npj | digital medicine

Article

Published in partnership with Seoul National University Bundang Hospital

https://doi.org/10.1038/s41746-025-01712-0

Multimodal fusion model for prognostic prediction and radiotherapy response assessment in head and neck squamous cell carcinoma

Ruxian Tian[1,15], Feng Hou[2,15], Haicheng Zhang[3,15], Guohua Yu[4,15], Ping Yang[4,15], Jiaxuan Li[1], Ting Yuan[1], Xi Chen[1], Ying Chen[5], Yan Hao[6], Yisong Yao[1], Hongfei Zhao[1], Pengyi Yu[1], Han Fang[1], Liling Song[5], Anning Li[7], Zhonglu Liu[1], Huaiqing Lv[8] ✉, Dexin Yu[7] ✉, Hongxia Cheng[9] ✉, Ning Mao[10,11] ✉ & Xicheng Song[1,12,13,14] ✉

Accurate prediction of prognosis and postoperative radiotherapy response is critical for personalized treatment in head and neck squamous cell carcinoma (HNSCC). We developed a multimodal deep learning model (MDLM) integrating computed tomography, whole-slide images, and clinical features from 1087 HNSCC patients across multiple centers. The MDLM exhibited good performance in predicting overall survival (OS) and disease-free survival in external test cohorts. Additionally, the MDLM outperformed unimodal models. Patients with a high-risk score who underwent postoperative radiotherapy exhibited prolonged OS compared to those who did not ($P = 0.016$), whereas no significant improvement in OS was observed among patients with a low-risk score ($P = 0.898$). Biological exploration indicated that the model may be related to changes in the cytochrome P450 metabolic pathway, tumor microenvironment, and myeloid-derived cell subpopulations. Overall, the MDLM effectively predicts prognosis and postoperative radiotherapy response, offering a promising tool for personalized HNSCC therapy.

Head and neck squamous cell carcinoma (HNSCC) is the sixth most common cancer worldwide[1,2]. Alarmingly, the incidence of HNSCC is projected to increase by 30%, equating to approximately 1.08 million new cases by 2030[1,2]. Despite advancements in multiple therapeutic modalities for HNSCC over the past decades, no significant improvement in survival outcomes has been observed[3,4]. Although postoperative radiotherapy can reduce recurrence and improve prognosis to a certain extent, not all patients can benefit from radiotherapy, and ineffective radiotherapy can even increase the risk of death[5]. Therefore, to develop personalized strategies, identifying patients with different risk outcome and who can benefit from radiotherapy is crucial.

Traditionally, a variety of malignancies, including HNSCC, are primarily risk-stratified based on the TNM staging system to assist in treatment planning[6]. However, judging TNM staging according to the American Joint Committee on Cancer (AJCC) staging criteria is subject to significant subjectivity owing to substantial inter-observer and intra-observer variability, as well as intratumoral heterogeneity[7,8]. Consequently, notable disparities still exist in prognostic outcomes, even among patients with identical TNM stages. Currently, enhancing the capacity for prognostic risk stratification of patients with HNSCC remains an ongoing challenge for clinicians.

Recent studies have focused on developing genomic prediction models through bioinformatics analysis[9–11]. However, these studies often rely on public databases to establish predictive signatures, and the lack of independent external validation undermines their reliability. Moreover, genomic examinations are time-consuming, labor-intensive, and costly, limiting their applicability for large-scale clinical applications. Radiomics is a rapidly evolving field that can identify associations with clinical outcomes by capturing tissue and lesion characteristics, such as heterogeneity and shape from medical images[12]. Computed tomography (CT) is a routine diagnostic tool for patients with HNSCC, providing valuable information about tumor characteristics and surrounding structures in a cost-effective and time-

A full list of affiliations appears at the end of the paper. ✉e-mail: lvhuaiqing@126.com; yudexin0330@sina.com; hongxiacheng322@hotmail.com; maoning@pku.edu.cn; drxchsong@163.com

efficient manner. The CT-based radiomics model has demonstrated promising outcomes in prognostic prediction for colon, lung, and gastric cancers, and other malignant tumors[13–16] Currently, researchers have initiated exploring the application value of utilizing machine learning to establish CT-based radiomics prognostic models for HNSCC[17,18]. However, the lack of external validation leads to the poor generalization of these models. Additionally, relying solely on predefined image features to build models restricts the amount of information that can be extracted from images, which may compromise the predictive performance.

The deep learning (DL) technique can automatically learn representative information from medical image and has shown great potential in clinical applications[19]. Studies on DL for predicting gene mutation as well as histological differentiation grade and lymph node metastasis of HNSCC provides a valuable reference for exploring its application in prognosis[20–22]. However, relying solely on unimodal images limits the predictive performance of DL model. Emerging evidence suggests that the integration of multimodal data has the potential to complement tumor heterogeneity at various scales and enhance the predictive capability of existing models[23,24]. Researchers have observed that histopathological whole-slide imaging (WSI) can provide information reflecting the molecular characteristics or genetic patterns of tumors. Previous studies have shown that WSI-based DL can effectively predict prognosis and treatment response[25,26]. A recent study reported that a radiopathomics model, integrating magnetic resonance (MR) and WSI, was superior to unimodal models in predicting the response of rectal cancer to neoadjuvant therapy[27]. However, the predictive value of radiopathology in the prognosis and radiotherapy response of patients with HNSCC remains unclear. Furthermore, DL operates as a black box, and its underlying mechanisms remain unclear. In recent years, researchers have explored the biological mechanisms underlying DL models through bulk RNA sequencing[28,29]. However, because bulk RNA-seq is based on sequencing entire tissues or cell populations, it often fails to capture specific transcriptome information of single cells. Single-cell RNA-sequencing (scRNA-seq) can overcome these limitations by measuring the whole transcriptome at single-cell resolution and distinguishing different cell types in tumor tissue[30]. Additionally, Feng et al. demonstrated that the combination of scRNA-seq and bulk RNA-seq can offer a more comprehensive and in-depth understanding of the underlying mechanisms of DL models [31].

In this study, we developed and validated a multimodal DL model (MDLM) based on contrast-enhanced CT, WSI, and clinical features to predict the prognosis and postoperative radiotherapy response in patients with HNSCC. Additionally, we explored the underlying biological mechanisms of MDLM using scRNA-seq and bulk RNA-seq analyses to enhance the interpretability of the model.

Results

Baseline characteristics

A total of 1087 eligible patients were included in this study. Supplementary Fig. 1 provides detailed information on the patient enrollment process and cohort design. To enhance the adaptability of our model to patients of diverse origins and improve its generalizability, we amalgamated 551 patients from Yantai Yuhuangding Hospital (227 patients) and the Affiliated Hospital of Qingdao University (324 patients), from January 1, 2014, to December 31, 2019. Subsequently, these patients were randomly allocated to either a training cohort (386 patients; median age, 64 years; range, 37–91 years) or an internal test cohort (165 patients; median age, 62 years; range, 40–81 years) in a ratio of 7:3. External test cohort 1 (219 patients; median age, 63 years; range, 29–78 years) was recruited from Shandong University Qilu Hospital between January 1, 2016, and December 31, 2019. External test cohort 2 (101 patients; median age, 63 years; range, 42–83 years) was recruited from Shandong Provincial Hospital between January 1, 2016, and December 31, 2019. External test cohort 3 (160 patients; median age, 63.5 years; range, 41–83 years) was recruited from Linyi People's Hospital between January 1, 2016, and December 31, 2019. Additionally, 56 patients with HNSCC (median age, 61.5 years; range, 24–83 years) were collected from the Cancer Imaging Archive (TCIA, https://www.cancerimagingarchive.net/) to further assess the predictive performance across different ethnic populations. Notably, the TCIA cohort exhibited no disease-free survival (DFS). Male predominance was observed across all cohorts (97.15% for the training cohort, 98.18% for the internal test cohort, 97.72% for external test cohort 1, 93.07% for external test cohort 2, 95.62% for external test cohort 3, and 62.50% for the TCIA cohort). The distributions of sex, smoking history, drinking history, pathological differentiation, T stage, and N stage were significantly imbalanced across the different cohorts (all $P < 0.001$), which was attributed to the inclusion of the TCIA cohort. A detailed presentation of the patients' clinical features is provided in Supplementary Table 1. Supplementary Fig. 2 shows that there are significant differences in the overall survival (OS) or DFS among the different cohorts (all $P < 0.001$).

Performance of MDLM

Cox hazard proportional regression was used to construct the MDLM by integrating the CT-based DL signatures, WSI-based DL signatures, and clinical features. The workflow of this study is illustrated in Fig. 1. Supplementary Table 2 shows the detailed results of the univariate Cox regression analysis. Multivariate Cox regression showed that the CT risk score, WSI risk score, age, and pathological differentiation status were independent risk factors for OS, whereas only the CT risk score, WSI risk score and N stage were independent risk factors for DFS (Figs. 2a and 2b). The proposed MDLM exhibited good performance in predicting OS and DFS in the internal test cohort, with a C-indices of 0.745 (0.713–0.789) and 0.741 (0.708–0.781), respectively. The receiver operating characteristic (ROC) curves showed that MDLM had high accuracy in predicting 1-year, 2-year, and 3-year OS in the internal test cohort, with area under the curve (AUC) values of 0.778 (0.716–0.832), 0.770 (0.727–0.831), and 0.793 (0.737–0.863), respectively (Fig. 3b). Furthermore, the model could accurately predict the 1-year, 2-year, and 3-year DFS in the internal test cohort, with AUC values of 0.788 (0.722–0.857), 0.795 (0.721–0.855), and 0.760 (0.697–0.846), respectively (Fig. 3h).

