Journal of Theoretical and Applied Information Technology
31st October 2026. Vol.104. No.20
© Little Lion Scientific
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

ONCOGRAPH-RLNET: BIOLOGICALLY CONSTRAINED GRAPH REINFORCEMENT LEARNING FOR REPRODUCIBLE BIOMARKER DISCOVERY, RADIOTHERAPY PLAN ERROR DETECTION AND RISK-AWARE TREATMENT OPTIMISATION

CH SAI LAKSHMI[1], DR. CH. VENKATA RAMANA REDDY[2], DR L RAGHU KUMAR[3]

[1]Research Scholar, Department of CSE, Gandhi Institute of Engineering and Technology University, Gunupur, Odisha, India

[2]Professor, Department of CSE, Gandhi Institute of Engineering and Technology University, Gunupur, Odisha, India

[3]Associate Professor, Department of CSE, KG Reddy College of Engineering, Hyderabad, India

Corresponding author email :chikkalasai.lakshmi@giet.edu

ABSTRACT

Background: Treatment decisions in oncology depend on both tumour biology and the quality of the radiotherapy plan that delivers treatment, yet these are modelled separately, without biological network structure or sequential, risk-aware decision-making. **Methods:** OncoGraph-RLNet combines (i) gated multi-omics fusion tolerant to missing modalities, (ii) a graph attention network whose logits are biased by a gene–pathway knowledge graph, (iii) LP-Stability, a resampling-based score that selects reproducible gene signatures, (iv) a patient-conditioned CNN–Transformer for plan-error detection and dose prediction, and (v) a conservative offline distributional RL agent optimising a mean–CVaR objective under a toxicity budget. Evaluation uses TCGA, TCIA and OpenKBP with leakage-safe splits. **Results:** OncoGraph-RLNet reached 96.2 ± 0.8 % stratification accuracy, a C-index of 0.892 and a signature stability index of 0.872 (best baseline, GAT: 93.4 %). It detected injected plan errors with AUROC 0.981 and lowered the dose score to 0.087 Gy (best baseline 0.118 Gy). The learned policy reduced cost violations to 4.2 % (7.4 % for quantile CQL) at 97.1 % guideline concordance. Removing LP-Stability, the graph branch, the biological prior or CVaR each degraded the corresponding metric.
Conclusions: Coupling biologically constrained graph learning, plan-error detection and risk-aware offline RL yields more stable signatures, more sensitive plan screening and safer policies than single-purpose models; clinical use requires prospective validation.

Keywords: *Precision Oncology, Multi-Omics, Graph Attention Network, Offline Reinforcement Learning, Radiotherapy Quality Assurance*

Highlights

- Unified graph reinforcement-learning framework linking tumour biology and radiotherapy quality.
- Biologically biased graph attention with gene gates yields interpretable molecular signatures.
- LP-Stability ranks biomarkers by connectivity, pathway, clinical relevance and resampling stability.
- Plan-error detector conditioned on patient biology and trained with controlled error injection.
- Conservative, CVaR-aware offline policy learning evaluated with doubly robust estimators.

1. INTRODUCTION

Even patients sharing a specific histological diagnosis experience varied clinical trajectories due to somatic and regulatory changes, rewired pathways, microenvironment of the tumor and treatment tolerance [1, 2]. Thus, precision medicine in oncology requires personalised therapy based on molecular characteristics, radiological phenotype, and clinical evolution [3]. Machine learning algorithms are effective for screening, pathology, and dose prediction [4-6]; however, open databases such as TCGA [7], TCIA [8], and OpenKBP [6] facilitate research, while physician decision-making remains underrepresented [9].

There are three major deficiencies in current research: (1) structure-oblivious multimodal fusion when modalities are concatenated or voted without accounting for genes' and pathways' topologies [10, 11]. (2) unstable biomarkers where rankings of univariate associations are inconsistent across

cohorts and repetitions [12, 13]. (3) separation of plan quality and decision-making, while radiotherapy models predict dose [14, 15] and sequential decision models do not account for plan uncertainty and biological factors of the tumour [16, 17].

