npj Precision Oncology

https://doi.org/10.1038/s41698-026-01346-9

Article in Press

Multimodal data-based graph convolutional networks for predicting outcomes in ovarian cancer receiving neoadjuvant chemotherapy

Received: 21 July 2025

Accepted: 19 February 2026

Cite this article as: Zhang, S., Liu, Y., Liu, Z. *et al.* Multimodal data-based graph convolutional networks for predicting outcomes in ovarian cancer receiving neoadjuvant chemotherapy. *npj Precis. Onc.* (2026). https://doi.org/10.1038/s41698-026-01346-9

Shimin Zhang, Yinlong Liu, Zhuonan Liu, Xinyue Li, Guan Wang, Zhuo Yang, Yutong Liu, Meiyao Li, Jiarui Wang, Jiage Zhang, Bosinan Chen, Jingyi Liu, Yi Zhang, Jiangdian Song & Xin Zhou

Title: Multimodal Data-Based Graph Convolutional Networks for Predicting Outcomes in Ovarian Cancer Receiving Neoadjuvant Chemotherapy

Authors and Affiliations

Shimin Zhang[1], Yinlong Liu[2], Zhuonan Liu[3], Xinyue Li[1], Guan Wang[4], Zhuo Yang[5], Yutong Liu[1], Meiyao Li[1], Jiarui Wang[1], Jiage Zhang[1], Bosinan Chen[6], Jingyi Liu[6], Yi Zhang[7]*, Jiangdian Song[8]*, Xin Zhou[1]*.

1. Department of Obstetrics and Gynecology, Shengjing Hospital of China Medical University, Shenyang, Liaoning, 110004, China
2. Faculty of Data Science, City University of Macau, Room S504, Stanley Ho Building, Macau, China
3. Department of Radiology, Shengjing Hospital of China Medical University, Shenyang, Liaoning, 110004, China
4. Department of Radiology, The First Hospital of China Medical University, Shenyang, Liaoning, 110001, China
5. Department of Gynaecology, Cancer Hospital of China Medical University; Cancer Hospital of Dalian University of Technology; Liaoning Cancer Hospital & Institute, Shenyang, Liaoning, 110001, China
6. The First Hospital of China Medical University, Shenyang, Liaoning,110001, China
7. Department of Gynecology, The First Hospital of China Medical University, Shenyang, Liaoning, 110001, China
8. School of Health Management, China Medical University, Shenyang, Liaoning, 110122, China

*Corresponding author

Prof. Jiangdian Song

China Medical University

No.77 Puhe Road, Shenbei New District, Shenyang, Liaoning, 110122 China

song.jd0910@gmail.com

Abstract

Although neoadjuvant chemotherapy (NACT) is commonly used for advanced ovarian cancer, patient outcomes vary substantially. We developed a graph convolutional network (GCN) that integrates patient-specific baseline clinical variables and computed tomography–derived radiomic features while modeling inter-patient relationships to improve outcome prediction beyond standard models. The GCN operates without reliance on high-performance computing resources and predicts long-term overall survival (OS) while stratifying short-term surgical outcomes (R0 resection). The GCN was compared with CA-125 ELIMination rate constant K (KELIM) score and three Cox-based comparator models. Model performance was evaluated using the concordance index (C-index) for OS, area under the receiver operating characteristic curve for 3-year OS, Kaplan–Meier survival analysis, and R0 resection stratification. The GCN demonstrated strong OS prognosis performance (C-index = 0.73, 0.72, and 0.70 across the training and two external test datasets), stratified surgical outcomes, and identified 16.30% of patients with low KELIM scores but favorable survival.

Introduction

Ovarian cancer is the most lethal gynecologic malignancy, and more than 70% of patients are diagnosed at an advanced stage with regional or distant metastasis [1]. Despite advances in treatment, the prognosis for advanced ovarian cancer remains poor, with a 5-year survival rate of less than 40% [2], highlighting the urgent need for improved prognostic tools and personalized treatment strategies. For patients ineligible for primary cytoreductive surgery, interval cytoreductive surgery (ICS) following neoadjuvant chemotherapy (NACT) has become a widely adopted treatment approach [3]. Between 2010 and 2021, the proportion of patients undergoing ICS after NACT increased from 16.6% to 40.8% [4]. Although NACT followed by interval cytoreduction can enhance resectability and reduce perioperative complications compared to primary cytoreduction, determining which patients will derive optimal survival benefit from NACT and appropriate timing of treatment remain critical clinical challenges[5-7]. Accurately identifying which patients are likely to benefit from NACT is critical for optimizing treatment strategies.

Predictive models have been developed to inform treatment decisions in ovarian cancer; however, those focused on prognostication for patients undergoing NACT have largely relied on a single data modality, resulting in limited improvements in precision [8-11]. Indicators such as changes in CA-125 levels during NACT, the neutrophil-to-lymphocyte ratio, the platelet-to-lymphocyte ratio, and computed tomography (CT) semantic features have been correlated with clinical outcomes in these patients [12-16]. Integrated predictive models, which combine imaging data that provide intuitive visual assessments of disease status with non-imaging data that capture complementary inter-individual associations, have demonstrated superior predictive performance across various cancer types [17-19]. In current clinical practice and related research, prediction models used to support NACT

decision-making are typically constructed from clinicopathologic characteristics, laboratory indicators, or imaging-based assessments [20]. Commonly applied tools include biomarker-based approaches such as the CA-125 ELIMination rate constant K (KELIM) score [12] and clinicoradiologic scoring systems such as the Suidan score [21].

Previous predictive models typically handle each patient independently and assume features are uncorrelated [20,22,23]. This limits their ability to capture complex relationships in heterogeneous multimodal data and overlooks population-level patterns that may influence individual outcomes. A graph convolutional network (GCN) addresses these limitations by representing patients as nodes in a population graph, with edges encoding similarities based on multimodal features. This allows the model to integrate both individual characteristics and inter-patient relationships, capturing subtle patterns that may affect treatment response and prognosis. Clinically, this is important because patients with similar baseline features can still exhibit heterogeneous outcomes, making population-informed prediction a valuable approach. Dependence on high-performance computing resources has become a major barrier to the clinical deployment of current deep learning methods [24,25]. In contrast, the GCN can be trained and deployed using clinically accessible central processing unit (CPU) resources, facilitating its use in real-world clinical environments without reliance on specialized graphics processing unit (GPU) infrastructure.

In this study, we developed and optimized a GCN model based on a population graph constructed from multimodal medical data to predict two clinically meaningful outcomes in patients with ovarian cancer undergoing NACT. Overall survival (OS) was defined as the primary endpoint, representing the long-term prognostic outcome, while the probability of achieving R0 resection was defined as the secondary endpoint, reflecting the

key short-term surgical outcome after NACT. Together, these endpoints capture both the long-term impact of therapy and the immediate response to NACT, providing a clinically relevant framework for treatment decision-making. To provide a comprehensive and clinically meaningful comparison framework, we included clinical, radiomic, and combined clinical–radiomic models, alongside the KELIM score, as the primary comparator models. Additionally, the Suidan score, a widely used clinical scoring system for predicting cytoreductive outcomes, was incorporated as an extra comparator specifically for short-term surgical outcome assessment. The model was rigorously trained to forecast clinical outcomes, and its performance was evaluated across multiple external datasets and compared with a series of current advanced models to assess its added value over conventional clinical, biomarker, and radiomics-based approaches.

Results

Overview of the framework

We collected data from 853 patients with ovarian cancer treated with NACT combined with ICS across multiple centers. The detailed patient inclusion and exclusion criteria are illustrated in Figure 1. The training dataset consisted of 489 participants, with two external test datasets 1 and 2 comprising 198 and 166 patients, respectively. The overall study workflow is outlined in Figure 2.

Baseline characteristics of the study population in the training and two external test datasets are summarized in Table 1. The median (interquartile range [IQR]) age of the patients was 58 (51–64) in the training dataset, 59 (53–64) in external test dataset 1, and 61 (53–67) in external test dataset 2. Most patients in all three datasets were diagnosed with serous ovarian cancer, accounting for 77.51% (n = 379), 85.86% (n = 170), and 88.55% (n = 147), respectively. High-grade tumors were predominant, with proportions of

73.42% (n = 359), 80.30% (n = 159), and 72.29% (n = 120) across the three datasets. All included patients were diagnosed with International Federation of Gynecology and Obstetrics (FIGO) stage III or IV disease. The median (IQR) OS was 34.57 (20.07–58.90), 27.02 (14.35–50.30), and 37.50 (23.97–60.03) months for the training, external test dataset 1, and external test dataset 2, respectively.

Evaluation of prognostic efficacy across multiple models

In the training dataset (five-fold cross validation), pairwise association encoder (PAE) achieved a sensitivity and specificity of 93.63% and 85.34%, respectively. The sensitivity and specificity were 90.26% and 86.62% and 88.92% and 90.18%, respectively, for external test datasets 1 and 2, respectively. For OS prediction, the GCN model achieved a C-index of 0.73 (95% confidence interval [CI]: 0.67–0.79) in the training dataset, and 0.72 (95% CI: 0.65-0.79) and 0.70 (95% CI: 0.59–0.81) in the two external test datasets 1 and 2, respectively (Figure 3a). The prognostic performance of the GCN model was compared with that of the four independent comparator models (clinical, radiomic, combined clinical–radiomic, and KELIM). In both the training dataset and two independent external test datasets, the GCN model consistently achieved higher C-index values than the other models, with all comparisons yielding statistically significant differences ($P < 0.01$, Figure 3a and Table S1). Comparison of baseline characteristics between the high- and low-GCN score groups showed that patients in the high-GCN score group had significantly longer overall survival and were relatively younger. No significant differences were observed between the two groups with respect to other clinical characteristics, including histologic subtype, tumor grade, and FIGO stage (Table S2). Further, we evaluated each model's ability to predict 3-year OS using the area under the receiver operating characteristic curve (AUC). The GCN model achieved the highest AUC values across the datasets: 0.88 (95% CI: 0.84–0.92), 0.84 (95% CI: 0.78–0.89), and 0.76 (95%

CI: 0.67–0.85), respectively, outperforming all comparator models ($P < 0.001$ for the training and external test dataset 1, and $P < 0.05$ for the external test dataset 2, Figure 3b–d and Table S3).

Kaplan–Meier (K–M) analysis showed that patients in the high GCN prediction score group had significantly longer OS in both the training and external test datasets (training dataset: hazard ratio [HR] = 0.36 [95% CI: 0.28–0.46], P < 0.001; external test dataset 1: HR = 0.31 [95% CI: 0.21–0.44], P < 0.001; external test dataset 2: HR = 0.31 [95% CI: 0.20–0.51], P < 0.001; Figure S1). KELIM ≥1 was associated with longer OS and showed statistical significance in the training dataset and external test dataset 2 (training dataset: HR = 0.73 [95% CI: 0.56–0.96], $P = 0.03$; external test dataset 2: HR = 0.55 [95% CI: 0.31–0.98], $P = 0.04$; Figure S2); however, its prognostic capability was inferior to that of the GCN model ($P < 0.001$). For the combined model, the high prediction score group was associated with shorter OS only in the training dataset and external test dataset 2 under the combined model (training dataset: HR = 1.94 [95% CI: 1.45–2.60], $P < 0.001$; external test dataset 2: HR = 2.04 [95% CI: 1.01–4.15], $P = 0.04$; Figure S3).