Next, we verified the predicted performance of the MDLM in the external test and TCIA cohorts. In the external test cohort 1, the MDLM showed strong robustness in predicting OS and DFS, with C-indices of 0.717(0.669–0.758) and 0.674(0.628–0.732), respectively. In external test cohort 2, the model also showed good performance in predicting OS and DFS, with C-indices of 0.713 (0.676–0.772) and 0.721 (0.677–0.786), respectively. In external test cohort 3, the C-indices of the model for predicting OS and DFS were 0.682 (0.638–0.729) and 0.747 (0.707–0.786), respectively. In the TCIA cohort, the C-index of the model for predicting OS decreased slightly to 0.636 (0.596–0.677). In addition, the ROC curve further demonstrates the good predictive performance of the MDLM. The respective AUC values for predicting 1-year, 2-year, and 3-year OS were 0.644 (0.592–0.689), 0.739 (0.688–0.775), 0.742 (0.687–0.796) in external test cohort 1, 0.724 (0.679–0.782), 0.741 (0.683–0.812), 0.750 (0.688–0.825) in external test cohort 2, 0.841 (0.799–0.882), 0.719 (0.663–0.762), 0.728 (0.678–0.785) in external test cohort 3, and 0.552 (0.511–0.606), 0.695 (0.638–0.743), 0.679 (0.626–0.751) in TCIA cohort (Fig. 3c–f). Similarly, the model exhibited high accuracy in predicting DFS. The respective AUC values for predicting 1-year, 2-year, and 3-year DFS were 0.597 (0.566–0.649), 0.695 (0.638–0.741), 0.699 (0.659–0.743) in external test cohort 1, 0.706 (0.656–0.749), 0.792 (0.738–0.851), 0.709 (0.659–0.768) in external test cohort 2, and 0.701 (0.649–0.762), 0.784 (0.723–0.832), 0.779 (0.728–0.825) in external test cohort 3 (Fig. 3i–k).

In addition, we divided patients into low- and high-risk groups based on the median MDLM risk score in the training cohort as a cutoff value and applied the same cutoff value to the test cohorts. In the internal test cohort, the Kaplan–Meier curve demonstrated that patients in the low-risk group exhibited prolonged OS and DFS compared to those in the high-risk group (all $P < 0.0001$; Fig. 4b, h). Furthermore, we observed significant prognostic differences between different risk groups in external test cohort 1, external test cohort 2, external test cohort 3, and the TCIA cohort (all $P < 0.05$; Figs. 4c–f and 4i–3k). The distribution of clinical features differed among patients

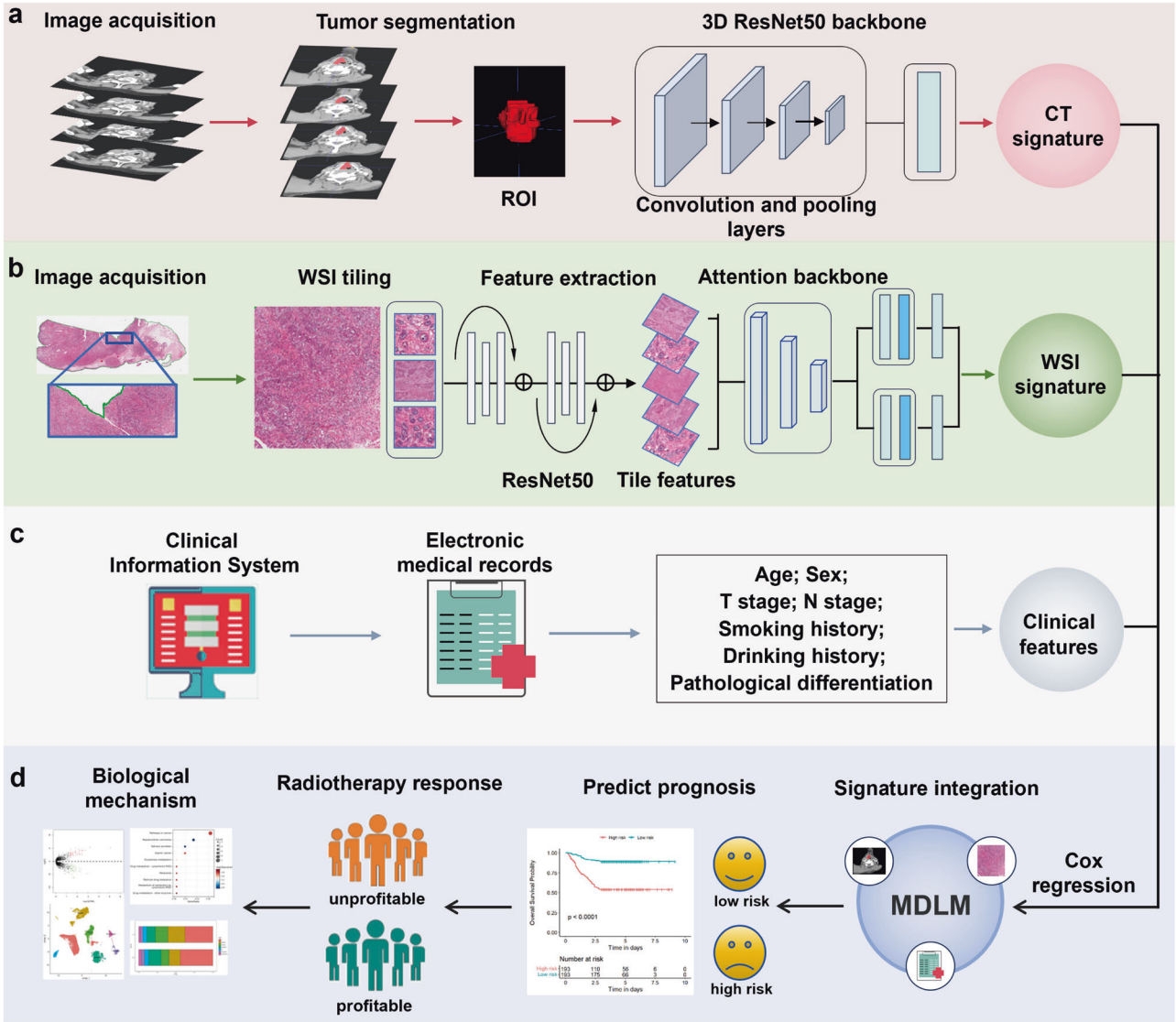

Fig. 1 | The overall workflow of MDLM. a Three-dimensional ResNet was utilized to establish CT signature for predicting prognosis. The ROI was input into the 3D ResNet50 model, where features were extracted through multiple 3D convolutional and pooling layers. These features were then condensed into a fixed-length vector via fully connected layers, forming the CT signature—a quantitative representation of the tumor phenotype. **b** The attention-based multiple instance learning algorithm was used to establish WSI signature for predicting prognosis. WSI were divided into tiles, encoded into 1024-dimensional feature vectors using pretrained ResNet50, then aggregated into a 512-dimensional slide-level representation through attention-weighted pooling for survival prediction. **c** Clinical pathological information. **d** MDLM was established using Cox hazard proportional regression to integrate CT signature, WSI signature, and clinical features to predict prognosis and radiotherapy response. Single cell RNA-sequencing and bulk RNA-sequencing analysis were employed to investigate the underlying biological mechanisms of MDLM. MDLM multimodal deep learning model, CT computed tomography, WSI whole slide images, ROI region of interest.

in the different risk groups, as shown in Supplementary Table 3 and Supplementary Table 4. Furthermore, we investigated the risk stratification capability based solely on either CT or WSI risk scores. In the training cohort, both the CT risk score and WSI risk score successfully stratified patients into low- and high-risk groups with significant prognostic differences when using median cutoffs (all $P < 0.05$). However, in external test cohort 1 and TCIA cohort, the CT risk score failed to stratify patients into risk groups with significantly different OS (Supplementary Fig. 3). Similarly, in the internal test cohort, external test cohort 1, and external test cohort 2, the CT risk score could not differentiate patients into groups with significantly different DFS (Supplementary Fig. 3). Furthermore, in external test cohort 2 and TCIA cohort, the WSI risk score was unable to stratify patients into risk groups with significantly different OS (Supplementary Fig. 4). Additionally, in internal test cohort, external test cohort 1, and external test cohort 3, the WSI risk score failed to distinguish patients with significantly different DFS (Supplementary Fig. 4). These results suggest that the MDLM exhibits superior risk stratification capability compared to unimodal prediction models.

Comparison with unimodal model

We compared the MDLM with the unimodal models for all test sets. The MDLM outperformed the unimodal models in predicting OS and DFS in all cohorts (Supplementary Fig. 5). In the internal test cohort, the C-index of the MDLM (0.745 [0.713–0.789]) for the prediction of OS was higher than those of the CT-based DL signature (0.700 [0.657–0.754]), WSI-based DL signature (0.653 [0.610–0.692]), and clinical signature (0.659 [0.607–0.711]). Consistent results were also observed in external test cohort 1, external test cohort 2, external test cohort 3, and the TCIA cohort (Table 1). In addition, the C-index of the MDLM (0.741[0.708–0.781]) in predicting DFS was higher than that of the CT-based DL signature (0.676 [0.637–0.712]), WSI-based DL signature (0.607 [0.572–0.646]), and clinical signature (0.627 [0.563–0.673]) in the internal test cohort. This finding was

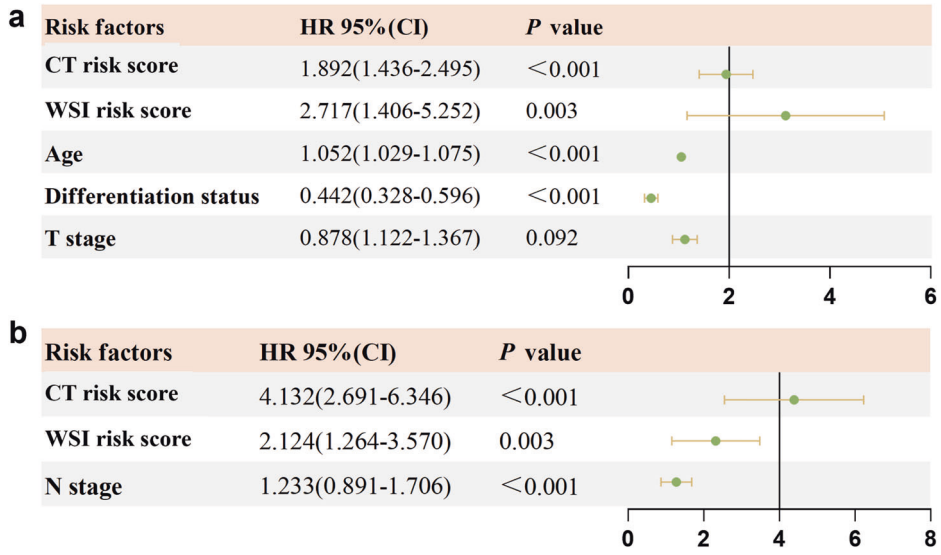

Fig. 2 | Forest plot displays results of multivariate Cox regression analysis. a Independent risk factors associated with overall survival, and with (**b**) disease-free survival. HR hazard ratio, CI confidence interval, CT computed tomography, WSI whole slide images.

further validated in external test cohort 1, external test cohort 2, and external test cohort 3 (Table 1). Furthermore, a gene-based prediction model constructed in the Yuhuangding Hospital cohort demonstrated good performance (C-index = 0.895 [0.859–0.938]) for OS prediction during training but exhibited poor generalizability in the TCIA cohort (C-index = 0.481 [0.355–0.606]), indicating that its cross-cohort applicability was inferior to that of the MDLM.