We propose OncoGraph-RLNet, which addresses these gaps with the following contributions:

(1) a unified, mathematically specified pipeline for biological representation learning, plan-error detection and risk-aware sequential decision support (Section 4);
(2) graph attention with a biological prior in its logits, which interpolates between standard GAT and random-walk propagation (Proposition 2), and LP-Stability, a bounded gene score with a signature-level stability index;
(3) a patient-conditioned CNN–Transformer trained with controlled error injection, and a conservative mean–CVaR offline RL objective with a toxicity budget and doubly robust evaluation;
(4) a leakage-safe protocol with ablations, robustness analyses and pre-specified statistics on public data.

The study addresses three research questions. RQ1: does a gene–pathway prior inside graph attention, together with a resampling-based stability score, give more accurate and more reproducible molecular stratification than structure-agnostic or prior-free models? RQ2: does conditioning plan-error detection and dose prediction on patient biology improve sensitivity, calibration and dose accuracy over anatomy-only models? RQ3: can a conservative, CVaR-aware offline policy lower toxicity-budget violations while remaining guideline-concordant? Accordingly, the objectives are (O1) to build and validate the biologically constrained representation and LP-Stability signatures, (O2) to develop and evaluate the patient-conditioned plan-error and dose branch, (O3) to learn and evaluate a risk-aware treatment policy, and (O4) to establish by ablation and robustness analysis which components produce the gains.

2. RELATED WORK

Multi-omics and graph learning. Reviews stress that omics layers are complementary but need harmonisation, missing-data handling and external validation [18, 19]. Similarity network fusion [20], multimodal survival networks [10] and Pathomic Fusion [21] integrate modalities; graph neural networks [22, 23] add relational structure, as in MOGONET, which fuses sample-similarity graphs per omics layer [11]. Curated gene topology is rarely used to bias attention, and radiotherapy is absent.

Radiotherapy AI. Knowledge-based planning predicts achievable dose from anatomy [24, 25], now with deep networks [26] and the OpenKBP benchmark [6, 14, 15]. Detecting erroneous plans and relating them to patient biology has received little attention.

Reinforcement learning. RL has been applied to sequential treatment choice [16, 27], with guidance on reward design and off-policy evaluation [28]. Offline methods that penalise unsupported actions [29, 30], distributional critics [31] and doubly robust estimators [32] suit retrospective data but have not been combined with biological graph representations. Attribution methods [33–35] complement the intrinsic explanations provided here. Table 1 positions the closest studies.

Critical appraisal. Established multi-omics integrators have clear strengths but also limits for the needs of this work. Similarity network fusion [20] and MOGONET [11] derive their graphs from sample similarity rather than curated gene–pathway topology, so the relations learned are data-driven and not biologically constrained, and biomarker rankings from such models are typically not accompanied by a resampling-based reproducibility measure [12, 13]. Multimodal survival networks [10] and Pathomic Fusion [21] handle missing or heterogeneous modalities, but remain static predictors with no radiotherapy component and no downstream decision. In radiotherapy, dose predictors benchmarked on OpenKBP [6, 14, 15] are judged on dose accuracy alone; they neither flag erroneous plans nor condition on tumour biology. Sequential-decision work in medicine [16, 28] has mainly reported mean return and, in its best-known example, addressed intensive care rather than oncology, while conservative [29, 30] and distributional [31] offline methods address distribution shift or risk separately from biological representations. No single study therefore connects biological structure, plan reliability and risk-constrained decisions, which is the need this work addresses.

Table 1: Positioning Of Representative Studies.