Further, we explored whether the prediction scores differed significantly between patients who achieved R0 resection and those who did not. In the training dataset and both external test datasets, the GCN and KELIM scores showed statistically significant differences between the R0 and non-R0 groups (training dataset: $P = 0.04$, $P < 0.001$; external test dataset 1: $P < 0.01$, $P < 0.001$; external test dataset 2: $P = 0.04$, $P < 0.01$). The Suidan score showed statistically significant differences only in the two external test datasets, whereas no significant difference was observed in the training dataset (training dataset: $P = 0.18$; external test dataset 1: $P = 0.04$; external test dataset 2: $P = 0.03$). The combined model showed a significant difference only in the training dataset. In contrast, the clinical-only and radiomic-only

models did not demonstrate significant discrimination between R0 and non-R0 groups in any of the datasets (Figure 3e–g and Table S4). A systematic SHAP analysis was performed for the GCN score, clinical score, radiomics score, combined clinical–radiomic score, and KELIM score (Figure S4). The SHAP results showed that the GCN score exhibited the widest distribution of SHAP values and the most pronounced separation in both magnitude and direction, indicating the strongest individual-level contribution to outcome prediction among all evaluated models. Calibration curves and decision curve analysis were conducted, and the results are shown in Figure S5.

To assess whether the graph structure provided additional prognostic value beyond the underlying features, ablation experiments were conducted; detailed results are provided in Supplementary Appendix A1. In addition, we performed a comparative analysis benchmarking the graph convolutional network against a standard deep learning architecture without graph connectivity (ResNet50) using identical prognostic endpoints, in order to explicitly isolate the contribution of the graph structure itself; the results are summarized in Supplementary Table S5.

GCN identified favorable-prognosis patients with low KELIM scores

Given that the KELIM score is widely used for prognostic assessment in patients with ovarian cancer receiving NACT, we conducted a detailed comparison between the GCN model and the KELIM score in predicting OS. Notably, the GCN model identified a subset of patients with low KELIM scores who exhibited similar clinical benefits from NACT as those with high KELIM scores. This finding was consistently observed in the training dataset and both external test datasets (Figure 4a–c). These patients with low KELIM but favorable-prognosis identified by the GCN model accounted for 17.79% (n = 87/489) of the training dataset, and 15.66% (n = 31/198) and 12.65% (n = 21/166) of the two external test datasets, respectively. Overall, this subgroup represented 16.30% (n = 139/853) of the total study population. Specifically,

irrespective of whether a patient had a high or low KELIM score, those with high GCN prediction scores had significantly better OS, whereas those with low GCN prediction scores had worse outcomes.

Furthermore, we evaluated the ability of both the GCN model and the KELIM score to predict short- and long-term survival. The GCN model demonstrated strong predictive performance for 1-, 3-, and 5-year OS, with AUCs of 0.80 (95% CI: 0.74–0.86), 0.88 (95% CI: 0.84–0.92), and 0.74 (95% CI: 0.69–0.79) in the training dataset. Similar superior performance was observed in the external test datasets, with the GCN consistently outperforming the KELIM score across all time points (Figure 4d–i and Table S6). The GCN model demonstrated a 14.91–31.74% improvement in 3-year OS AUC over the KELIM score across multiple datasets. Time-dependent AUCs at 1-, 3-, and 5-year overall survival were compared between the GCN model and KELIM across different datasets, with multiplicity adjustment using the Holm–Bonferroni method (Table S7). Compared with patients with a high KELIM score, those in the high GCN score group exhibited significantly longer OS (49.67 vs. 34.77 months, $P < 0.001$). Additionally, we compared survival probabilities between high and low GCN score groups at different time points following NACT (1, 2, 3, 4, and 5 years). High GCN scores were consistently associated with significantly better OS at all time points ($P < 0.001$, Figure 4j). In contrast, high KELIM scores failed to consistently predict improved survival across the time points (Figure 4k).

Monotonic relationship between GCN scores and outcomes

To explore the relationship between GCN scores and outcomes, we stratified participants into four quartiles based on increasing GCN scores (0–25%, 25–50%, 50–75%, and 75–100%). A clear positive trend was observed: as the GCN score increased, OS significantly improved. Each adjacent quartile comparison showed statistically significant differences, with the HR between

the lowest (0–25%) and highest (75–100%) quartiles reaching 6.94 (95% CI: 5.25–9.17, $P < 0.001$) (Figure 5a). Conversely, when participants were stratified into quartiles based on increasing KELIM scores, no such monotonic trend was observed (Figure 5b), and the survival difference between the lowest and highest quartiles was significantly lower than that observed with the GCN model (HR: 2.04 [95% CI: 1.51–2.76] and 6.94 [95% CI: 5.25–9.27], $P < 0.001$). Subsequently, we investigated the relationship between the GCN scores of participants and their short-term outcomes. We found that as the GCN score increased, the probability of achieving R0 resection continued to rise, ranging from 0% to nearly 100%. A similar monotonic relationship was observed between the KELIM score and the probability of achieving R0 resection; however, for patients with a probability of R0 resection less than 25%, the KELIM score was unable to identify them. (Figure 5c–d).

Independent prognostic value of the GCN model

To evaluate whether the GCN model has independent prognostic value, we performed multivariate Cox regression analyses in the training dataset and two independent external test datasets. The results demonstrated that patients classified into the high prediction score group by the GCN model had significantly better OS, even after adjusting for potential confounders, including age, tumor stage, grade, and histological subtype (training dataset: HR = 0.32 [95% CI: 0.23–0.43], $P < 0.001$; external test dataset 1: HR = 0.30 [95% CI: 0.19–0.47], $P < 0.01$; external test dataset 2: HR = 0.17 [95% CI: 0.07–0.40], $P < 0.001$) (Figure 6). In contrast, the KELIM score was significantly associated with better OS only in the training dataset and external test dataset 2, and in both datasets, the HRs for KELIM were higher than those for the GCN model.

Discussion

In this study, we developed a population-based GCN model capable of identifying individuals with low KELIM scores who are likely to significantly benefit from NACT. The proposed model effectively integrates multimodal data, leveraging both radiological features and routinely collected clinical non-imaging information to model inter-individual relationships through an adaptive population graph. Using GCNs, the model enables accurate prediction of long-term outcomes (OS) and stratification of short-term surgical outcomes (R0) in patients with ovarian cancer undergoing NACT. Compared with multiple models, our GCN framework demonstrated superior predictive performance. Unlike previous deep learning models that relied on the efficacy of GPUs [26,27], our GCN model was trained and tested entirely on CPU, substantially reducing computational resource consumption in practical clinical settings. This design enhances its potential for clinical implementation on current low-cost hospital workstations.

The GCN-derived prognostic scores significantly outperformed the widely used KELIM score in predicting NACT outcomes. Across multiple datasets, the GCN model achieved a 14.91% to 31.74% increase in 3-year OS AUC compared to the KELIM score. K–M analyses showed that the HRs between the high- and low-prediction score groups defined by the GCN scores were consistently more favorable than those defined by KELIM. Notably, the GCN model identified a subgroup of patients with low KELIM scores but favorable prognoses—comparable to those with high KELIM scores—accounting for 16.30% of the total study population. A previous study reported a median OS of 40.8 months in patients with a KELIM score >1 [12]. In the present study, the high-score group predicted by the GCN model showed a median OS of 49.67 months, extending survival by 8.87 months compared to prior KELIM-based estimates. Additionally, the C-index of our GCN model for OS prediction surpasses not only that of the KELIM score in this study but also previously reported KELIM performance metrics (test dataset: 0.70 vs. 0.60 and 0.62) [28].

Furthermore, the GCN score exhibited a monotonic association with both OS and the probability of achieving R0 resection. After adjusting for multiple variates, the GCN score remained an independent prognostic factor ($P < 0.001$). These findings may assist clinicians in better identifying patients who are likely to benefit from NACT and support more informed treatment decision-making.

Previous research has demonstrated the potential of multimodal data integration for individualized prognosis prediction across various diseases [29-33]. In contrast to prior studies that were limited to single-source data [16,34], our model synergistically integrates radiomic and non-radiomic data. Moreover, rather than benchmarking against a single comparator, we comprehensively evaluated our model against four alternatives: a clinical model, a radiomic model, a combined clinical–radiomic model, and the KELIM score [35-37].

Predicting cytoreduction outcomes in ovarian cancer remains a clinical challenge. Additionally, selecting optimal candidates for cytoreductive surgery is an unmet need [38]. Although several tools have been proposed, their clinical utility remains limited, and they have not been widely adopted [39,40]. Compared with existing approaches [41,42], our study incorporates both short-term (R0) and long-term (OS) outcome measures. Unlike other prognostic indicators reported in previous studies [43], the GCN score not only stratified patients effectively according to long-term prognosis but also showed statistically significant differences in R0 resection rates between the high- and low-prediction score groups.

Traditional multi-modal schemes typically concatenate features from different sources for each patient and feed them into a neural network, however, such approach ignores inter-patient relationships and population-level structure. Unlike traditional multivariate Cox models or previous deep-learning

approaches, the GCN leverages non-imaging features to model inter-individual relationships through a patient–patient graph, where edges reflect similarities in baseline laboratory characteristics. Biologically, this captures shared disease mechanisms, such as comparable inflammatory profiles or tumor burden–related abnormalities, allowing the model to identify prognostic patterns that traditional models may miss. Clinically, it enables the model to match information from similar patients, improving outcome prediction and risk stratification in heterogeneous populations receiving NACT, even when individual features alone are not sufficiently discriminative.

Our ablation results clarify that the superiority of the GCN stems from the relational information encoded by the graph rather than from feature quantity alone. When the similarity structure was attenuated, prognostic performance declined even though the underlying clinical variables were unchanged, indicating that the graph captures dependencies not discerned from individual features. Moreover, once the similarity network reached adequate fidelity, expanding the feature set no longer improved GCN performance. These findings show that graph-based representation learning contributes distinct and meaningful prognostic value beyond traditional feature-based modelling.

A low GCN score predicts limited benefit from standard NACT, thus prompting a multidisciplinary reassessment to optimize management. The primary recommendation is to re-evaluate feasibility of primary cytoreduction, with surgery indicated if resectability is confirmed. For patients deemed unsuitable for surgery, GCN-guided risk stratification may support consideration of alternative strategies—such as targeted therapy combinations[44,45], rather than persistence with a predicted suboptimal standard approach. Notably, the GCN model identified 16.30% of patients who exhibited favorable survival despite having low KELIM scores, indicating that the GCN can facilitate more precise

stratification of patients for treatment selection. While prospective validation is required, these findings highlight the potential of GCN-guided, risk-adapted treatment pathways to move beyond one-size-fits-all strategies.

This study has certain limitations. First, all analyses were retrospective, and prospective validation in multi-institutional cohorts is needed to confirm the generalizability and clinical utility of the model. Second, the current model relies on laboratory and imaging features without incorporating molecular biomarkers such as HRD status and BRCA mutations [46,47]. While we conducted comprehensive searches of publicly available datasets (TCGA and TCIA) and our institutional database, no cohorts were identified that contained concurrent pretreatment CT imaging, complete clinical laboratory parameters, and molecular/genetic profiles. Previous studies have demonstrated that models integrating molecular-level and imaging data achieve strong predictive performance while also elucidating underlying biological mechanisms[48,49]. Future iterations of our graph neural network should incorporate these molecular biomarkers as such multimodal datasets become increasingly available, aligning the model with precision oncology paradigms for ovarian cancer management. Third, although neoadjuvant chemotherapy regimens were largely standardized across centers and predominantly based on platinum–taxane combinations, some variations in specific regimens among centers may potentially contribute to inter-patient heterogeneity in prognosis. Additionally, although the model is computationally efficient, its practical deployment in clinical settings faces challenges. One such challenge is the integration of radiomics feature extraction tools into existing clinical software environments, such as picture archiving and communication systems. Once this integration is achieved, the GCN can be smoothly embedded into clinical practice without requiring hospitals to invest in additional computational resources.