Prognostic value of the MDLM in subgroups

We also evaluated the prognostic capacity of MDLM in subgroups defined by various clinical features. In the internal test cohort, the MDLM could distinguish patients in the same T stage, N stage, or pathological differentiation (moderate and well) subgroups into low- and high-risk groups with significantly different OS (all $P < 0.05$; Fig. 5a). Regarding DFS, only the T stage, N0 stage, and pathological differentiation (moderate and well) subgroups showed prognostic differences (all $P < 0.05$; Fig. 5b). In the external test cohort 1, significant differences in OS were observed in different risk groups within the T1-T2 stage and N0 stage subgroups ($P < 0.01$), whereas no significant differences were observed among the other subgroups (all $P > 0.05$, Supplementary Fig. 6). In addition, subgroup analysis showed no significant differences in DFS between the different risk groups in the external test cohort 1 (all $P > 0.05$; Supplementary Fig. 7). In the external test cohort 2, only the N0 stage subgroup showed OS differences ($P < 0.01$, Supplementary Fig. 8). Regarding DFS, subgroup analysis showed no significant differences between the different risk groups in the external test cohort 2 (all $P > 0.05$; Supplementary Fig. 9). In external test cohort 3, significant difference in OS were observed in different risk groups within the T1-T2 stage, N0 stage, and moderately pathological differentiation subgroups (all $P < 0.01$, Supplementary Fig. 10). Regarding DFS, T3-T4 stage and pathological differentiation (moderate and well) subgroups showed prognostic differences (all $P < 0.05$, Supplementary Fig. 11). These results demonstrate the complementary value of the MDLM to traditional clinical prognostic features. The inconsistent results of the subgroup analysis between the internal and external test cohorts may be attributed to the limited sample size of the external test cohorts. Additionally, in the external test cohorts, subgroups could not be analyzed appropriately based on the pathological differentiation status owing to insufficient sample size.

Predict postoperative radiotherapy response

In addition to assessing the MDLM to predict prognosis, we further evaluated the model's ability to predict the benefits of postoperative radiotherapy. First, 1:1 propensity score matching (PSM) was initially conducted to ensure that patient characteristics were comparable between those who underwent postoperative radiotherapy and those who did not (Supplementary Table 5 and Supplementary Table 6). After matching, we compared the outcomes of the patients in each risk group who received postoperative radiotherapy with those who did not. In the high-risk group, postoperative radiotherapy was associated with improved OS ($P = 0.016$), whereas no significant improvement in DFS was observed ($P = 0.712$; Fig. 6a). In contrast, postoperative radiotherapy did not yield any significant improvement in OS ($P = 0.898$) or DFS ($P = 0.732$; Fig. 6b) for patients classified as low-risk. Additionally, the above analysis was conducted in all patients without PSM, and similar results were obtained (Supplementary Fig. 12).

Furthermore, patients who received postoperative radiotherapy were divided into radiotherapy-sensitive or radiotherapy-resistant groups based on whether they benefited from radiotherapy (OS ≥ or < 3 years), and compared the MDLM risk score of the two groups. Among patients receiving postoperative radiotherapy, the MDLM risk scores of radiotherapy-sensitive patients were significantly higher than those of radiotherapy-resistant patients ($P < 0.001$; Fig. 6c). Additionally, the percentage of radiotherapy-sensitive patients in the high-risk group was significantly higher than that in the low-risk group ($P = 0.001$; Fig. 6d).

Model visualization

To visualize the model prediction process and enhance interpretability, we performed a visual analysis of the CT and WSI features derived from the DL algorithm. As shown in Supplementary Fig. 13a and 13b, for high-risk patients, the WSI tiles primarily focused on by DL algorithm were tumor parenchyma regions with significant nuclear atypia. For low-risk patients, the WSI tiles that the DL algorithm focuses on included normal squamous epithelium, tumor tissue without obvious nuclear atypia, muscle tissue, lymphocytes, and red blood cells. Additionally, we conducted SHapley Additive exPlanations (SHAP) analysis on the CT features extracted by the DL algorithm to calculate the contribution of each CT feature to the model output. The SHAP plot displayed the top 10 CT features that significantly contributed to the model output (Supplementary Fig. 13c).

Biological mechanism

To explore the potential biological basis of MDLM, we performed bulk RNA-seq and scRNA-seq analyses. Bulk RNA-seq analysis was conducted on 50 patients (21 in the low-risk group and 29 in the high-risk group) using RNA-seq data from the TCIA cohort. A total of 82 differentially expressed genes (DEGs) were identified (47 upregulated and 35 downregulated) between the low- and high-risk groups (Fig. 7a). Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis indicated that the DEGs were enriched in the metabolism of xenobiotics by cytochrome P450, drug

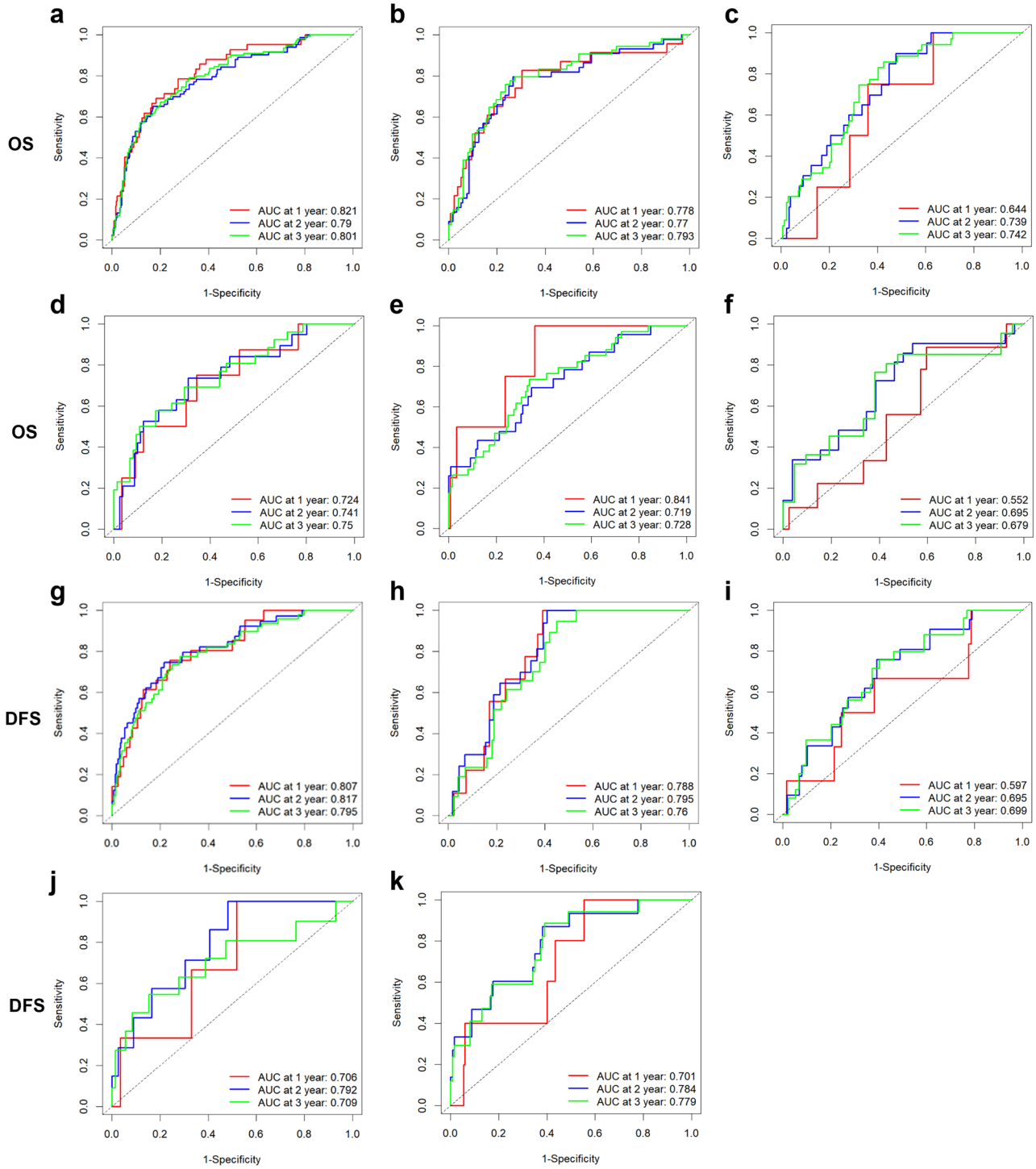

Fig. 3 | Performance of MDLM for predicting prognosis. ROC curves illustrate the predictive performance of MDLM for OS in (**a**) training cohort, (**b**) internal test cohort, (**c**) external test cohort 1, (**d**) external test cohort 2, (**e**) external test cohort 3, and (**f**) TCIA cohort. ROC curves illustrate the predictive performance of MDLM for DFS in (**g**) training cohort, (**h**) internal test cohort, (**i**) external test cohort 1, (**j**) external test cohort 2, and (**k**) external test cohort 3. AUC areas under the receiver operating characteristic curve, OS overall survival, DFS disease-free survival, TCIA The Cancer Imaging Archive, MDLM multimodal deep learning model, ROC receiver operating characteristic.

metabolism–cytochrome P450, and pathways in cancer metabolism (Fig. 7b). Gene ontology (GO) analysis showed that DEGs were primarily enriched in cellular components, such as the extracellular region, extracellular space, extracellular exosome, extracellular vesicle, and extracellular organelle (Fig. 7c). Additionally, we selected 128 patients (63 in the low-risk group and 65 in the high-risk group) with available RNA-seq data from the Yuhuangding Hospital cohort and repeated the above analysis. In this cohort, we identified 23 DEGs (10 upregulated and 13 downregulated). Notably, the functional enrichment analysis results differed from those obtained in the TCIA cohort, which may be attributed to differences in ethnic composition between the two cohorts (Supplementary Fig. 14).

