[BIOINFO 2025]

Predicting Nottingham grade in breast cancer digital pathology using a foundation model

Jun Seo Kim[1†], Jeong Hoon Lee[2†], Yousung Yeon[1], Doyeon An[1], Seok Jun Kim[1], Suehyun Lee[1*] and Myung-Giun Noh[3*]

[1]Department of Computer Engineering, Gachon University,
[2]Department of Radiology, Stanford University School of Medicine, USA,
[3]Department of Pathology, School of Medicine, Ajou University

*Corresponding author: *mgnoh@ajou.ac.kr*

The Nottingham histologic grade is crucial for assessing severity and predicting prognosis in breast cancer, a prevalent cancer worldwide. Traditional grading systems rely on subjective expert judgment and require extensive pathological expertise, are time-consuming, and often lead to inter-observer variability. To address these limitations, we develop an AI-based model to predict Nottingham grade from whole-slide images of hematoxylin and eosin (H&E)-stained breast cancer tissue using a pathology foundation model. From TCGA database, we trained and evaluated using 521 H&E breast cancer slide images with available Nottingham scores through internal split validation, and further validated its clinical utility using an additional set of 597 cases without Nottingham scores. The model leveraged deep features extracted from a pathology foundation model (UNI) and incorporated 14 distinct multiple instance learning (MIL) algorithms. The best-performing model achieved an F1 score of 0.731 and a multiclass average AUC of 0.835. The top 300 genes correlated with model predictions were significantly enriched in pathways related to cell division and chromosome segregation, supporting the model's biological relevance. The predicted grades demonstrated statistically significant association with 5-year overall survival ($p < 0.05$). Our AI-based automated Nottingham grading system provides an efficient and reproducible tool for breast cancer assessment, offering potential for standardization of histologic grade in clinical practice.