Future research should address these limitations by conducting prospective trials and large-scale external validation. Incorporating multi-omics and longitudinal data could enhance both predictive robustness and mechanistic understanding. Finally, building a clinical software prototype that integrates seamlessly into hospital information systems could pave the way for the translation of graph-based models into practical tools for risk stratification and treatment planning in ovarian cancer.

In conclusion, this study proposes a method that integrates both imaging features and clinical variables into a CPU-compatible graph network, assisting in improving the accuracy of clinical NACT decision-making for ovarian cancer. Both the training dataset and the two external validation datasets showed that the proposed GCN holds the potential to refine classifications for patients with low KELIM scores, effectively identifying those who are likely to benefit from NACT and thereby enhancing patient survival benefits. The advantages of the GCN in multimodal data modeling, compatibility with low-cost computational resources, and patient re-stratification underscore its potential to guide precise clinical decisions in ovarian cancer treatment.

Methods

Patient cohorts

This multicenter retrospective study utilized data collected between January 1, 2012 and March 31, 2024 from three tertiary hospitals in China. A total of 78 laboratory variables commonly used in routine gynecological practice were selected for pairwise association encoder [50] construction. In addition to laboratory indicators, pre-treatment CT images were obtained for radiomic feature extraction, which together with laboratory variables were used in GCN development. Data from Shengjing Hospital of China Medical University were used as the training dataset, whereas data from Liaoning Cancer Hospital &

Institute and the First Hospital of China Medical University served as external test datasets 1 and 2, respectively. The enrolled patients were all pathologically diagnosed with ovarian cancer and received NACT prior to ICS. The inclusion criteria were: (1) age ≥18 years; (2) availability of a pre-NACT pelvic CT scan within 1 month before NACT; and (3) availability of laboratory test results within 1 month before NACT. The exclusion criteria were (1) insufficient image quality; (2) presence of systemic inflammatory or other major diseases; (3) concurrent other malignancies; and (4) missing clinical data.

Approval for the study from the institutional ethics review committee at the Shengjing Hospital of China Medical University (approval number 2024PS969K), the First Hospital of China Medical University (approval number [2024] 560), and Liaoning Cancer Hospital and Institute (approval number LH20250304). This work was complied with principles of the Helsinki Declaration. All patient data were anonymized, and the requirement for written informed consent was waived by the ethics committees.

Clinical data collection

Demographic, clinicopathological, and laboratory data were extracted from the electronic medical records of the three retrospective cohorts. A comprehensive set of routine gynecologic laboratory indicators (including both blood and urine tests obtained before the initiation of NACT) was initially collected. To ensure data quality and reduce bias from missingness, laboratory variables with more than 20% missing values were excluded during preprocessing [51,52] , resulting in 78 laboratory indicators used for model construction. For the remaining features, missing values were imputed in a center-wise and feature-wise manner, using the median of each feature calculated within the corresponding center. All imputations were performed independently for each center to avoid potential data leakage. Laboratory

measurements from different centers were standardized before analysis. The laboratory indicators and their missingness are detailed in Table S8. Tumor staging was based on the International Federation of Gynecology and Obstetrics (FIGO) criteria for ovarian cancer. OS was defined as the time from histopathological diagnosis to death from any cause or the last follow-up. R0 status was defined as the condition where cytoreduction resulted in no macroscopic residual disease [3].

Image segmentation and radiomic feature

All CT images from the three centers were retrieved from their respective picture archiving and communication systems. Tumor regions of interest were manually segmented using open-source software (ITK-SNAP, http://www.itksnap.org/) with an abdominal window setting (level: 50, width: 400). Two experienced radiologists independently performed the segmentations, and discrepancies were resolved by a senior radiologist with over 20 years of experience. Radiomic features were extracted using the PyRadiomics [53]. All CT volumes were resampled to an isotropic resolution of 1×1×1 mm^3 using B-spline interpolation, and gray-level discretization was performed using a fixed bin width of 25. Feature extraction was performed on the original images, on Laplacian of Gaussian (LoG) filtered images with sigma values of 1.0–5.0 mm, and on all eight-wavelet decomposition sub-bands. A total of 1,218 radiomic features were generated, including first-order, shape-based, and texture features, as well as LoG and wavelet-derived features. The complete feature list is provided in Table S9. All features were standardized using Z-scores, and highly collinear features with a Pearson correlation coefficient >0.9 were excluded from further analysis.

Prediction model construction

To construct the population graph, both pre-treatment laboratory variables and radiomic features from patients who received NACT were incorporated. The

PAE was used to estimate inter-patient similarities based on non-radiomic data (e.g., laboratory results), which defined the edge weights (W_{ij}) in the population graph. Supervised learning was performed using a 90-day[54,55] survival difference as the cutoff value, and the PAE predicted the survival similarity between patients based on their laboratory indicators. The PAE constructed 59622, 10118, and 9322 graphical connections for the training and two external test datasets, respectively. Subsequently, radiomic features served as node attributes within the graph. After constructing the adaptive population graph, it was used in the graph convolutional neural networks[56] for graph representation learning and prognosis prediction of patients undergoing NACT. The model architectures and schematic illustrations of the PAE and GCN are provided in the Supplementary Methods. Patients were stratified into high- and low-prediction score groups based on the median prediction score in the training dataset to evaluate differences in survival outcomes. The entire training and testing process of the model was conducted on the CPU, without the need for high-performance GPU hardware.

For comparison, four independent prognostic models were constructed: a clinical model, a radiomic model, a combined clinical–radiomic model, and the KELIM score. To complement this prognostic comparison framework with an assessment of short-term cytoreductive outcomes, the Suidan score was incorporated as an auxiliary comparator. The Suidan score was calculated according to its original criteria using preoperative clinical and radiologic parameters[21]. In this study, it was used solely as an additional comparator for short-term surgical outcome analysis and was not considered one of the primary prognostic models. The clinical model was constructed using clinicopathological features and laboratory variables. Laboratory data were standardized via Z-score normalization. Significant predictors ($P < 0.05$) were identified through univariate and multivariate Cox regression analyses and subsequently used to calculate clinical prediction scores. The radiomic

prediction model was constructed based on radiomic features. These radiomic features were standardized using Z-scores, and collinear features with Pearson correlation coefficients >0.9 were excluded. Subsequently, univariate and multivariate Cox regression analyses were performed to select significant radiomic features for building the radiomic prediction model and calculating prediction scores. A combined clinical–radiomic model was also developed by incorporating both clinical and radiomic predictors using the same Cox-based approach. A detailed description of the variables included in the clinical, radiomic, and combined clinical–radiomic predictive models is provided in Supplementary Appendix A2. In addition, the commonly used KELIM score in current research was included as a reference model. The KELIM score, calculated using an online tool (https://www.biomarker-kinetics.org/CA-125-neo) [57], requires at least three CA-125 measurements taken within the first 100 days (or fewer) of chemotherapy initiation.

Statistical analysis

All statistical analyses were conducted using R software (http://www.R-project.org/). The software versions and R packages used in this study are provided in Table S10 and Table S11, respectively. Continuous variables are presented as medians with IQRs, whereas categorical variables are expressed as counts and percentages. Univariate and multivariate Cox proportional hazards models were used to estimate HRs and 95% CIs. The multivariate analysis was adjusted for potential confounders, including age, histological subtype, clinical stage, tumor grade, and KELIM score. For KELIM, patients were stratified into high- and low-prediction score groups using a cutoff of 1, as previously described [28,38]. For all other models, patients were stratified based on the median predicted score from the training dataset to evaluate differences in survival outcomes. All OS-related analyses (C-index, time-dependent AUCs, Kaplan–Meier curves, HRs, risk-stratified survival comparisons, and monotonicity analyses) are presented as

performance assessments for the primary endpoint. Analyses comparing prediction scores between R0 and non-R0 groups and examining monotonic trends in R0 probability evaluate performance for the secondary endpoint. All statistical tests were two-sided, and statistical significance was set at $P < 0.05$. The code related to this study has been released on GitHub repository (https://github.com/jessiezhang1021/GCN-OC).

Acknowledgments

This work was supported by the National Natural Science Foundation of China [82072885, 92259104]; Noncommunicable Chronic Diseases-National Science and Technology Major Project [2025ZD0545600, 2025ZD0545601]; Xingliao Talent Program of Liaoning Province [XLYC2403102] and the Science and Technology Plan Joint Plan of Liaoning Province [2023JH2/101700193].

Data Availability Statement

The datasets used and analyzed in this study are available from the corresponding author upon reasonable request.

Code Availability Statement

The code related to this study has been released on GitHub repository (https://github.com/jessiezhang1021/GCN-OC).

Competing Interests

The authors declare no competing interests.

Author Contributions

Designing the study: SM.Z., JD.S., X.Z. Methodology: SM.Z., YL.L., JD.S., X.Z. Data curation: SM.Z., ZN.L., XY.L., YT.L., MY.L., JR.W., JG.Z. Writing—Original Draft: SMZ. Writing—Review & Editing: JD.S., X.Z., YL.L. Funding acquisition: JD.S., X.Z. Resources: G.W., Z.Y., Y.Z., J.S., X.Z. All authors read and approved the final manuscript.

References

1 *Cancer stat facts: ovarian cancer 2025*, <https://seer.cancer.gov/statfacts/html/ovary.html> (

2 Siegel, R. L., Giaquinto, A. N. & Jemal, A. Cancer statistics, 2024. *CA Cancer J Clin* **74**, 12-49, doi:10.3322/caac.21820 (2024).

3 Gaillard, S. *et al.* Neoadjuvant Chemotherapy for Newly Diagnosed, Advanced Ovarian Cancer: ASCO Guideline Update. *J Clin Oncol* **43**, 868-891, doi:10.1200/JCO-24-02589 (2025).

4 Bercow, A. *et al.* Utilization of Primary Cytoreductive Surgery for Advanced-Stage Ovarian Cancer. *JAMA Netw Open* **7**, e2439893, doi:10.1001/jamanetworkopen.2024.39893 (2024).

5 Vergote, I. *et al.* Neoadjuvant chemotherapy or primary surgery in stage IIIC or IV ovarian cancer. *The New England Journal of Medicine* **363**, 943-953, doi:10.1056/NEJMoa0908806 (2010).

6 Coada, C. A. *et al.* Optimal number of neoadjuvant chemotherapy cycles prior to interval debulking surgery in advanced epithelial ovarian cancer: a systematic review and meta-analysis of progression-free survival and overall survival. *J Gynecol Oncol* **34**, e82, doi:10.3802/jgo.2023.34.e82 (2023).

7 Cho, J. H., Kim, S. & Song, Y. S. Neoadjuvant chemotherapy in advanced ovarian cancer: optimal patient selection and response evaluation. *Chin Clin Oncol* **7**, 58, doi:10.21037/cco.2018.10.11 (2018).

8 Cioffi, R. *et al.* Neoadjuvant chemotherapy in high-risk ovarian cancer patients: Role of age. *Tumori* **105**, 168-173, doi:10.1177/0300891618792468 (2019).