Tumor samples were obtained from 10 patients with HNSCC for scRNA-seq analysis, of which six were predicted by MDLM as low-risk and four as high-risk. After the quality control assessment, we acquired the

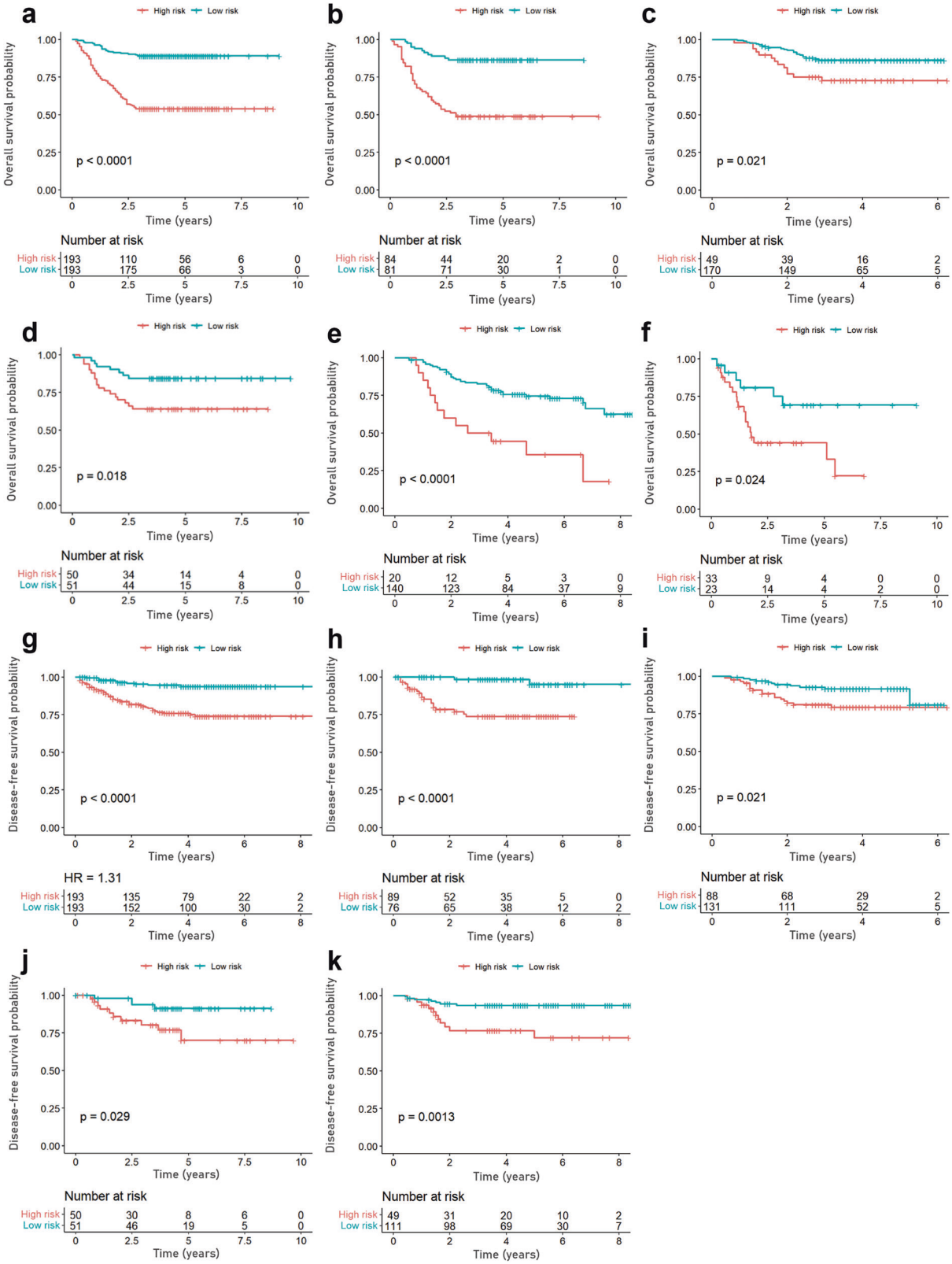

Fig. 4 | Kaplan–Meier analysis based on the risk score of MDLM. Comparison of overall survival between the low- and high-risk groups in (**a**) training cohort, (**b**) internal test cohort, (**c**) external test cohort 1, (**d**) external test cohort 2, (**e**) external test cohort 3, and (**f**) TCIA cohort. Comparison of disease-free survival between low- and high-risk groups in (**g**) training cohort, (**h**) internal test cohort, (**i**) external test cohort 1, (**j**) external test cohort 2, and (**k**) external test cohort 3. TCIA The Cancer Imaging Archive, MDLM multimodal deep learning model.

Table 1 | The C-index of OS and DFS predicted by various models

	Predictive models	Training cohort	Internal test cohort	External test cohort 1	External test cohort 2	External test cohort 3	TCIA cohort[a]
OS	MDLM	0.767 (0.726–0.802)	0.745 (0.713–0.789)	0.717 (0.669–0.758)	0.713 (0.676–0.772)	0.682(0.638–0.729)	0.636 (0.596–0.677)
	CT-based DL signature	0.678 (0.648–0.729)	0.700 (0.657–0.754)	0.644(0.592–0.677)	0.641 (0.608–0.689)	0.669(0.622–0.715)	0.620 (0.592–0.683)
	WSI-based DL signature	0.631 (0.610–0.667)	0.653 (0.610–0.692)	0.642(0.585–0.706)	0.583(0.550–0.627)	0.623(0.588–0.660)	0.616 (0.572–0.656)
	Clinical signature	0.723 (0.594–0.678)	0.659 (0.607–0.711)	0.646(0.609–0.688)	0.622 (0.578–0.688)	0.618(0.579–0.647)	0.493 (0.468–0.523)
DFS	MDLM	0.790 (0.742–0.845)	0.741 (0.708–0.781)	0.674(0.628–0.732)	0.721 (0.677–0.786)	0.747(0.707–0.786)	/
	CT-based DL signature	0.775 (0.736–0.815)	0.676 (0.637–0.712)	0.657(0.611–0.725)	0.681 (0.639–0.727)	0.719(0.673–0.758)	/
	WSI-based DL signature	0.611 (0.579–0.678)	0.607 (0.572–0.646)	0.603(0.576–0.663)	0.594 (0.561–0.638)	0.632(0.594–0.673)	/
	Clinical signature	0.609 (0.569–0.653)	0.627 (0.563–0.673)	0.617(0.582–0.656)	0.544 (0.513–0.579)	0.674(0.647–0.713)	/

MDLM multimodal deep learning model, *DL* deep learning, *CT* computed tomography, *WSI* whole slide imaging, *TCIA* the cancer imaging archive, *OS* overall survival, *DFS* disease-free survival.
[a]TCIA cohort exhibited no disease-free survival.

transcriptomes of 64,639 single cells (29,772 cells in the low-risk group, 34867 cells in the high-risk group) from 10 participants. Subsequently, 30 major cell clusters were identified based on t-distributed stochastic neighbor embedding (t-SNE) analysis (Supplementary Fig. 15). Seven distinct cell populations were identified based on the t-SNE analysis and canonical markers, including T cells, myeloid-derived cells (MDCs), plasma cells, epithelial cells, B cells, endothelial cells, and fibroblasts (Fig. 7d). As shown in Fig. 7e, these cell populations were distributed differently among the different risk groups. Further analysis showed that cell subpopulations with gene-specific expression in MDCs were unevenly distributed among the different risk groups (Fig. 7f and g). The proportion of macrophages with high expression of GPNMB was higher in the high-risk group than in the low-risk group. In contrast, the proportion of macrophages with high expression of FCN1 and mast cells with high expression of CPA3 was lower in the high-risk than in low-risk group. The detailed distribution of the MDCs subpopulations is shown in Fig. 7g.

Discussion

In this study, we established an MDLM that effectively predicts the prognosis and postoperative radiotherapy response of patients with HNSCC by incorporating CT, WSI, and clinical features. Our multimodal model is superior to the unimodal model and exhibits strong generalization and robustness. To our knowledge, this is the first study to develop and validate a DL-based multimodal model for predicting prognosis and radiotherapy response in patients with HNSCC. This was a multicenter study with the largest sample size to date in the field of HNSCC prognosis. Additionally, scRNA-seq and bulk RNA-seq analyses showed that poor prognosis predicted by MDLM may be related to changes in the cytochrome P450 metabolic pathway, tumor microenvironment, and MDCs subpopulation.

Considering the importance of predicting prognosis to guide the clinical treatment of HNSCC, many efforts have been dedicated to this topic. Abe et al. constructed a radiomics model based on conventional machine learning to predict OS using multi-ventricular CT but only achieved a C-index of 0.667[17]. The study conducted by Liu et al. demonstrated that the integration of clinicopathological characteristics with radiomics features of pre-treatment positron emission tomography (PET)/CT or post-treatment PET/CT improved the prediction of OS and DFS in patients with HNSCC[32]. However, these studies were constrained by small sample sizes and the absence of multicenter validation, thereby potentially resulting in overfitting and uncertain generalization. In addition, model performance is limited by traditional machine learning methods. One advantage of our study was the inclusion of a large sample of participants from multiple centers to improve the robustness and generalizability of the predictive model. In addition, we utilized DL, which is a branch of machine learning capable of automatically extracting features suitable for research tasks, in the modeling process.

When establishing the prognostic model, we observed that the performance of the CT-based signature was insufficient (C-index, < 0.70). Previous studies have indicated that the combination of multimodal data could further improve the model's prediction ability[23,24]. Andreas et al. reported that multimodal artificial intelligence-based pathogenomics improved survival prediction in oral squamous cell carcinoma[33]. Although the high cost of genetic testing makes this method unsuitable for widespread implementation, utilizing histopathology remains a promising approach. Unlike radiographic imaging, which can capture macrostructural information of the tumor, histopathology provides in-depth microstructural information about cellular properties and microenvironment characteristics within local lesions. Capturing the features of histopathological images using DL has been shown to be capable of predicting lesion types in HNSCC[34,35]. Our WSI-based signature for predicting HNSCC prognosis, although not satisfactory in terms of performance (C-index, <0.70), provided complementary information about the tumor microstructure related to prognosis. Recent studies have demonstrated that the integration of MR, CT, and histopathological images yields good performance in predicting prognosis and therapeutic response[27,36]. Moreover, the study by Boehm et al.