Study	Data	Method	Contribution	Gap addressed
Wang et al. [11]	Multi-omics	Sample-graph GCN + fusion	Classification, biomarkers	No plan quality; no decisions
Cheerla & Gevaert [10]	TCGA pan-cancer	Multimodal survival network	Robust to missing modalities	No graph prior; static
Chen et al. [21]	TCGA histology + genomics	Pathomic Fusion	Gated fusion	No radiotherapy; no policy
Babier et al. [6]	OpenKBP	Benchmark	Common dose protocol	No error detection or biology
Gronberg [14]; Liu [15]	OpenKBP	3D U-Net variants	Accurate dose prediction	Anatomy only
Komorowski et al. [16]	ICU records	Off-policy RL	Policy from records	Non-oncology; mean return only
This work	TCGA + TCIA + OpenKBP	Biased GAT + LP-Stability + conditioned RT + CVaR offline RL	Biology, plan reliability and risk-aware decisions	Retrospective, in-silico policy evaluation

3. PROBLEM FORMULATION

Let $\mathcal{M} = \{g, t, c, v, d\}$ index mutation/copy-number, transcriptome, clinical, imaging-derived and radiotherapy plan/dose modalities. Patient $i \in \{1, \dots, N\}$ has vectors $\boldsymbol{x}_i^{(m)}$ for $m \in \mathcal{M}_i \subseteq \mathcal{M}$, a response label y_i, a survival pair (t_i, δ_i) and, when available, a plan-error label y_i^e and reference dose D_i. Three coupled tasks are addressed: **T1** learn $p(y_i \mid \cdot)$ and a hazard score η_i, and extract a K-gene signature $\mathcal{T}_K$ that is stable under resampling; **T2** estimate the plan-error probability e_i and dose $\hat{D}_i$; **T3** learn a policy $\pi(a \mid s)$ over a clinician-defined action set $\mathcal{A}$ that maximises expected benefit and lower-tail return subject to a toxicity budget. Symbols are listed in Table 2.

Table 2: Nomenclature.

Symbol	Meaning	Symbol	Meaning
N, p, d	Patients, genes, latent dimension	$\boldsymbol{z}_i^{(m)}, \boldsymbol{z}_i$	Modality and fused embeddings
$a_i^{(m)}, \omega_i^{(m)}$	Fusion weight, observation mask	$\mathcal{G}, \hat{\boldsymbol{W}}$	Gene–pathway graph, sparsified weights
α_{jk}, κ	Attention; strength of biological prior	m_j	Learnable gene gate
u_j, S_j, GS_j	Base importance, stability, LP-Stability score	$\mathcal{T}_K, KI$	Signature, Kuncheva index
$\hat{\boldsymbol{e}}_i, e_i$	Error posterior, error probability	s_t, a_t, r_t, c_t	State, action, reward, toxicity cost
Z_n, Q^{risk}	Return quantiles, risk-adjusted value	ζ, ς	Mean–CVaR weight, CVaR tail level
π, μ	Target, behaviour policy	$\lambda_\bullet, \gamma$	Loss weights, discount factor

4. METHODOLOGY

4.1 Overview

OncoGraph-RLNet (see Fig. 1) consists of modality encoders with gating fusion, biological biased graph attention head with LP-Stability scoring, patient-conditioned radiotherapy head, and risk-aware offline RL policy whose state represents information from the other heads. All normalisations, filters, graph metrics, and feature scores are calculated based on the training dataset only.

4.2 Encoders And Adaptive Gated Fusion

Features are standardised with training statistics and mapped by modality-specific encoders f_m to a shared space. An attention score and an observation mask $\omega_i^{(m)} \in \{0,1\}$ give patient-specific modality weights, so that high-dimensional expression vectors cannot dominate low-dimensional but decisive variables such as stage:

$$\tilde{x}_{ij}^{(m)} = \frac{x_{ij}^{(m)} - \mu_j^{(m)}}{\sigma_j^{(m)} + \epsilon}, \boldsymbol{z}_i^{(m)} = f_m\left(\tilde{\boldsymbol{x}}_i^{(m)}; \boldsymbol{\theta}_m\right) \in \mathbb{R}^d \quad (1)$$

$$s_i^{(m)} = \boldsymbol{q}^\top tanh\left(\boldsymbol{W}_a \boldsymbol{z}_i^{(m)} + \boldsymbol{b}_a\right), a_i^{(m)} = \frac{\omega_i^{(m)} exp\left(s_i^{(m)}\right)}{\sum_{m'} \omega_i^{(m')} exp\left(s_i^{(m')}\right)}$$
$$\boldsymbol{z}_i = \sum_{m \in \mathcal{M}} a_i^{(m)} \boldsymbol{z}_i^{(m)} \quad (2)$$