9 Tajik, P. *et al.* The FIGO Stage IVA Versus IVB of Ovarian Cancer: Prognostic Value and Predictive Value for Neoadjuvant Chemotherapy. *Int J Gynecol Cancer* **28**, 453-458, doi:10.1097/IGC.0000000000001186 (2018).

10 Fagotti, A. *et al.* Randomized trial of primary debulking surgery versus neoadjuvant chemotherapy for advanced epithelial ovarian cancer (SCORPION-NCT01461850). *Int J Gynecol Cancer* **30**, 1657-1664, doi:10.1136/ijgc-2020-001640 (2020).

11 Bryant, A. *et al.* Impact of residual disease as a prognostic factor for survival in women with advanced epithelial ovarian cancer after primary surgery. *Cochrane Database Syst Rev* **9**, CD015048, doi:10.1002/14651858.CD015048.pub2 (2022).

12 You, B. *et al.* CA-125 ELIMination Rate Constant K (KELIM) Is a Marker of Chemosensitivity in Patients with Ovarian Cancer: Results from the Phase II CHIVA Trial. *Clin Cancer Res* **26**, 4625-4632, doi:10.1158/1078-0432.CCR-20-0054 (2020).

13 Kessous, R. *et al.* CA-125 reduction during neoadjuvant chemotherapy is associated with success of cytoreductive surgery and outcome of patients with advanced high-grade ovarian cancer. *Acta Obstet Gynecol Scand* **99**, 933-940, doi:10.1111/aogs.13814 (2020).

14 Liontos, M. *et al.* Neutrophil-to-lymphocyte ratio and chemotherapy response score as prognostic markers in ovarian cancer patients treated with neoadjuvant chemotherapy. *J Ovarian Res* **14**, 148, doi:10.1186/s13048-021-00902-0 (2021).

15 Tuntinarawat, P., Tangmanomana, R. & Kittisiam, T. Association between alteration of neutrophil to lymphocyte ratio, platelet to lymphocyte ratio, cancer antigen-125 and surgical outcomes in advanced stage ovarian cancer patient who received neoadjuvant chemotherapy. *Gynecol Oncol Rep* **52**, 101347, doi:10.1016/j.gore.2024.101347 (2024).

16 Wood, N. *et al.* Association between CT-based body composition assessment and patient outcomes during neoadjuvant chemotherapy for epithelial ovarian cancer. *Gynecol Oncol* **169**, 55-63, doi:10.1016/j.ygyno.2022.11.024 (2023).

17 Mao, N. *et al.* A multimodal and fully automated system for prediction of pathological complete response to neoadjuvant chemotherapy in breast cancer. *Sci Adv* **11**, eadr1576, doi:10.1126/sciadv.adr1576 (2025).

18 Schutte, K. *et al.* An artificial intelligence model predicts the survival of solid tumour patients from imaging and clinical data. *Eur J Cancer* **174**, 90-98, doi:10.1016/j.ejca.2022.06.055 (2022).

19 Chen, S. *et al.* Deep learning-based multi-model prediction for disease-free survival status of patients with clear cell renal cell carcinoma after surgery: a multicenter cohort study. *Int J Surg* **110**, 2970-2977, doi:10.1097/JS9.0000000000001222 (2024).

20 Crispin-Ortuzar, M. *et al.* Integrated radiogenomics models predict response to neoadjuvant chemotherapy in high grade serous ovarian cancer. *Nature communications* **14**, 6756, doi:10.1038/s41467-023-41820-7 (2023).

21 Suidan, R. S. *et al.* A multicenter assessment of the ability of preoperative computed tomography scan and CA-125 to predict gross residual disease at primary debulking for advanced epithelial ovarian cancer. *Gynecol Oncol* **145**, 27-31, doi:10.1016/j.ygyno.2017.02.020 (2017).

22 Yin, R. *et al.* Predicting Neoadjuvant Chemotherapy Response and High-Grade Serous Ovarian Cancer From CT Images in Ovarian Cancer with Multitask Deep Learning: A Multicenter Study. *Acad Radiol* **30 Suppl 2**, S192-S201, doi:10.1016/j.acra.2023.04.036 (2023).

23 AlSomairi, A., Himayda, S., Altelmesani, A., Lee, Y. J. & Lee, J. Y. Prognostic value of HE4 in advanced-stage, high-grade serous ovarian cancer: Analysis of HE4 kinetics during NACT, predicting surgical outcome and recurrence in comparison to CA125. *Gynecol Oncol* **181**, 155-161, doi:10.1016/j.ygyno.2023.12.021 (2024).

24 Dinsdale, N. K. *et al.* Challenges for machine learning in clinical translation of big data imaging studies. *Neuron* **110**, 3866-3881, doi:10.1016/j.neuron.2022.09.012 (2022).

25 Tobore, I. *et al.* Deep Learning Intervention for Health Care Challenges: Some Biomedical Domain Considerations. *JMIR Mhealth Uhealth* **7**, e11966, doi:10.2196/11966 (2019).

26 Pandey, M. *et al.* The transformational role of GPU computing and deep learning in drug discovery. *Nature Machine Intelligence* **4**, 211-221 (2022).

27 Gawehn, E., Hiss, J. A., Brown, J. B. & Schneider, G. Advancing drug discovery via GPU-based deep learning. *Expert opinion on drug discovery* **13**, 579-582 (2018).

28 Bouvarel, B. *et al.* Clinical impact of CA-125 ELIMination rate constant K (KELIM) on surgical strategy in advanced serous ovarian cancer patients. *Int J Gynecol Cancer* **34**, 574-580, doi:10.1136/ijgc-2023-004872 (2024).

29 Keyl, J. *et al.* Decoding pan-cancer treatment outcomes using multimodal real-world data and explainable artificial intelligence. *Nat Cancer* **6**, 307-322, doi:10.1038/s43018-024-00891-1 (2025).

30 Ding, S. *et al.* HGMSurvNet: A two-stage hypergraph learning network for multimodal cancer survival prediction. *Med Image Anal* **104**, 103661, doi:10.1016/j.media.2025.103661 (2025).

31 Baheti, B. *et al.* Multimodal Explainable Artificial Intelligence for Prognostic Stratification of Glioblastoma Patients. *Mod Pathol*, 100797, doi:10.1016/j.modpat.2025.100797 (2025).

32 Tian, R. *et al.* Multimodal fusion model for prognostic prediction and radiotherapy response assessment in head and neck squamous cell carcinoma. *NPJ Digit Med* **8**, 302, doi:10.1038/s41746-025-01712-0 (2025).

33 Wang, L. *et al.* Role of artificial intelligence in medical image analysis. *Chin Med J (Engl)* **138**, 2879-2894, doi:10.1097/CM9.0000000000003824 (2025).

34 Guo, W. *et al.* Patients with macroscopic lymph node metastasis expect poor prognosis after neoadjuvant chemotherapy in advanced ovarian cancer: a retrospective cohort study based on a single gynecological team. *BMC Cancer* **25**, 832, doi:10.1186/s12885-025-14237-2 (2025).

35 Zannoni, G. F. *et al.* Chemotherapy Response Score (CRS): A comprehensive review of its prognostic and predictive value in High-Grade Serous Carcinoma (HGSC). *Gynecol Oncol* **194**, 1-10, doi:10.1016/j.ygyno.2025.01.012 (2025).

36 Li, C. *et al.* CA-125 elimination rate constant K (KELIM) as a promising predictor of complete cytoreduction after neoadjuvant chemotherapy in advanced ovarian cancer patients: a retrospective study from two Chinese hospitals. *BMC Cancer* **24**, 609, doi:10.1186/s12885-024-12252-3 (2024).

37 Trewin-Nybraten, C. B., Leithe, S., Paulsen, T., Langseth, H. & Fortner, R. T. Ovarian cancer survival by residual disease following cytoreductive surgery: a nationwide study in Norway. *Br J Cancer* **132**, 1158-1166, doi:10.1038/s41416-025-03018-0 (2025).

38 Piedimonte, S. *et al.* Validation of the KELIM score as a predictor of response to neoadjuvant treatment in patients with advanced high grade serous ovarian cancer. *Gynecol Oncol* **167**, 417-422, doi:10.1016/j.ygyno.2022.10.014 (2022).

39 Narasimhulu, D. M. *et al.* Using an evidence-based triage algorithm to reduce 90-day mortality after primary debulking surgery for advanced epithelial ovarian cancer. *Gynecol Oncol* **155**, 58-62, doi:10.1016/j.ygyno.2019.08.004 (2019).

40 Kumar, A. *et al.* Models to predict outcomes after primary debulking surgery: Independent validation of models to predict suboptimal cytoreduction and gross residual disease. *Gynecol Oncol* **154**, 72-76, doi:10.1016/j.ygyno.2019.04.011 (2019).

41 Crispin-Ortuzar, M. *et al.* Integrated radiogenomics models predict response to neoadjuvant chemotherapy in high grade serous ovarian cancer. *Nat Commun* **14**, 6756, doi:10.1038/s41467-023-41820-7 (2023).

42 Postl, M. *et al.* The Predictive Value of the Fibrinogen-Albumin-Ratio Index on Surgical Outcomes in Patients with Advanced High-Grade Serous Ovarian Cancer. *Cancers (Basel)* **16**, doi:10.3390/cancers16193295 (2024).

43 Bregar, A. *et al.* CT prediction of surgical outcome in patients with advanced epithelial ovarian carcinoma undergoing neoadjuvant chemotherapy. *Gynecol Oncol* **152**, 568-573, doi:10.1016/j.ygyno.2018.12.012 (2019).

44 Wright, A. A. *et al.* Neoadjuvant Chemotherapy for Newly Diagnosed, Advanced Ovarian Cancer: Society of Gynecologic Oncology and American Society of Clinical Oncology Clinical Practice Guideline. *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology* **34**, 3460-3473, doi:10.1200/JCO.2016.68.6907 (2016).

45 Colombo, N. *et al.* ESMO-ESGO consensus conference recommendations on ovarian

cancer: pathology and molecular biology, early and advanced stages, borderline tumours and recurrent disease†. *Annals of oncology : official journal of the European Society for Medical Oncology* **30**, 672-705, doi:10.1093/annonc/mdz062 (2019).

46 Colombo, N. *et al.* Consensus statements and treatment algorithm to guide clinicians in the selection of maintenance therapy for patients with newly diagnosed, advanced ovarian carcinoma: Results of a Delphi study. *Gynecologic Oncology* **175**, 182-189, doi:10.1016/j.ygyno.2023.05.065 (2023).

47 Ledermann, J. A. *et al.* ESGO-ESMO-ESP consensus conference recommendations on ovarian cancer: pathology and molecular biology and early, advanced and recurrent disease. *Annals of oncology : official journal of the European Society for Medical Oncology* **35**, 248-266, doi:10.1016/j.annonc.2023.11.015 (2024).

48 Fathi Kazerooni, A. *et al.* Multiparametric MRI along with machine learning predicts prognosis and treatment response in pediatric low-grade glioma. *Nat Commun* **16**, 340, doi:10.1038/s41467-024-55659-z (2025).

49 Li, G. *et al.* An MRI radiomics approach to predict survival and tumour-infiltrating macrophages in gliomas. *Brain* **145**, 1151-1161, doi:10.1093/brain/awab340 (2022).