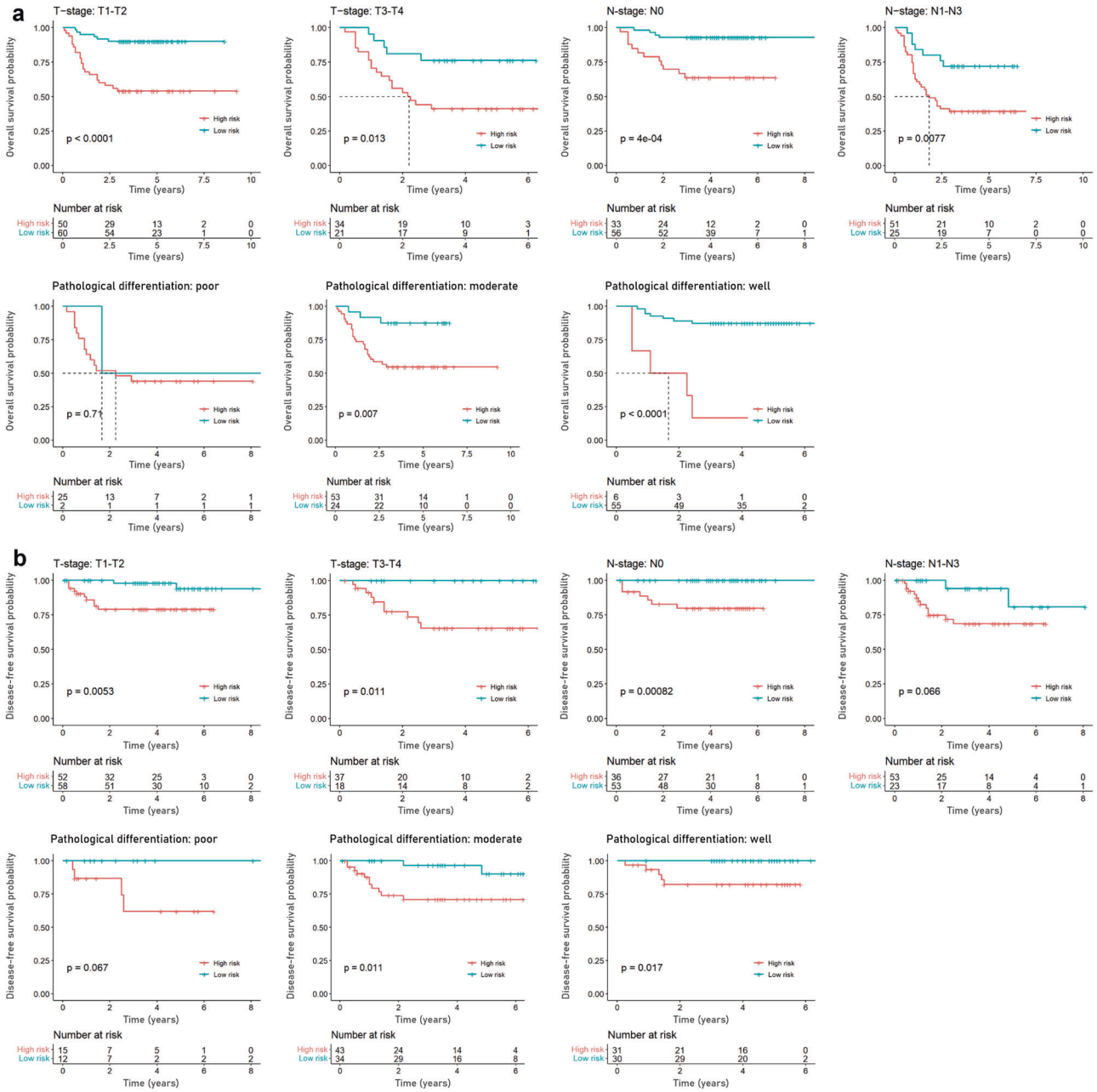

Fig. 5 | Kaplan–Meier survival analysis according to the risk score within each subgroup in the internal test cohort. a Overall survival. **b** Disease-free survival.

demonstrated that integrating radiological, histopathological, and genomic models could significantly improve risk stratification in high-grade serous ovarian cancer[37]. Similar to these studies, our MDLM showed better performance than other unimodal models, which was attributed to the integration of the macrostructural and microstructural characteristics of the tumor.

Despite the fact that approximately 75% of patients with HNSCC benefit from radiotherapy, there remains a subset of individuals who exhibit resistance to this treatment[5]. For these radiotherapy-resistant patients, radiotherapy is not efficacious and may cause severe adverse effects. Previous studies have used bioinformatics to establish a genetic signature for predicting radiotherapy response in HNSCC[38,39]. However, the clinical utility of prediction models derived from mining public RNA-seq databases is limited. In recent years, many researchers have tried to establish DL models based on medical images to predict treatment response in other tumors and have achieved good results[25,27,40]. For instance, Feng et al. demonstrated that a DL model based on MR and pathological images has high accuracy in predicting pathological response to neoadjuvant chemoradiotherapy in rectal cancer[27]. In our study, MDLM can also identify groups of patients with HNSCC who benefit from radiotherapy, which is of great clinical significance.

DL, as a black box, remains a stubborn conundrum to elucidate its underlying biological mechanisms. Recently, Feng et al. revealed the association between the radiomics model and tumor macrophage infiltration by analyzing of scRNA-seq and bulk RNA-seq data[31]. The combination of scRNA-seq and bulk RNA-seq can provide more comprehensive functional information, helping researchers gain a deeper understanding of cellular biological processes[30]. Bulk RNA-seq indicated that poor prognosis predicted by MDLM may be related to the tumor-related pathways such as cytochrome P450 and the cell microenvironment, both of which are closely related to tumor progression[41,42]. Notably, although the functional enrichment analysis results were inconsistent between the two cohorts, both were associated with prognosis. The observed discrepancies in enrichment outcomes may be attributed not only to limited sample sizes but also to

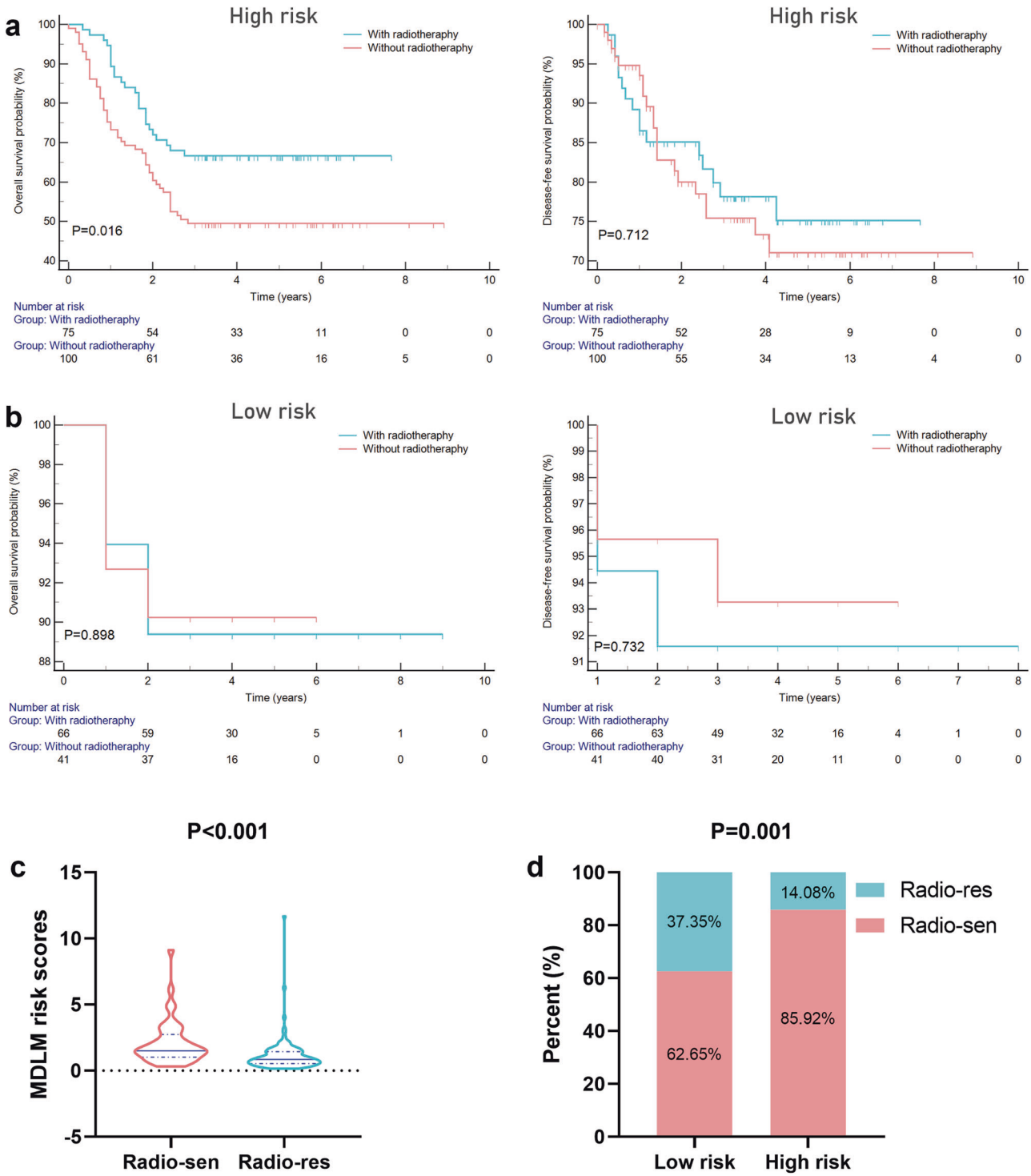

Fig. 6 | Relationship between MDLM score and benefit from postoperative radiotherapy in matched patients with HNSCC. Kaplan–Meier curves of overall survival and disease-free survival for patients stratified based on their receipt of postoperative radiotherapy in (**a**) high-risk and (**b**) low-risk groups. **c** Comparison of the MDLM score between radio-sen and radio-res patients. **d** Comparison of radiotherapy sensitivity between low-risk and high-risk patients. MDLM multimodal deep learning model, Radio-sen radio-sensitive, Radio-res radio-resistant, HNSCC head and neck squamous cell carcinoma.

potential molecular heterogeneity among HNSCC patients from different ethnic backgrounds[43]. The MDLM integrates molecular features (reflecting tumor microstructures) and morphological features (reflecting tumor macrostructures), which may partially compensate for genetic heterogeneity[23], thereby conferring superior generalizability compared to the genetic prediction model. ScRNA-seq suggested that the proportion of macrophages with high GPNMB expression was higher in the high-risk group, while the proportion of macrophages with high FCN1 expression and mast cells with high CPA3 expression was higher in the low-risk group. Studies have shown that GPNMB is linked to poor prognosis in malignant tumors, whereas FCN1 and CPA3 are associated with improved prognosis[44–46]. These findings are consistent with the results of our study.

