PanMETAI: A High-Performance Tabular Foundation Model for Early Pancreatic Cancer Diagnosis and Its Clinical Translation via Biofluid Evaluation

Chun-Mei Hu
胡春美

Genomics Research Center, Academia Sinica, Taipei, Taiwan

Abstract:

Late diagnosis and the lack of effective early detection techniques contribute to the poor prognosis of pancreatic ductal adenocarcinoma (PDAC). To address this challenge, we developed PanMETAI, an AI-driven metabolomics platform employing the Tabular Foundation Model (TabPFN) framework. This platform integrates ^{1}H NMR (600MHz) derived serum metabolomic profiles—including small-molecule metabolites and lipoproteins—with clinical parameters (age, CA19-9) and Activin A.

Evaluated across diverse cohorts, PanMETAI demonstrated exceptional diagnostic efficacy. In the Taiwanese training and validation cohort (n=902), it achieved an AUC of 0.99 (95% CI: 0.98–0.99). Its robustness was confirmed in a Lithuanian external validation cohort (n=322), yielding an AUC of 0.93 (0.90-0.95). Notably, PanMETAI accurately identifies key metabolite features to improve early-stage (I/II) PDAC diagnosis, performing robustly even with small sample sizes.

To establish analytical versatility for routine clinical implementation, we conducted a comparative biofluid study using 291 matched serum and plasma pairs (comprising 98 high-risk controls, 97 Stage I/II PDAC, and 96 Stage III/IV PDAC). Validated through 100-iteration Monte Carlo cross-validation, PanMETAI exhibited robust predictive performance across both matrices. While both serum and plasma yielded exceptional accuracy (AUC 0.97–0.98), our analysis revealed that EDTA in plasma induces significant spectral interference, particularly by distorting the NMR signatures of lipoproteins and low-molecular-weight metabolites. This interference introduced a modest bias—elevating sensitivity at the expense of specificity—which ultimately compromised the analytical consistency relative to serum.

Conclusion: PanMETAI offers a rapid, highly generalizable, and non-invasive solution for early PDAC detection. This matched-pair evaluation confirms that while the platform remains versatile, the superior spectral integrity and specificity of serum establish it as the gold-standard matrix for clinical translation and large-scale metabolic screening.