50 Huang, Y. & Chung, A. C. S. Disease prediction with edge-variational graph convolutional networks. *Med Image Anal* **77**, 102375, doi:10.1016/j.media.2022.102375 (2022).

51 Nguyen, T. J. *et al.* Effect of immediate reconstruction on postmastectomy surgical site infection. *Ann Surg* **256**, 326-333, doi:10.1097/SLA.0b013e3182602bb7 (2012).

52 Wang, H. *et al.* Development and Internal Validation of a Nomogram to Predict Mortality During the ICU Stay of Thoracic Fracture Patients Without Neurological Compromise: An Analysis of the MIMIC-III Clinical Database. *Front Public Health* **9**, 818439, doi:10.3389/fpubh.2021.818439 (2021).

53 van Griethuysen, J. J. M. *et al.* Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res* **77**, e104-e107, doi:10.1158/0008-5472.CAN-17-0339 (2017).

54 Abel, M. K. *et al.* Neoadjuvant Chemotherapy, Case Volume, and Mortality in Advanced Ovarian Cancer. *JAMA Netw Open* **8**, e2523434, doi:10.1001/jamanetworkopen.2025.23434 (2025).

55 Aletti, G. D., Dowdy, S. C., Podratz, K. C. & Cliby, W. A. Relationship among surgical complexity, short-term morbidity, and overall survival in primary surgery for advanced ovarian cancer. *American journal of obstetrics and gynecology* **197**, 676. e671-676. e677 (2007).

56 Feng, Y., You, H., Zhang, Z., Ji, R. & Gao, Y. in *Proceedings of the AAAI conference on artificial intelligence.* 3558-3565.

57 *Biomarker-Kinetics: https://www.biomarker-kinetics.org/CA-125-neo*.

Figure and Table Legends

Figure 1

Patient enrollment flowchart.

SJ cohort = patients from Shengjing Hospital of China Medical University, SZL cohort = patients from Liaoning Cancer Hospital & Institute, YDY cohort = patients from the First Hospital of China Medical University. CT, computed tomography; NACT, neoadjuvant chemotherapy.

Figure 2

Study flowchart.

(a) The pairwise association encoder (PAE) was constructed using non-imaging data prior to treatment to compute edge weights (W_{ij}). These were integrated with radiomic features extracted from pre-treatment imaging data to construct a population graph. Graph convolutional networks (GCNs) were then trained to generate prediction scores for each patient. (b) Four additional comparative models were developed, and prediction scores were calculated for each. Patients were stratified into high and low prediction score groups. Model performance was evaluated using multiple metrics, and differences in short-term and long-term outcomes between the two groups were compared. HR, hazard ratio; GCN, graph convolutional network; KELIM, CA-125 ELIMination rate constant K.

Figure 3

Comparison of the predictive performance of multiple models.

(a) Comparison of C-index values for overall survival (OS) prediction across different models. (b–d) Receiver operating characteristic curves for predicting 3-year OS in the training dataset and two external test datasets, along with the corresponding area under the curve values for each model. (e–g) Distribution of prediction scores for different models in the R0 and non-R0 groups within the training dataset and two external test datasets. Box boundaries represent the first and third quartiles; the center line marks the median. Whiskers extend to the furthest non-outlier points within 1.5 times the interquartile range. ‘ref’ indicates reference, ‘*’ indicates $P < 0.05$, ‘**’ indicates $P < 0.01$, ‘***’ indicates $P < 0.001$, ‘NS’ indicates not significant. CI, confidence interval; GCN, graph convolutional network; KELIM, CA-125 ELIMination rate constant K.

Figure 4

Further comparison of the predictive performance of the GCN model and KELIM score.

(a–c) Kaplan–Meier analysis for overall survival (OS) in the training set and two external test sets. The GCN high and low prediction score groups were divided based on the median prediction score in the training set, while the KELIM score high and low groups were divided using a cutoff of 1. ‘L’ indicates low, ‘H’ indicates high, ‘NS’ indicates not significant. (d–f) Time-dependent receiver operating characteristic (ROC) curves and the area under the ROC (AUC) values for 1-year, 3-year, and 5-

year OS predictions using the GCN model in the three datasets. (g–i) Time-dependent ROC curves and AUC values for KELIM score predictions. (j) Comparison of 1-, 2-, 3-, 4-, and 5-year OS based on GCN prediction score stratification, divided into high and low GCN prediction score groups. (k) Comparison of 1-, 2-, 3-, 4-, and 5-year OS based on KELIM score stratification. '*' indicates $P < 0.05$, '***' indicates $P < 0.001$. GCN, graph convolutional network; KELIM, CA-125 ELIMination rate constant K.

Figure 5

Monotonic relationship between GCN prediction scores and outcomes.
(a, b) Kaplan–Meier analysis for GCN prediction scores and KELIM scores, categorized from low to high according to percentiles. (c, d) Relationship between GCN prediction scores and KELIM scores with the probability of achieving R0 resection. The red shaded area represents the 95% confidence interval. HR, hazard ratio; GCN, graph convolutional network; KELIM, CA-125 ELIMination rate constant K.

Figure 6

Independent prognostic value of the GCN model.
(a–c) Forest plots of multivariate Cox regression analyses for overall survival (OS) in the training and external test cohorts. Green circles represent hazard ratios (HRs) less than 1, while orange circles represent HRs greater than 1. The horizontal bars indicate the 95% confidence intervals. 'L' indicates low, 'H' indicates high. HR, hazard ratio.

Table 1

Baseline characteristics of the included population.

Table 1. Baseline characteristics of the included population

Variable	Training dataset	External test dataset 1	External test dataset 2	*P* value [a]	*P* value [b]	*P* value [c]
No. patients	489	198	166			
Age (years)				0.23	<0.01*	0.07
Median (IQR)	58 (51-64)	59 (53-64)	61 (53-67)			
FIGO stage (%)				0.87	0.07	0.15
III	303(61.96)	124(62.63)	116(69.88)			
IV	186(38.04)	74(37.37)	50(30.12)			
Histological subtype (%)				0.09	<0.01*	0.19
Serous	379(77.51)	170(85.86)	147(88.55)			
Other	82(16.77)	24(12.12)	13(7.83)			

N/A	28(5.73)	4(2.02)	6(3.61)			
Tumor grade (%)				0.10	0.06	0.54
I	18(3.68)	16(8.08)	14(8.43)			
II	38(7.77)	14(7.07)	16(9.64)			
III	359(73.42)	159(80.30)	120(72.29)			
N/A	74(15.13)	9(4.55)	16(9.64)			
Complete gross resection (%)				<0.001*	<0.01*	0.26
True	322(65.85)	101(51.01)	94(56.63)			
False	155(31.70)	93(46.97)	68(40.96)			
N/A	12(2.45)	4(2.02)	4(2.41)			
OS (days)				<0.001*	0.29	<0.001*
Median (IQR)	34.57 (20.07-58.90)	27.02 (14.35-50.30)	37.50 (23.97-60.03)			
CA125 (U/mL)				0.02*	0.79	0.08
Median (IQR)	1106.00 (515.00-2540.00)	1182.70 (670.90-3871.00)	1115.00 (575.00-2668.00)			
HE4 (pmol/L)				0.87	0.03*	0.03*
Median (IQR)	732.00 (341.80-1500.00)	700.20 (346.30-1375.20)	548.10 (247.00-866.50)			

IQR, Interquartile range; OS, overall survival; CA125, cancer antigen 125; HE4, human epididymis protein 4. 'a' indicates the comparison of baseline characteristics between the training dataset and external test dataset 1, 'b' indicates the comparison between the training dataset and external test dataset 2, 'c' indicates the comparison between external test dataset 1 and external test dataset 2, '*' denotes P-value less than 0.05.

Patients with ovarian cancer after NACT(2012-2024)

Inclusion criteria:
(1) Age ≥ 18
(2) Pelvic CT within 1 month prior to NACT
(3) Laboratory test results within 1 month prior to NACT

Center 1 (SJ Cohort)
n=547

58 excluded
18 with insufficient image quality
12 with systemic inflammation or major diseases
11 with other malignancies
17 with missing clinical data

Training set
n=489

Center 2 (SZL Cohort)
n=255

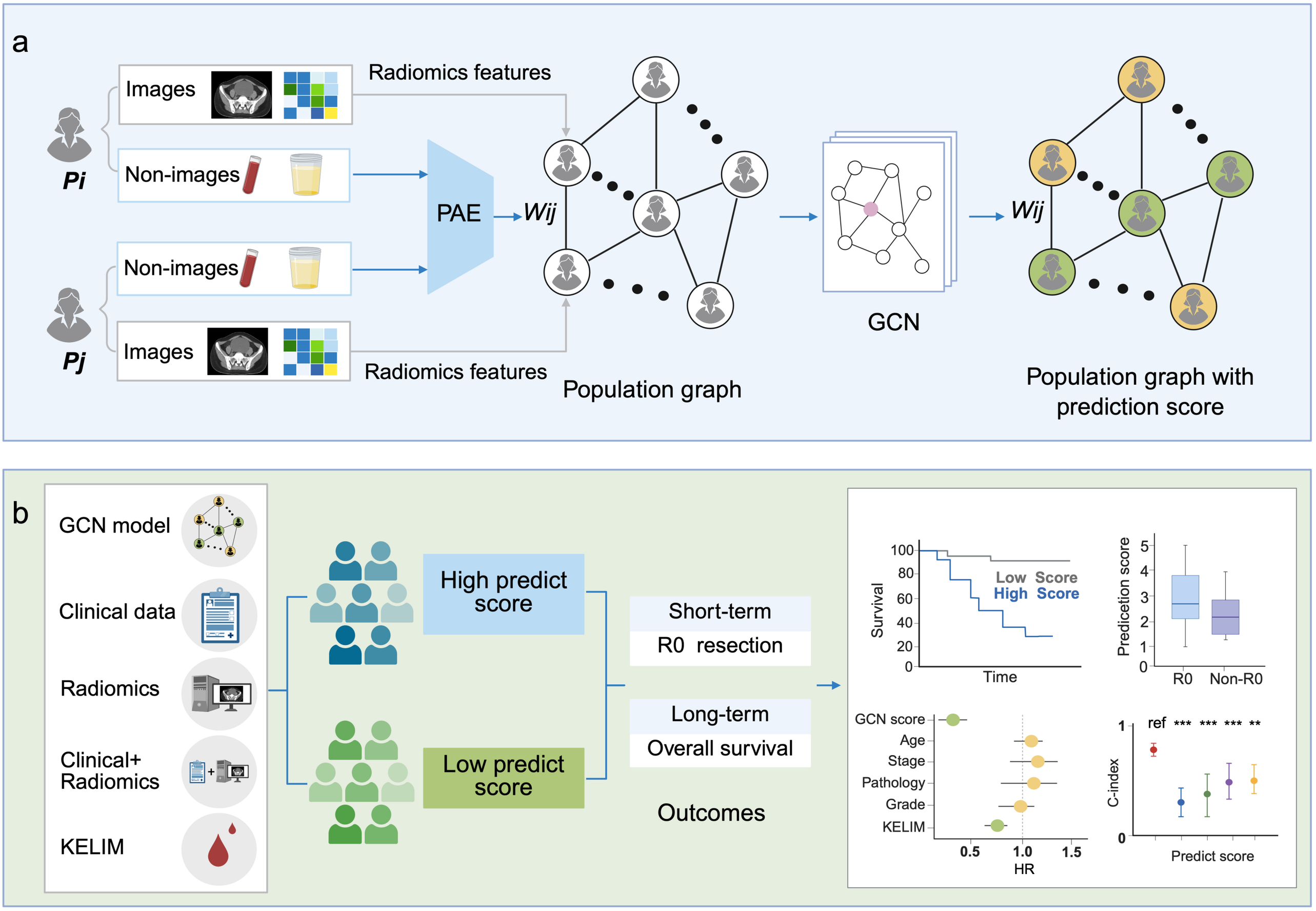
47 excluded
17 with insufficient image quality
8 with systemic inflammation or major diseases
16 with other malignancies
6 with missing clinical data