Despite these promising findings, our study has some limitations. First, all tissue slides were obtained using the same type of digital scanner. Future

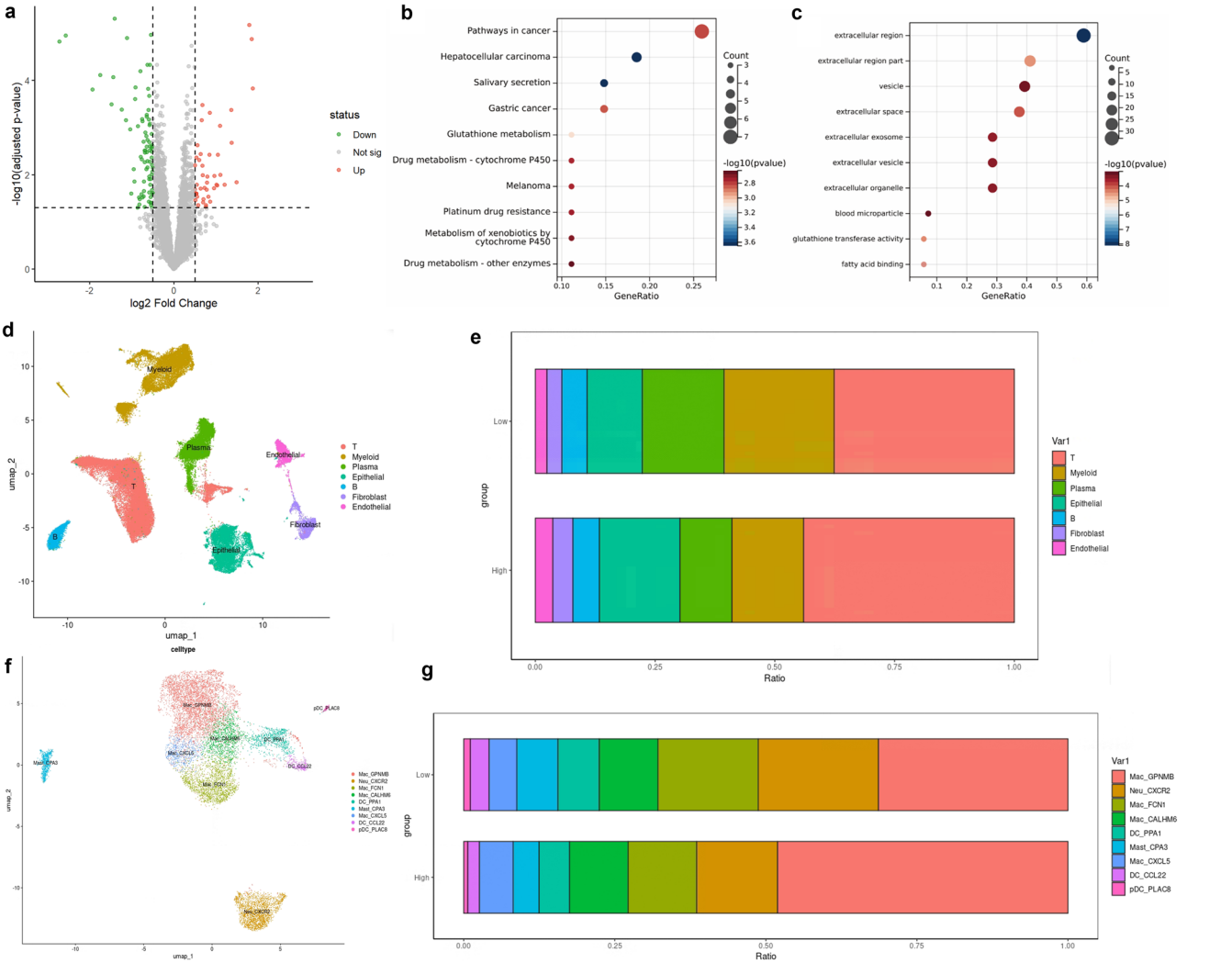

Fig. 7 | Bioinformatics analysis for exploring the underlying biological mechanism of the multimodal deep learning model. a Volcano pot of DEGs between low-risk and high-risk groups of patients. The red dots represent upregulated genes in the high-risk group, while the green dots are the reversed. **b** KEGG and (**c**) GO enrichment analysis of DEGs. Bubble plots were generated to visualize the top 10 enrichment pathways. **d** UMAP plot showing the clustering of different cell populations in HNSCC tissues. **e** Bar charts showing the proportion of each cell population in different risk groups. **f** UMAP plot showing the subpopulation of MDCs in HNSCC tissues. **g** Bar charts showing the proportion of each MDCs subpopulation in different risk groups. DEGs differentially expressed genes, UMAP uniform manifold approximation and projection, HNSCC head and neck squamous cell carcinoma, MDCs myeloid derived cells, KEGG Kyoto Encyclopedia of Genes and Genomes, GO Gene ontology.

research should include slides obtained using different types of scanners with different parameters to enhance the adaptability of the model. Second, manual segmentation of tumor lesions is a time-consuming and labor-intensive task, and because of the complex anatomical structure, there is no mature tool for automatic segmentation of HNSCC lesions. Therefore, we are working to develop an automatic segmentation system for HNSCC lesions to prepare for future large-scale applications. Third, although our study had the largest sample size for predicting HNSCC outcomes, incorporating more radiotherapy patients is needed to enhance the model's reliability in predicting radiotherapy response, while additional RNA-seq data would improve its interpretability. Fourth, the predictive performance of the model in the TCIA cohort is not as good as that in the domestic external test cohorts; therefore, it is necessary to add foreign cohorts when training the model to enhance generalization. Finally, multicenter retrospective studies have the potential for selection and inherent biases. Therefore, designing prospective studies to enhance the generalizability and clinical applicability of our model is imperative.

In conclusion, we developed an MDLM to predict the prognosis and postoperative radiotherapy response in patients with HNSCC by integrating CT, WSI, and clinical features. The model was validated in multicenter cohorts and showed superior predictive performance, robustness, and generality, providing valuable insights for clinicians to make informed treatment decisions. Moreover, scRNA-seq and bulk RNA-seq revealed the underlying biological basis of MDLM. Prospective studies with large sample sizes are required to confirm the clinical validity and utility of these findings.

Materials and methods

Ethics Statement

The Ethics Committee of Yantai Yuhuangding Hospital approved this retrospective analysis of anonymous patient data (NO. 2024-215). Data collection and analysis in this retrospective study were conducted anonymously, thereby obviating the requirement for informed consent. Written informed consent was obtained from each patient who was enrolled in the scRNA-seq and bulk RNA-seq analysis. The study design followed the guidelines of the Declaration of Helsinki and the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis.

Study design and participants

This study aimed to develop and validate an MDLM for predicting the prognosis and postoperative radiotherapy response in patients with HNSCC. A total of 1087 patients from five Chinese hospitals and one TCIA cohort were enrolled in this study. The TCIA serves as a comprehensive resource for cancer imaging data, offering access to CT images. WSI images, clinical features, follow-up information (excluding disease-free survival), and bulk RNA-seq data of patients in the TCIA cohort were downloaded from The Cancer Genome Atlas (TCGA, https://portal.gdc.cancer.gov/). The MDLM was developed to utilize Cox hazard proportional regression to integrate CT-based DL signature, WSI-based DL signature, and clinical features. Additionally, we explored the biological basis of MDLM.

The same inclusion and exclusion criteria were applied across all the institutions. The inclusion criteria were as follows[1]: histologically confirmed diagnosis of HNSCC[2], treatment with surgical resection[3], contrast-enhanced CT scans of the neck performed within 1 week prior to surgery[4], accessible histological slides of formalin-fixed paraffin-embedded material, and[5] obtainable clinical information and follow-up data. The exclusion criteria were as follows[1]: a history of other malignancies[2], previous anticancer treatment[3], inadequate CT image quality, and[4] incomplete clinical information. The preoperative clinical information for all patients included sex, age, T stage, N stage, pathological differentiation status, smoking history, and alcohol history. All patients were staged according to the eighth edition of the AJCC staging criteria.

Patients were followed up every 3–6 months for at least 3 years or until the clinical endpoint, i.e., death. Recurrence was defined as the presence of local recurrence, cervical lymph node metastasis, or distant metastasis as determined by reoperation, endoscopy, CT, MRI, or PET/CT. DFS was defined as the duration from surgery until tumor recurrence or the patient's last known progression-free status. OS was defined as the duration from surgery to death from any cause.

CT and WSI image acquisition

All patients underwent contrast-enhanced neck CT scans using a multislice spiral CT scanner, with the scanning range extending from the skull base to the 7th cervical vertebra level. Following a plain CT scan, the contrast material was injected intravenously, followed by flushing with normal saline. Contrast-enhanced imaging was obtained 50s–60 s following intravenous injection.

Tumor samples were embedded in paraffin and subsequently sliced at 4 mm intervals. For each patient, a senior pathologist with 15 years of experience evaluated all hematoxylin and eosin-stained slides, and 1–3 slides containing ample tumor regions were chosen for digital scanning using the three-dimensional (3D) Histech Digital Pathology System (Pannoramic SCAN; Budapest, Hungary) with a 40× objective to acquire the WSI for analysis.

