Prediction of Mutations and Outcome in Gastrointestinal Stromal Tumors with Deep Learning: A Multicenter, Multinational Study

Info

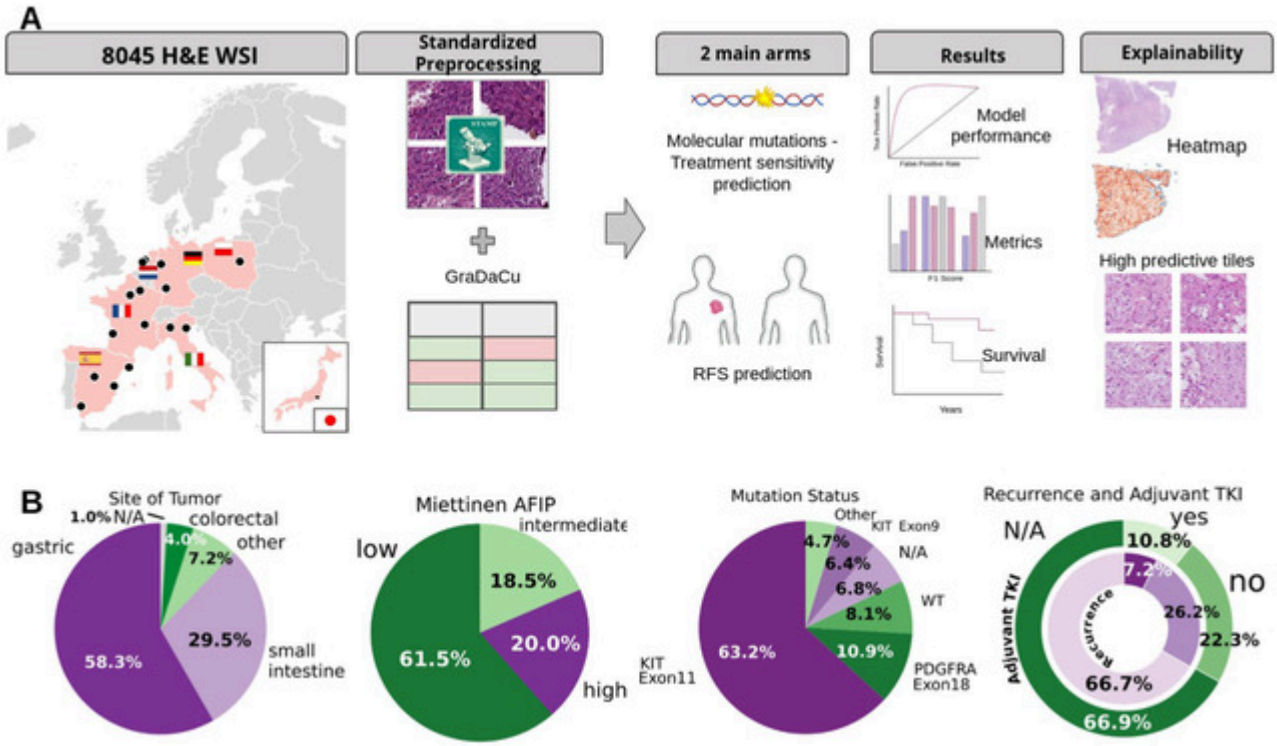
PUBLISHED DATE
2026-02-03

JOURNAL
preprint

LINK
https://doi.org/10.64898/2026.02.02.26345350

Authors

A. Bonetti V.L. Le Z. I. Carrero F. Wolf M. Gustav S.W. Lam L. Vanhersecke P. Sobczuk F. Le Loarer M. Lenarcik P. Rutkowski J. M. van Sabben N. Steeghs H. van Boven I. Machado S. Bagué S. Navarro E. Medina-Ceballos C. Agra F. Giner G. Tapia A. Hernández-Gallego G. Civantos Jubera M. Cuatrecasas S. Lopez-Prades R.E. Perret I. Soubeyran E. Khalifa L. Blouin E. Wardelmann A. Meurgey P. Collini A. Voloshin Y. Yatabe H. Hirano A. Gronchi T. Nishida O. Bouché J.F. Emile C. Ngo P. Hohenberger C. Cotarelo J. Jakob J.V.M.G. Bovee H. Gelderblom A. Szumera-Cieckiewicz M. Jean-Denis J. Bollard N. Lassau A. Lecesne J.Y. Blay A. Italiano A. Crombé J.M. Coindre **J. N. Kather**



Graphical Abstract

Overview of study design and dataset characteristics.

(A) Multinational collection of WSIs from seven countries (Spain, France, Italy, Germany, the Netherlands, Poland, and Japan), followed by standard image preprocessing with the STAMP pipeline and clinical data preprocessing/standardization via the Grammar Data Curation framework. The workflow was divided into two main branches: (i) molecular mutation and treatment sensitivity prediction, and (ii) RFS prediction. Model performance was evaluated using AUROC and F1 score for classification tasks, and Kaplan–Meier survival curves with hazard ratios for RFS. Model explainability was assessed through heatmaps of WSIs and identification of top predictive tiles. (B) Summary of clinical dataset composition: proportion of cases receiving adjuvant therapy, tumor location distribution, mutation distribution at the exon level, and mutation distribution at the codon level.

Abstract

Background: Gastrointestinal stromal tumor (GIST) is the most common gastrointestinal mesenchymal tumor, driven by tyrosine-protein kinase KIT and platelet-derived growth factor receptor A (PDGFRA) mutations. Specific variants, such as KIT exon 11 deletions, carry prognostic and therapeutic implications, whereas wild-type (WT) variants derive limited benefit from tyrosine kinase inhibitors (TKIs). Given the limited reproducibility of established clinicopathological risk models, deep learning (DL) applied to whole-slide images (WSIs) emerged as a promising tool for molecular classification and prognostic assessment.

Patients and methods: We analyzed 8398 GIST cases from 21 centers in 7 countries, including 7238 with molecular data and 2638 with clinical follow-up. DL models were trained on WSIs to predict mutations, treatment sensitivity, and recurrence-free survival (RFS).

Results: DL predicted mutational status in GIST from WSIs, with area under the curve (AUC) of 0.87 for KIT, 0.96 for PDGFRA. High performance was observed for subtypes, including KIT exon 11 del-inss 557–558 (0.67) and PDGFRA exon 18 D842V (0.93). For therapeutic categories, performance reached 0.84 for avapritinib sensitivity, 0.81 for imatinib sensitivity. DL models predicted RFS, with hazard-ratios (HR) of 8.44 (95%CI 6.14–11.61) in the overall cohort and 4.74 (95%CI 3.34–6.74) in patients receiving adjuvant therapy. Prognostic performance was comparable to pathology-based scores, with highest discrimination in the overall cohort and in patients without adjuvant therapy (9.44, 95%CI (5.87–15.20)).

Conclusion: DL applied to WSIs enables prediction of molecular alterations, treatment sensitivity, and RFS in GIST, performing comparably to established risk scores across international cohorts, providing a baseline for future multimodal predictors.

Keywords: GIST, molecular mutations, treatment, recurrence free survival, deep learning