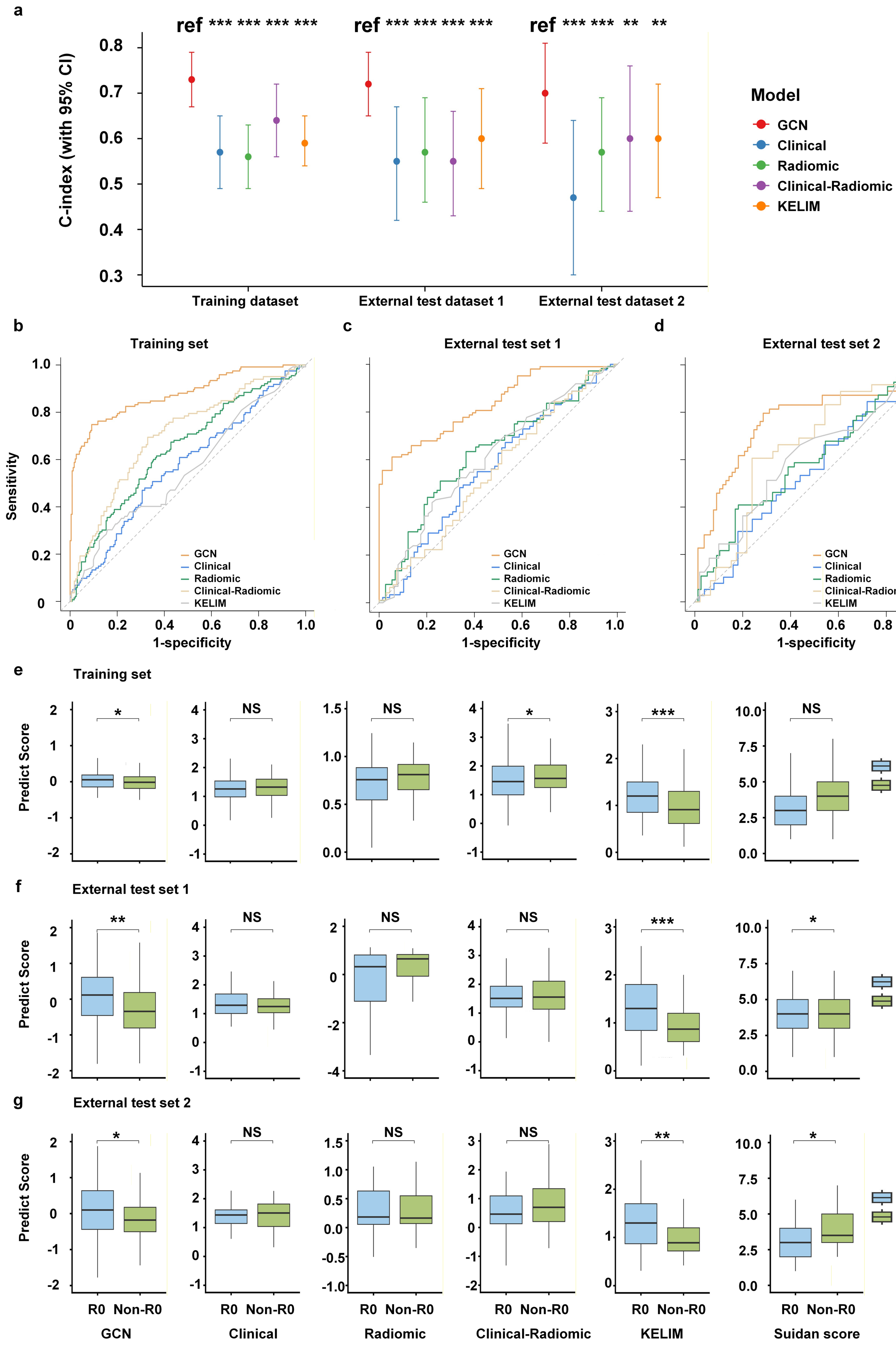
External test set 1
n=198

Center 3 (YDY Cohort)
n=205

39 excluded
11 with insufficient image quality
8 with systemic inflammation or major diseases
11 with other malignancies
9 with missing clinical data

External test set 2
n=166


a
Images
Non-images
Pi
Non-images
Images
Pj
Radiomics features
Radiomics features
PAE
Wij
Population graph
GCN
Wij
Population graph with prediction score
b
GCN model
Clinical data
Radiomics
Clinical+ Radiomics
KELIM
High predict score
Low predict score
Short-term
R0 resection
Long-term
Overall survival
Outcomes
Survival
Time
Low Score
High Score
Predicetion score
R0
Non-R0
GCN score
Age
Stage
Pathology
Grade
KELIM
HR
C-index
ref
Predict score


a
ref *** *** *** ***
ref *** *** *** ***
ref *** *** ** **
C-index (with 95% CI)
0.8
0.7
0.6
0.5
0.4
0.3
Training dataset
External test dataset 1
External test dataset 2
Model
GCN
Clinical
Radiomic
Clinical-Radiomic
KELIM
b
Training set
c
External test set 1
d
External test set 2
Sensitivity
1-specificity
e
Training set
f
External test set 1
g
External test set 2
Predict Score
R0
Non-R0
GCN
Clinical
Radiomic
Clinical-Radiomic
KELIM
Suidan score