Tumor annotation on CT

The CT images of all patients in native DICOM format were uploaded to ITK-SNAP (version 3.6; www.itksnap.org) for manual annotation of the tumor lesion to obtain the region of interest (ROI). To enhance the reliability and consistency of delineating the ROI boundaries, the tumor boundaries were delineated by two radiologists in a blinded manner. The initial delineation was independently performed by a junior radiologist with 8 years of experience and subsequently reviewed by a senior radiologist with 16 years of experience.

CT and WSI image preprocessing

To ensure high-quality and standardised CT images, the intensity values of the CT scans were normalised to a standard range, with the window width and window level set to 350 and 50, respectively. Subsequently, the CT images and delineated 3D masks were resampled using trilinear interpolation to ensure a consistent spatial resolution of $1 \times 1 \times 1$ mm. The outlined mask is then cropped from the original image and expanded outward by 5 pixels. All cropped images were normalized and resized to the same size ($14 \times 54 \times 54$). Additionally, to enhance the generalisability of the model, data augmentation techniques were employed, including random flipping, cropping, and scaling. In addition, data augmentation techniques, including random flipping, cropping, and scaling, were employed to enhance the generalisation capability of the model.

For the WSI images, automated tissue segmentation was performed for each WSI using the public CLAM repository. Tissue segmentation of the WSI (40×) was performed using Otsu thresholding. Non-overlapping tiling was performed, and the tiles were resized to 256×256 pixels.

CT-based DL signature development

The development of the CT-based DL signature utilized a 3D-ResNet50 architecture, adapted from the traditional 2D ResNet50 model to process 3D volumetric CT data. Key modifications included replacing 2D convolutional kernels with 3D counterparts to capture spatial relationships across height, width, and depth, while a residual learning framework with shortcut connections was incorporated to mitigate degradation in deep networks. These features were then condensed into a fixed-length vector via fully connected layers, forming the CT signature—a quantitative representation of the tumor phenotype. The resulting features were subsequently fed into a sigmoid layer to generate the final CT Signature. For model training, we adopted the Adam optimizer with a learning rate of 1e-4 and a batch size of 8, running for 100 epochs, using the Cox proportional hazards loss function for survival prediction.

WSI-based DL signature development

To predict survival from WSIs, we expanded on the attention-based multiple instance learning algorithm initially developed for weakly supervised classification. The method begins with data preprocessing and tiling, where WSIs were divided into multiple non-overlapping tiles, each serving as a basic input unit for the model. For tile feature extraction, each tile was encoded using a ResNet50 model pre-trained on the ImageNet dataset. After the third residual block, adaptive mean-spatial pooling was applied to transform each 256×256 patch into a 1024-dimensional feature vector, processed with a batch size of 128 per GPU while leveraging parallel computation across multiple GPUs. Subsequently, feature aggregation and attention mechanisms were employed, involving feature space mapping where each tile's feature vector is reduced to a 512-dimensional space using a linear encoder to enhance discriminative information. An attention module then calculates the attention score for each tile, reflecting its importance for the final prediction. The 512-dimensional feature vectors were weighted by their attention scores and aggregated to form a bag-level feature vector representing the entire WSI. For risk score prediction, this bag-level feature vector was passed through a fully connected layer to generate the final risk score for patient prognosis, with training optimized using a negative log-likelihood loss function. The model was trained for 100 epochs with a learning rate of 1e-3, employing the Adam optimizer and processing one WSI per iteration (batch size = 1).

MDLM development

The MDLM was established by Cox regression, integrating the CT risk score, WSI risk score, and clinical features. First, univariate Cox regression analysis was conducted on the CT risk score, WSI risk score, and clinical features to identify risk factors associated with prognosis. Subsequently, factors with a significance level of $P < 0.05$ were included in the multivariate Cox regression analysis to identify independent risk factors.

Model validation and performance comparison

The Harrell's concordance index (C-index) and the areas under the ROC were calculated to evaluate the model's predictive ability for OS and DFS. Bootstrapping (1000 bootstrap intervals) was used to calculate 95% confidence intervals (CIs) for these performance metrics. The predictive performance of the MDLM was compared with that of unimodal models and validated in the test cohorts.

Predict postoperative radiotherapy response

Owing to the limited number of patients who received radiotherapy in the external test cohort and the lack of radiotherapy information in the TCIA cohort, the ability of the MDLM to predict benefits from postoperative radiotherapy was evaluated in both the training and internal test cohorts. We compared the prognosis of patients who received postoperative radiotherapy and those who did not receive postoperative radiotherapy in the different risk groups.

Bulk RNA-sequencing

Total RNA was extracted using the TRIzol reagent (Invitrogen, CA, USA) according to the manufacturer's protocol. RNA purity and quantification were evaluated using the NanoDrop 2000 spectrophotometer (Thermo Scientific, USA). RNA integrity was assessed using the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). Then, the libraries were constructed using VAHTS Universal V6 RNA-seq Library Prep Kit according to the manufacturer's instructions. The transcriptome sequencing and analysis were conducted by OE Biotech Co., Ltd. (Shanghai, China).

The libraries were sequenced on an Illumina Novaseq 6000 platform, and 150 bp paired-end reads were generated. Raw reads of fastq format were firstly processed using fastp, and the low quality reads were removed to obtain the clean reads. Then, about 20 million clean reads for each sample were retained for subsequent analyses. The clean reads were mapped to the reference genome using HISAT2. FPKM of each gene was calculated, and the read counts of each gene were obtained by HTSeq-count.

Development of gene-based prognostic model

In the Yuhuangding Hospital cohort, 128 patients with available bulk RNA-seq data were selected to establish the gene-based prognostic model. To identify survival-associated genes, we first performed univariate Cox regression analysis ($P < 0.05$, FDR-adjusted), followed by LASSO regression with 10-fold cross-validation to select the most predictive genes while preventing overfitting. The optimal penalty parameter (λ) was determined by selecting the value yielding the minimum cross-validation error, and genes with non-zero coefficients at this λ threshold were retained for subsequent model construction. Subsequently, the retained genes were incorporated into a multivariate Cox proportional hazards model to construct a gene-based prognostic model for HNSCC patients:

$$h(t) = h_0(t) \cdot \exp\left(\beta 1 \cdot Gene1 + \beta 2 \cdot \text{Gene2} + \cdots + \beta n \cdot \text{Gene}n\right)$$

The predictive performance of the model was evaluated using the C-index and further validated in the TCIA cohort.

Biological explainability

To obtain scRNA-seq samples, fresh tumor tissues were obtained from 10 HNSCC patients, with 6 patients predicted to be low-risk and 4 as high-risk by MDLM. ScRNA-seq data were aligned and quantified using Cell Ranger Single Cell Software Suite (10× Genomics) and were processed with "Seurat" R package (version 4.9.9). After quality control and batch effect correction, FindNeighbors and FindClusters functions in "Seurat" R package (version 4.9.9) were applied for cell clustering. Cell types in each cluster were defined based on the expression of known marker genes. The Findallmarker function in "Seurat" R package (version 4.9.9) was used to identify DEGs in each subcluster.

The bulk RNA-seq data were obtained from 50 patients in the TCIA cohort and 128 patients in the Yuhuangding Hospital cohort. The "limma" R package was utilized to analysis the DEGs between low- and high-risk patients, with criteria set at $|\log2FC| > 1$ and adjusted $P < 0.05$. Subsequently, GO and KEGG enrichment analyses were conducted using the "clusterProfiler" R package to identify signaling pathways associated with MDLM.

Statistical analysis

Student's t-test was used to compare continuous variables. The Chi-square and Fisher's exact tests were used to compare categorical variables. Survival curves were generated according to the Kaplan–Meier analysis, and the log-rank test was employed to compare the curves. Univariate and multivariate analyses using the proportional hazard model were calculated to examine the prognostic ability of DL signatures and clinical parameters, with hazard ratios expressed as medians and 95% CIs. The Bonferroni correction was used to obtain the corrected P-value for multiple comparisons. All analyses with bilateral $P < 0.05$ were considered statistically significant. All statistical analyses were performed using SPSS (version 27.0), MedCalc (version 18.2.1), R Studio (version 4.3.0), and Python (version 3.10.12).

Data availability

The datasets analyzed during the current study are not publicly available due to patient privacy purposes but are available from the corresponding author on reasonable request. Public datasets can be accessed through their respective websites (TCIA, https://www.cancerimagingarchive.net/; TCGA, https://portal.gdc.cancer.gov/). All code analyses in this study were performed using R Studio (version 4.3.0) and Python (version 3.10.12). All code necessary for the analyses will be made available at https://github.com/yyyhd/MDLM without any access restrictions.

Code availability

All code necessary for the analyses will be made available at https://github.com/yyyhd/MDLM.

Received: 19 December 2024; Accepted: 10 May 2025;
Published online: 23 May 2025

References

1. Ferlay, J. et al. Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. *Int J. Cancer* **144**, 194–1953 (2019).
2. Bray, F. et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **68**, 394–424 (2018).
3. Cramer, J. D., Burtness, B., Le, Q. T. & Ferris, R. L. The changing therapeutic landscape of head and neck cancer. *Nat. Rev. Clin. Oncol.* **16**, 669–683 (2019).
4. Johnson, D. E. et al. Head and neck squamous cell carcinoma. *Nat. Rev. Dis. Prim.* **6**, 92 (2020).
5. Wang, X., Hu, C. & Eisbruch, A. Organ-sparing radiation therapy for head and neck cancer. *Nat. Rev. Clin. Oncol.* **8**, 639–648 (2011).
6. Amin, M. B. et al. The Eighth Edition AJCC Cancer Staging Manual: continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging. *CA Cancer J. Clin.* **67**, 93–99 (2017).
7. Chen, R. J. et al. Pan-cancer integrative histology-genomic analysis via multimodal deep learning. *Cancer Cell* **40**, 865–878 (2022).
8. Ljungberg, B. et al. EAU guidelines on renal cell carcinoma: 2014 update. *Eur. Urol.* **67**, 913–924 (2015).
9. Chi, H. et al. Natural killer cell-related prognosis signature characterizes immune landscape and predicts prognosis of HNSCC. *Front. Immunol.* **13**, 1018685 (2022).
10. Chi, H. et al. Circadian rhythm-related genes index: A predictor for HNSCC prognosis, immunotherapy efficacy, and chemosensitivity. *Front. Immunol.* **14**, 1091218 (2023).
11. Liu, D. et al. Developing a pyroptosis-related gene signature to better predict the prognosis and immune status of patients with head and neck squamous cell carcinoma. *Front. Gene* **13**, 988606 (2023).
12. Mayerhoefer, M. E. et al. Introduction to radiomics. *J. Nucl. Med.* **61**, 488–495 (2020).

