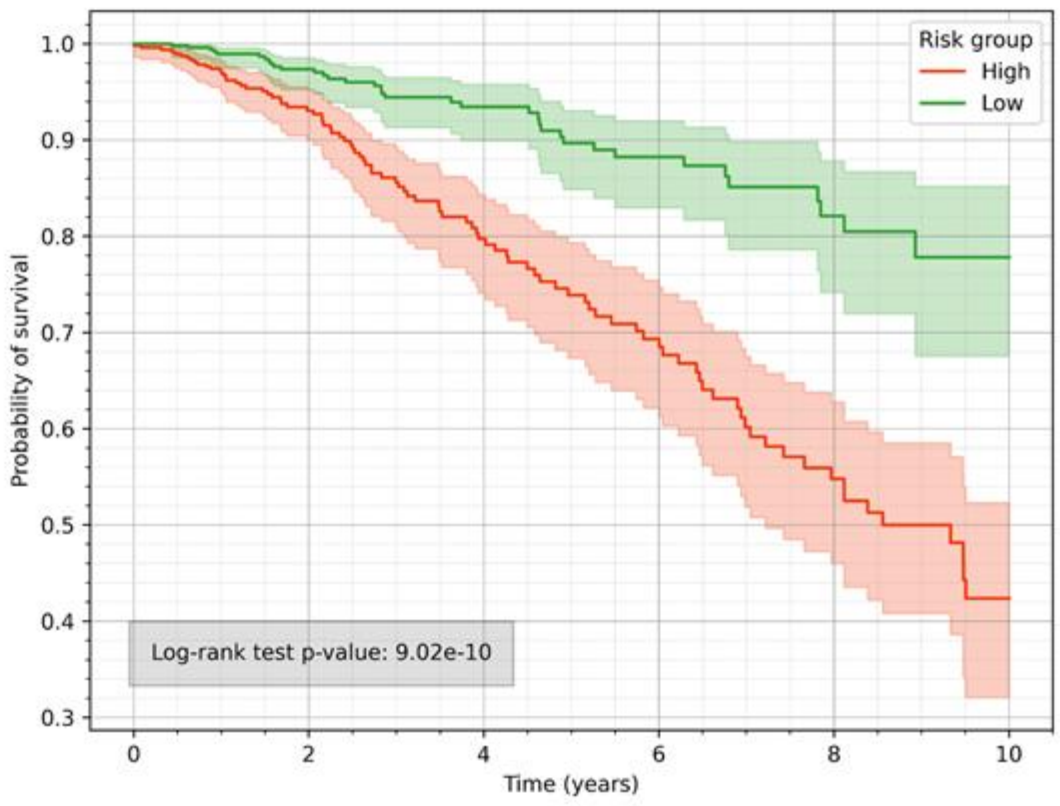

Figure 3 - Kaplan Meier curve of the testing set depending on the AI-based risk score