a

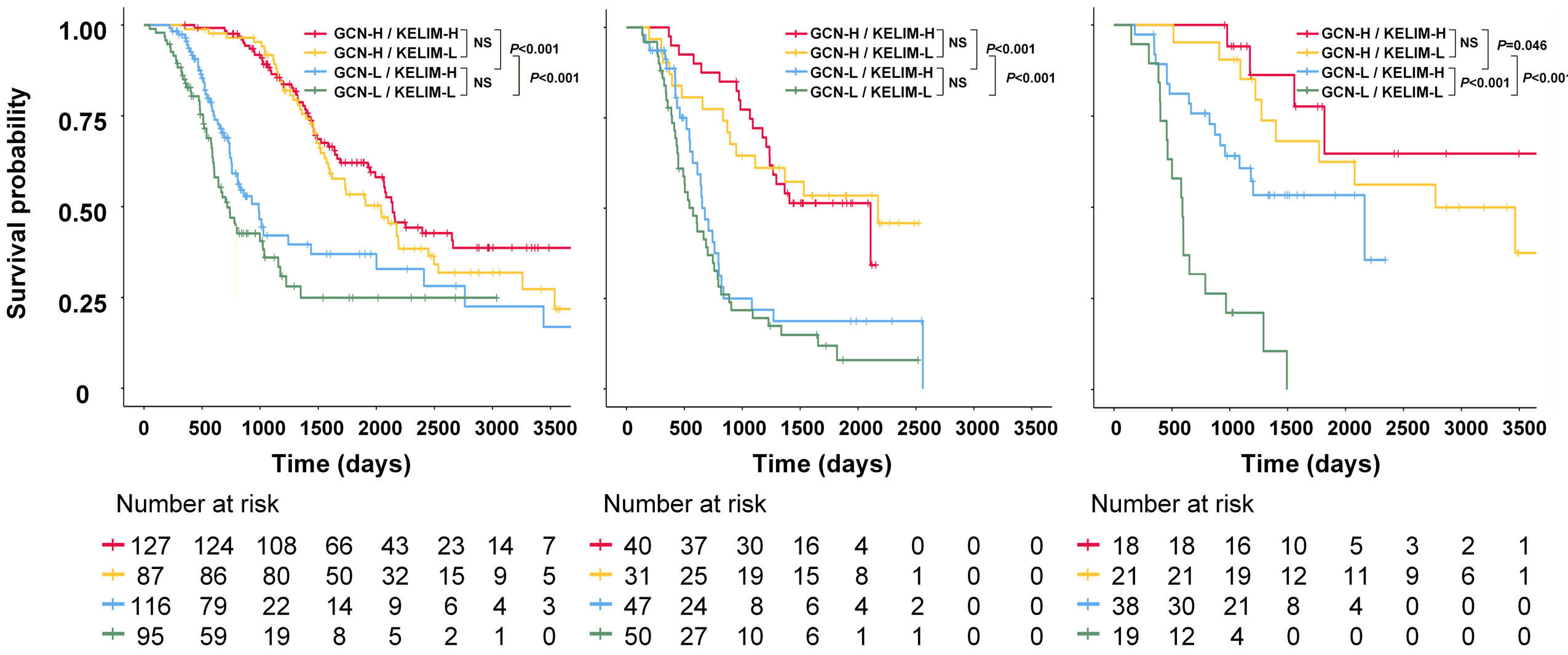
Training set

b

External test set 1

c

External test set 2



d

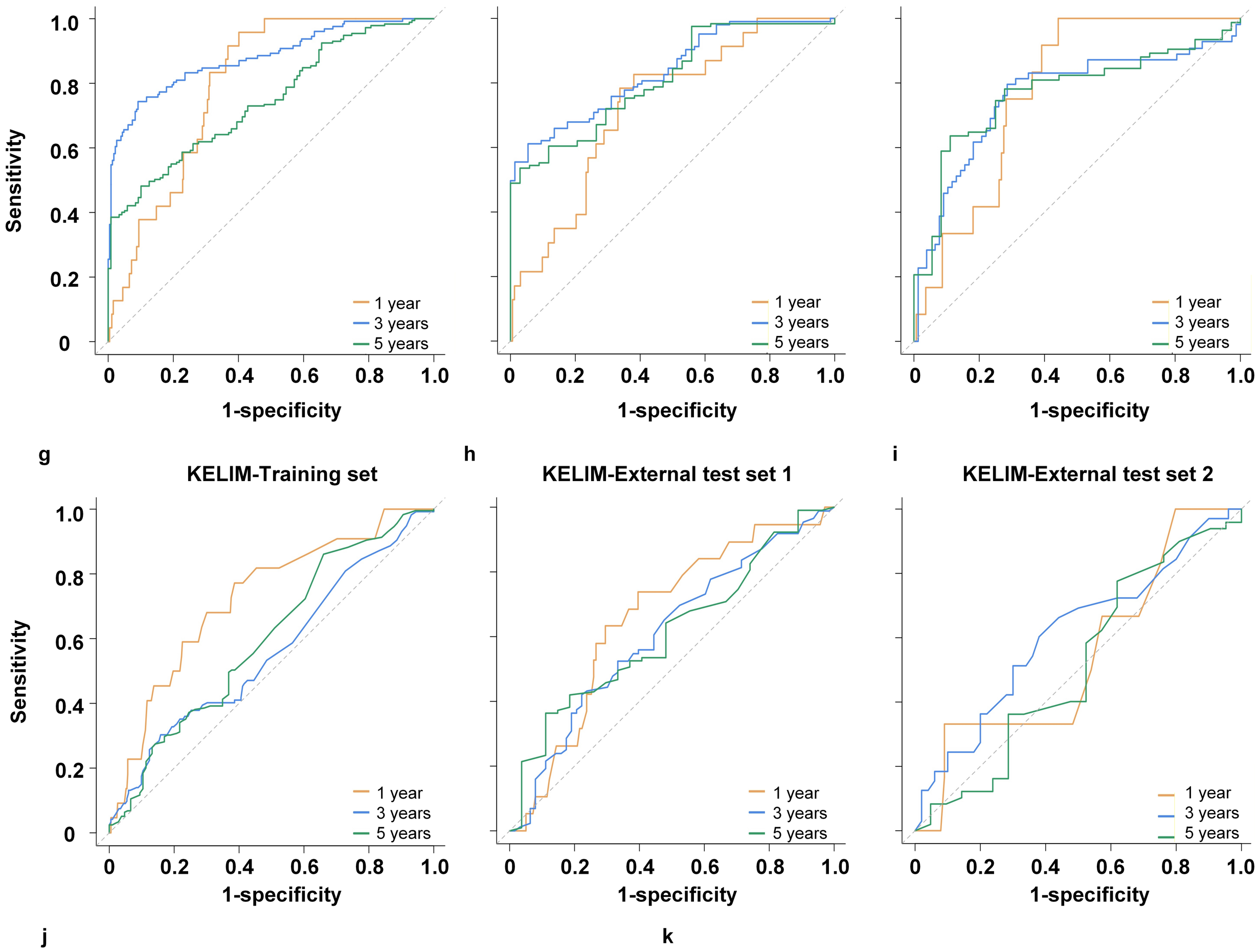
GCN-Training set

e

GCN-External test set 1

f

GCN-External test set 2



g

KELIM-Training set

h

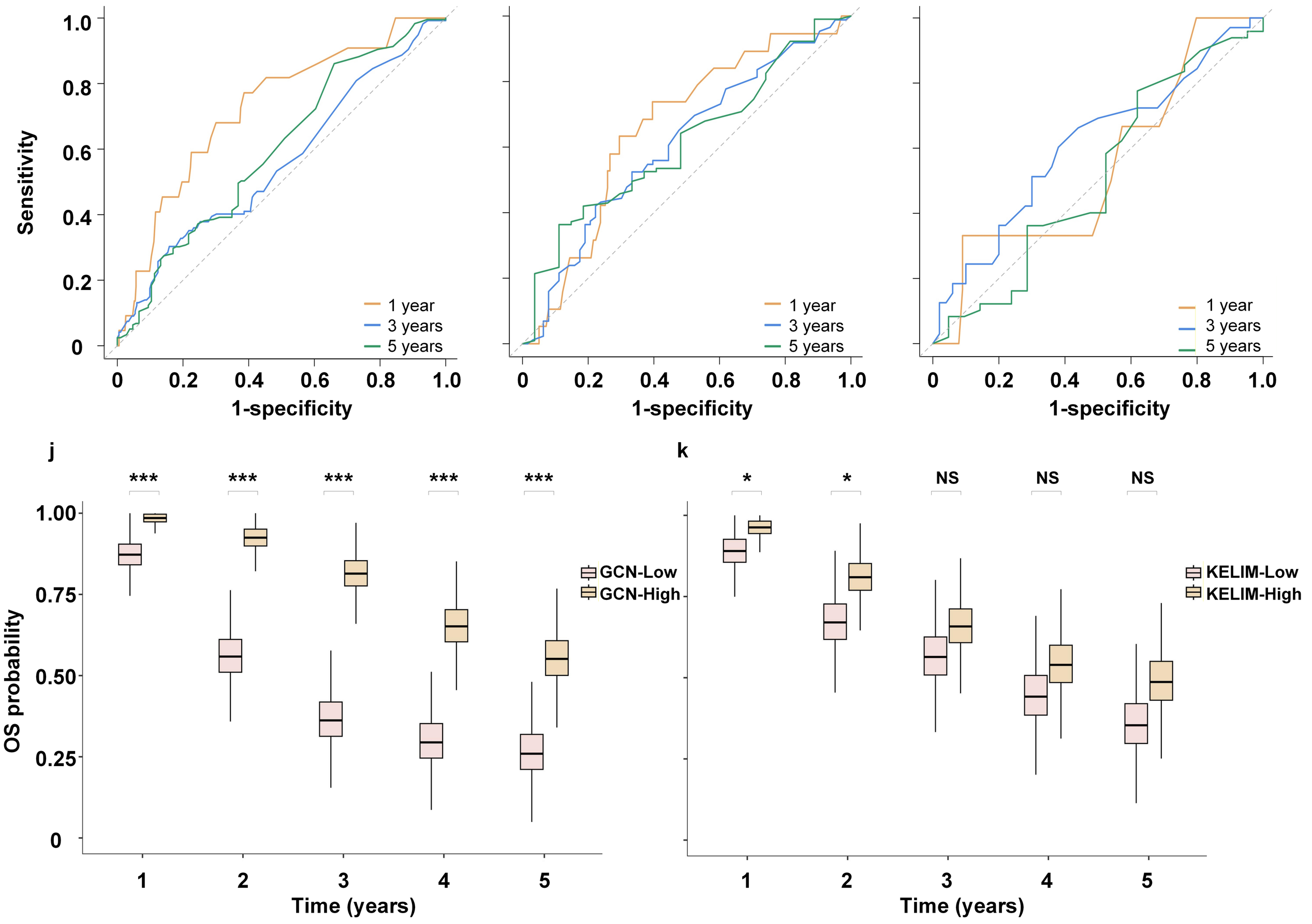
KELIM-External test set 1

i

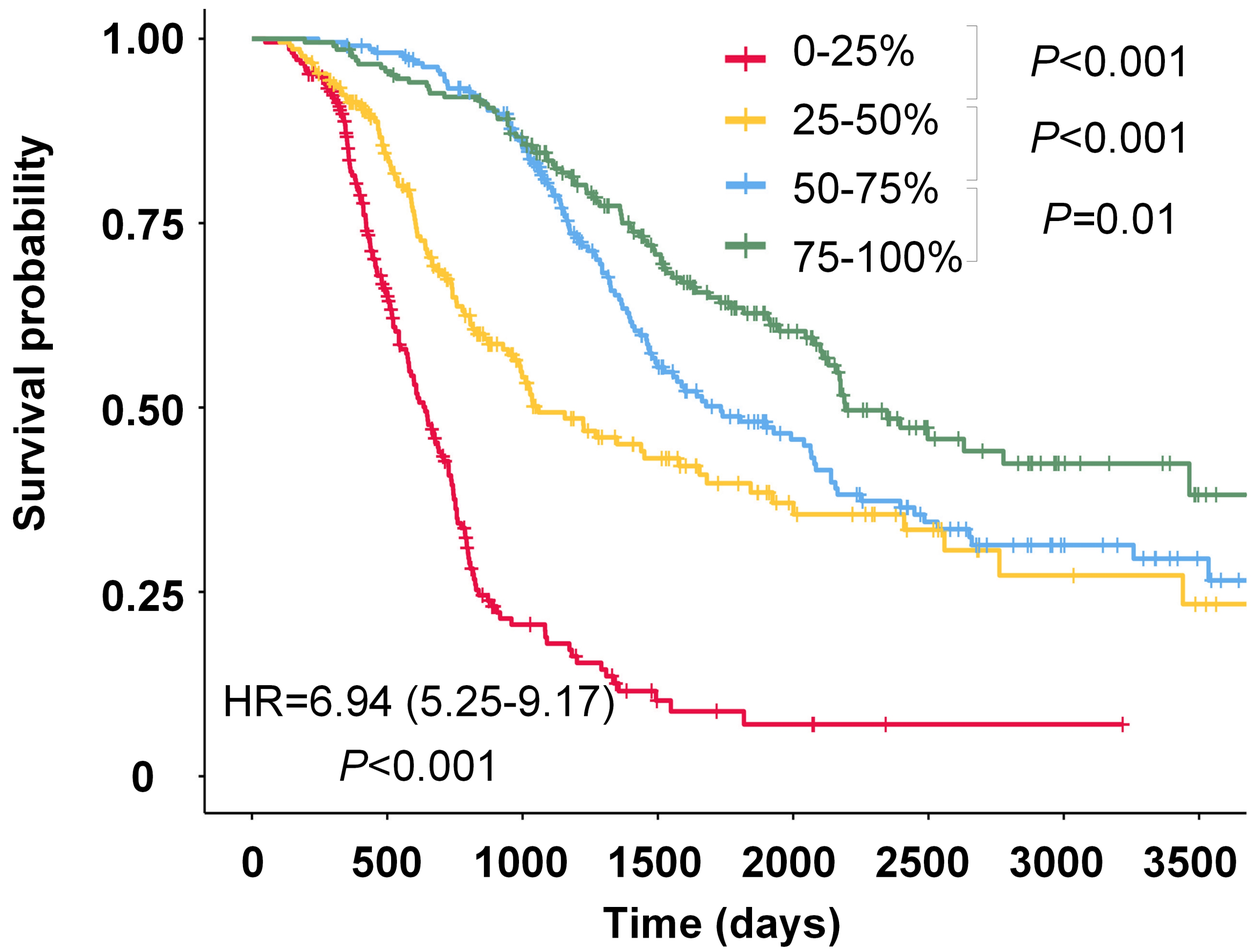
KELIM-External test set 2

j

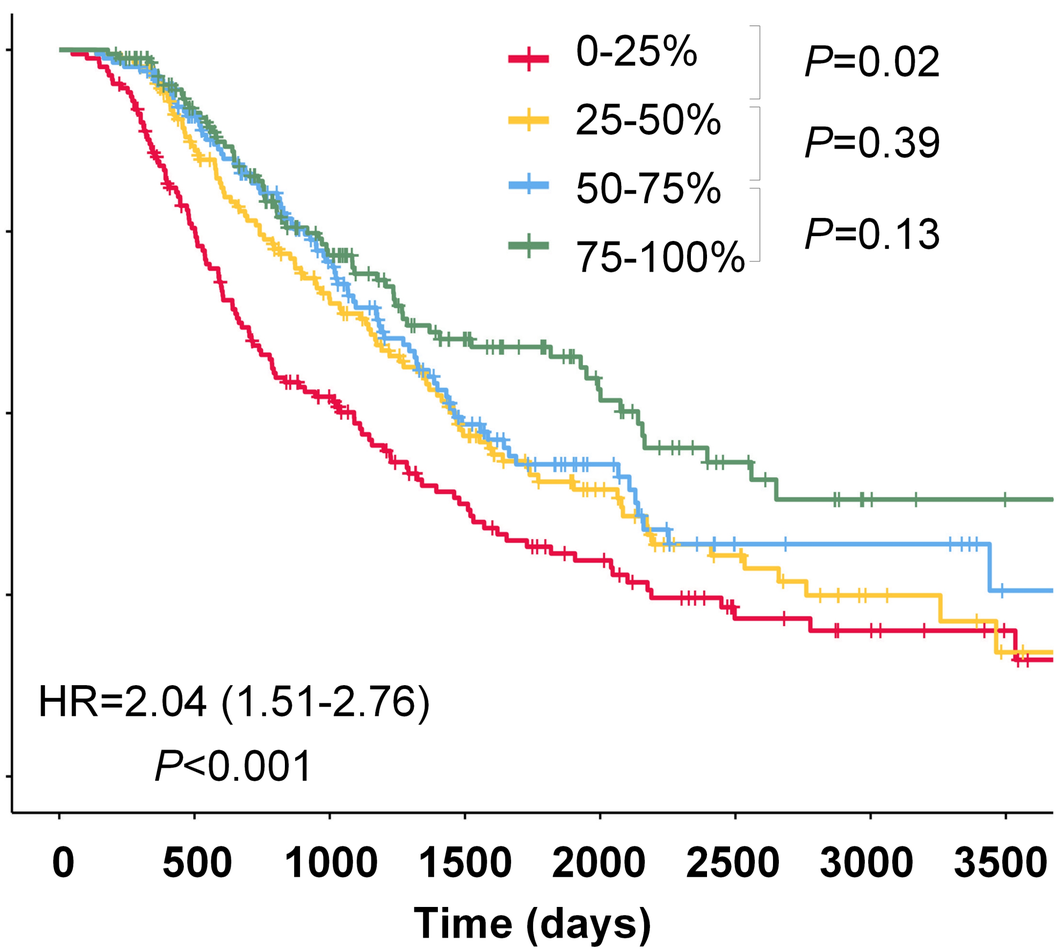
k



a



b



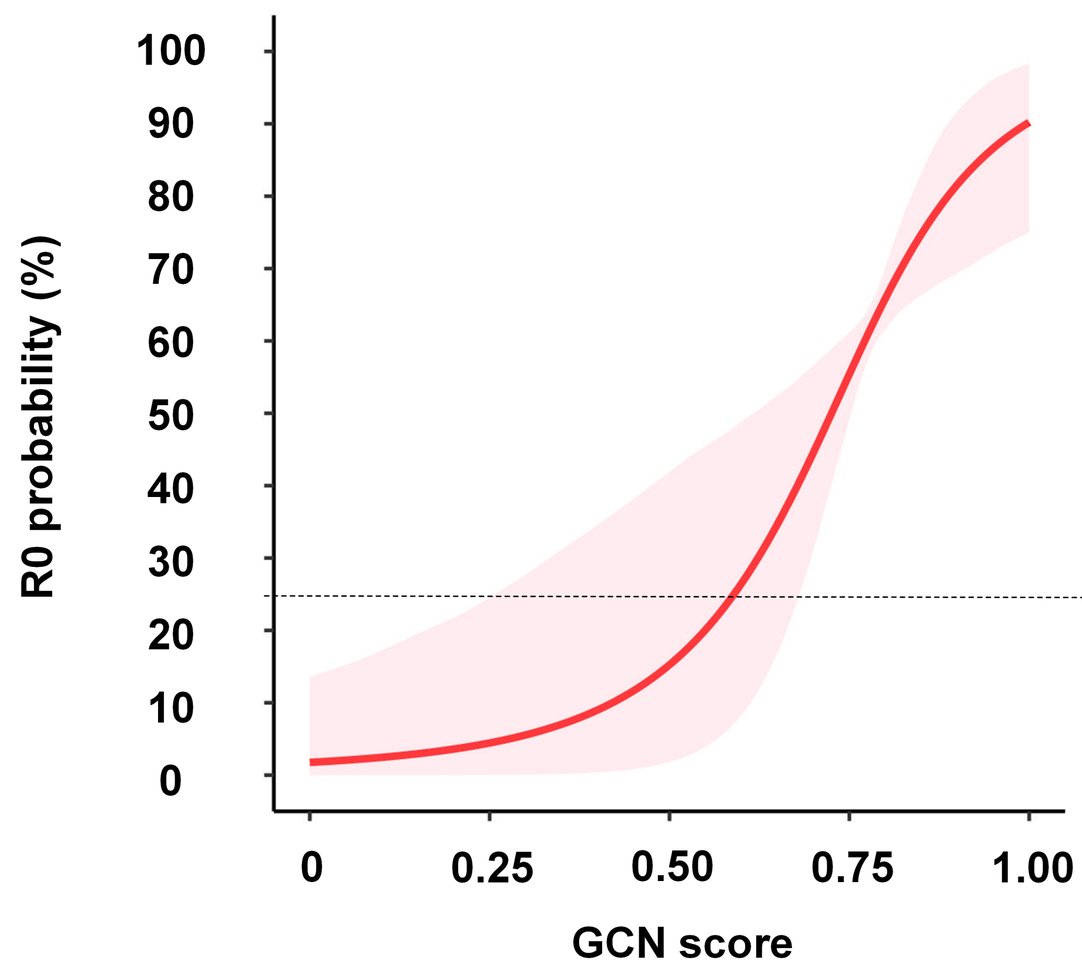
Number at risk

0-25%	210	114	25	7	4	1	1	0
25-50%	212	153	71	45	24	15	8	5
50-75%	213	206	166	90	55	35	20	10
75-100%	203	193	169	113	70	30	15	6

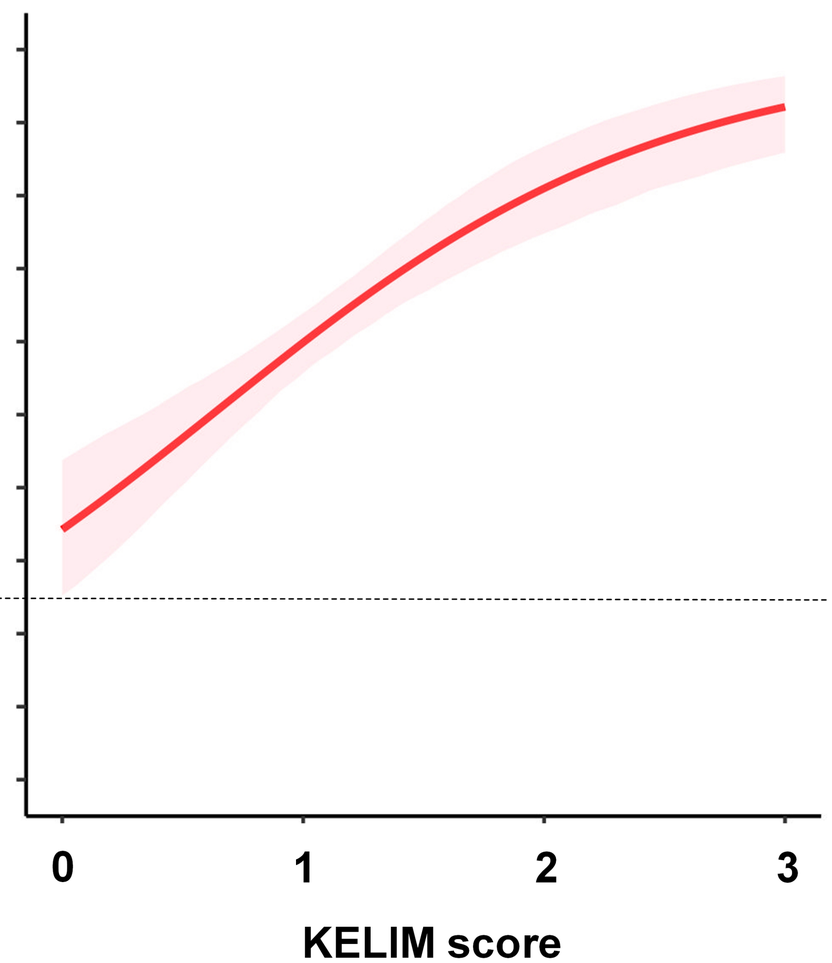
Number at risk