13. Xue, T. et al. A CT-based radiomics nomogram in predicting the postoperative prognosis of colorectal cancer: a two-center study. *Acad. Radio.* **29**, 1647–1660 (2022).
14. Wang, T. et al. Radiomics for survival risk stratification of clinical and pathologic stage IA pure-solid non-small cell lung cancer. *Radiology* **302**, 425–434 (2022).
15. Li, J. et al. CT-based delta radiomics in predicting the prognosis of stage IV gastric cancer to immune checkpoint inhibitors. *Front. Oncol.* **12**, 1059874 (2022).
16. Liu, X. et al. CT-based radiomics signature analysis for evaluation of response to induction chemotherapy and progression-free survival in locally advanced hypopharyngeal carcinoma. *Eur. Radio.* **32**, 7755–7766 (2022).
17. Abe, K. et al. Evaluation of the MVCT-based radiomic features as prognostic factor in patients with head and neck squamous cell carcinoma. *BMC Med. Imaging* **23**, 102 (2023).
18. Tang, F. H. et al. Radiomics from various tumour volume sizes for prognosis prediction of head and neck squamous cell carcinoma: a voted ensemble machine learning approach. *Life***12**, 1380 (2022).
19. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
20. Zheng, Y. M. et al. A CT-based deep learning radiomics nomogram for the prediction of EGFR mutation status in head and neck squamous cell carcinoma. *Acad. Radio.* **31**, 628–638 (2024).
21. Zheng, Y. M. et al. A CT-based deep learning radiomics nomogram to predict histological grades of head and neck squamous cell carcinoma. *Acad. Radio.* **30**, 1591–1599 (2023).
22. Chen, Z. et al. A deep learning and radiomics fusion model based on contrast-enhanced computer tomography improves preoperative identification of cervical lymph node metastasis of oral squamous cell carcinoma. *Clin. Oral. Investig.* **28**, 39 (2023).
23. Boehm, K. M., Khosravi, P., Vanguri, R., Gao, J. & Shah, S. P. Harnessing multimodal data integration to advance precision oncology. *Nat. Rev. Cancer* **22**, 114–126 (2022).
24. Shao, L. et al. Multiparametric MRI and whole slide image-based pretreatment prediction of pathological response to neoadjuvant chemoradiotherapy in rectal cancer: a multicenter radiopathomic study. *Ann. Surg. Oncol.* **27**, 4296–4306 (2020).
25. Zhang, Y. et al. Histopathology images-based deep learning prediction of prognosis and therapeutic response in small cell lung cancer. *NPJ Digit. Med.* **7**, 15 (2024).
26. Saillard, C. et al. Predicting Survival After Hepatocellular Carcinoma Resection Using Deep Learning on Histological Slides. *Hepatology* **72**, 2000–2013 (2020).
27. Feng, L. et al. Development and validation of a radiopathomics model to predict pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer: a multicentre observational study. *Lancet Digit. Health* **4**, e8–e17 (2022).
28. She, Y. et al. Deep learning for predicting major pathological response to neoadjuvant chemoimmunotherapy in non-small cell lung cancer: a multicentre study. *EBioMedicine* **86**, 104364 (2022).
29. Zhong, Y. et al. Multi-omics classifier for pulmonary nodules (MISSION) collaborative group. deep learning for prediction of N2 metastasis and survival for clinical stage I non-small cell lung cancer. *Radiology* **302**, 200–211 (2022).
30. Zhang, Y. et al. Single-cell RNA sequencing in cancer research. *J. Exp. Clin. Cancer Res.* **40**, 81 (2021).
31. Feng, Z. et al. CT radiomics to predict macrotrabecular-massive subtype and immune status in hepatocellular carcinoma. *Radiology* **307**, e221291 (2023).
32. Liu, Z. et al. Radiomics-based prediction of survival in patients with head and neck squamous cell carcinoma based on pre- and post-treatment ^{18}F-PET/CT. *Aging***12**, 14593–14619 (2020).
33. Vollmer, A. et al. Multimodal artificial intelligence-based pathogenomics improves survival prediction in oral squamous cell carcinoma. *Sci. Rep.* **14**, 5687 (2024).
34. Halicek, M. et al. Head and neck cancer detection in digitized whole-slide histology using convolutional neural networks. *Sci. Rep.* **9**, 14043 (2019).
35. Wang, R., Khurram, S. A., Walsh, H., Young, L. S. & Rajpoot, N. A novel deep learning algorithm for human papillomavirus infection prediction in head and neck cancers using routine histology images. *Mod. Pathol.* **36**, 100320 (2023).
36. Xie, Q. et al. A clinical-radiomic-pathomic model for prognosis prediction in patients with hepatocellular carcinoma after radical resection. *Cancer Med.* **13**, e7374 (2024).
37. Boehm, K. M. et al. Multimodal data integration using machine learning improves risk stratification of high-grade serous ovarian cancer. *Nat. Cancer* **3**, 723–733 (2022).
38. Foy, J. P. et al. A 13-gene expression-based radioresistance score highlights the heterogeneity in the response to radiation therapy across HPV-negative HNSCC molecular subtypes. *BMC Med.* **15**, 165 (2017).
39. Chen, L. et al. Prediction of radiotherapy response with a 5-microRNA signature-based nomogram in head and neck squamous cell carcinoma. *Cancer Med.* **7**, 726–735 (2018).
40. Jiang, Y. et al. Non-invasive tumor microenvironment evaluation and treatment response prediction in gastric cancer using deep learning radiomics. *Cell Rep. Med.* **4**, 101146 (2023).
41. McFadyen, M. C., Melvin, W. T. & Murray, G. I. Cytochrome P450 enzymes: novel options for cancer therapeutics. *Mol. Cancer Ther.* **3**, 363–371 (2004).
42. Xiao, Y. & Yu, D. Tumor microenvironment as a therapeutic target in cancer. *Pharm. Ther.* **221**, 10775 (2021).
43. Keck, M. K. et al. Integrative analysis of head and neck cancer identifies two biologically distinct HPV and three non-HPV subtypes. *Clin. Cancer Res.* **21**, 870–881 (2015).
44. Lazaratos, A. M., Annis, M. G. & Siegel, P. M. GPNMB: a potent inducer of immunosuppression in cancer. *Oncogene* **41**, 4573–4590 (2022).
45. Sun, L. et al. Comprehensive analysis of prognostic value and immune infiltration of ficolin family members in hepatocellular carcinoma. *Front. Genet.* **13**, 913398 (2022).
46. Xie, Z. et al. Single-cell analysis unveils activation of mast cells in colorectal cancer microenvironment. *Cell Biosci*. **13**, 217 (2023).

Acknowledgements

This work was supported by the Taishan Scholar Project (No.ts20190991, tsqn202103197), the Key R&D Project of Shandong Province (2022CXPT023), and the Natural Science Foundation of China (82371933). The funder played no role in study design, data collection, analysis, and interpretation of data, or the writing of this manuscript. We thank all the study participants.

Author contributions

Conceptualization: X.S., N.M., H.C., D.Y., and H.L. Methodology: R.T., F.H., H.Z. Data Curation: G.Y., P.Y., J.L., T.Y., X.C., Y.C., Y.H., Y.Y., H.Z., P.Y., L.S., A.L., and Z.L. Formal analysis: R.T., F.H., and H.Z. Investigation: R.T., F.H., H.Z., G.Y., P.Y., H.L., D.Y., H.C., N.M., and X.S. Visualization: R.T., H.Z., and H.F. Funding acquisition: X.S. and N.M. Project administration: X.S., N.M., H.C., D.Y., F.H., and H.L; Supervision: X.S. and N.M. Writing—original draft: R.T., H.Z., and H.F. Writing—review & editing: X.S., N.M., F.H., H.C., D.Y., and H.L. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41746-025-01712-0.

Correspondence and requests for materials should be addressed to Huaiqing Lv, Dexin Yu, Hongxia Cheng, Ning Mao or Xicheng Song.

[1]Department of Otorhinolaryngology, Head and Neck Surgery, Yantai Yuhuangding Hospital, Qingdao University, Yantai, China. [2]Department of Pathology, Affiliated Hospital of Qingdao University, Qingdao, China. [3]Key Laboratory of Biomedical Information Engineering of Ministry of Education, School of Life Science and Technology, Xi'an Jiaotong University, Xi'an, China. [4]Department of Pathology, Yantai Yuhuangding Hospital, Qingdao University, Yantai, China. [5]The Second School of Clinical Medicine of Binzhou Medical University, Yantai, China. [6]Shandong University of Traditional Chinese Medicine, Jinan, China. [7]Department of Radiology, Qilu Hospital of Shandong University, Jinan, China. [8]Linyi People's Hospital Affiliated to Shandong Second Medical University, Linyi, China. [9]Department of Pathology, Shandong Provincial Hospital Affiliated to Shandong First Medical University, Jinan, China. [10]Big Data and Artificial Intelligence Laboratory, Yantai Yuhuangding Hospital, Qingdao University, Yantai, China. [11]Department of Radiology, Yantai Yuhuangding Hospital, Qingdao University, Yantai, China. [12]Shandong Provincial Key Laboratory of Neuroimmune Interaction and Regulation, Yantai, China. [13]Shandong Provincial Clinical Research Center for Otorhinolaryngologic Diseases, Yantai, China. [14]Yantai Key Laboratory of Otorhinolaryngologic Diseases, Yantai, China. [15]These authors contributed equally: Ruxian Tian, Feng Hou, Haicheng Zhang, Guohua Yu, Ping Yang. ✉e-mail: lvhuaiqing@126.com; yudexin0330@sina.com; hongxiacheng322@hotmail.com; maoning@pku.edu.cn; drxchsong@163.com