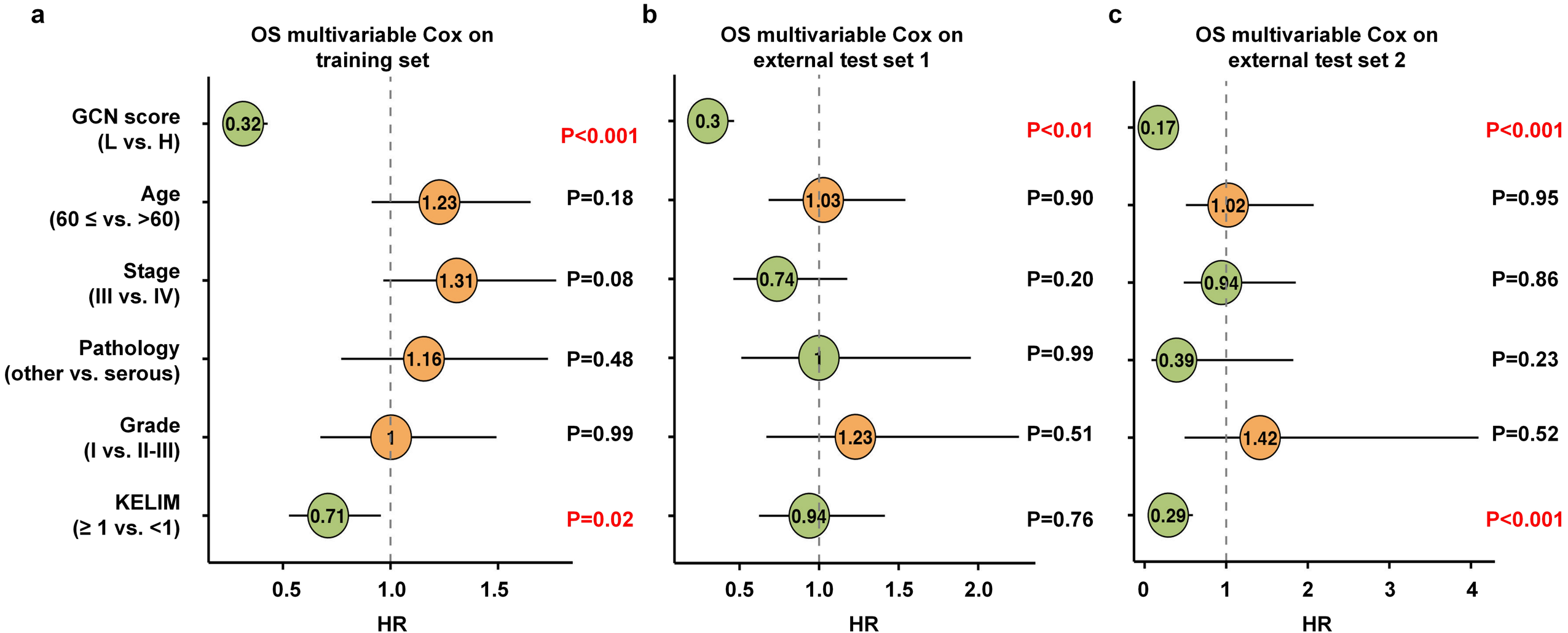
0-25%	171	122	75	45	31	14	10	5
25-50%	173	139	94	58	34	19	8	3
50-75%	171	141	93	49	26	10	9	3
75-100%	174	140	94	59	35	19	9	6

c



d




a
OS multivariable Cox on training set
GCN score (L vs. H)
Age (60 ≤ vs. >60)
Stage (III vs. IV)
Pathology (other vs. serous)
Grade (I vs. II-III)
KELIM (≥ 1 vs. <1)
0.32
1.23
1.31
1.16
1
0.71
P<0.001
P=0.18
P=0.08
P=0.48
P=0.99
P=0.02
0.5
1.0
1.5
HR
b
OS multivariable Cox on external test set 1
0.3
1.03
0.74
1
1.23
0.94
P<0.01
P=0.90
P=0.20
P=0.99
P=0.51
P=0.76
0.5
1.0
1.5
2.0
HR
c
OS multivariable Cox on external test set 2
0.17
1.02
0.94
0.39
1.42
0.29
P<0.001
P=0.95
P=0.86
P=0.23
P=0.52
P<0.001
0
1
2
3
4
HR