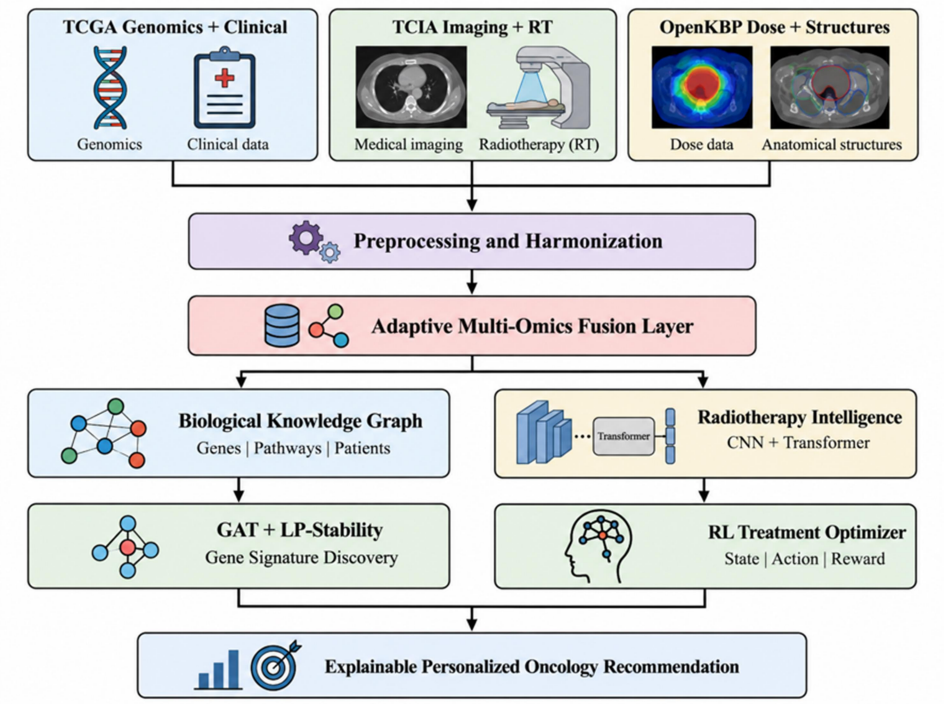


Figure 1: Architecture Of OncoGraph-RLNet. Encoded modalities are fused by gated attention; the representation drives a graph branch with LP-Stability scoring and a patient-conditioned plan-error branch; both feed a risk-aware offline RL agent that returns an explainable recommendation.

4.3 Biological Knowledge Graph

A single graph $\mathcal{G} = \left(\mathcal{V}_G \cup \mathcal{V}_P, \mathcal{E}, \hat{\boldsymbol{W}}\right)$ of genes ($|\mathcal{V}_G| = p$) and pathways is shared by all patients, keeping the model inductive. For genes j, k, edge weights combine training-cohort Spearman co-expression ρ_{jk}, interaction confidence b_{jk} [36] and the Jaccard index J_{jk} of KEGG pathway sets [37]; edges are pruned at a validation-chosen threshold τ, and pathway nodes connect to member genes with weight 1:

$$w_{jk} = \beta_1 \left|\rho_{jk}\right| + \beta_2 b_{jk} + \beta_3 J_{jk}$$
$$\hat{w}_{jk} = w_{jk} 1\left[w_{jk} \geq \tau\right], \boldsymbol{L} = \boldsymbol{I} - \boldsymbol{D}^{-1/2} \hat{\boldsymbol{W}} \boldsymbol{D}^{-1/2} \quad (3)$$

4.4 Biologically Biased Graph Attention With Gene Gates

For patient i, gene node j is initialised from its expression and mutation values scaled by a gate $m_j = \sigma\left(\phi_j\right)$, and pathway nodes from learned embeddings. Attention logits add a scaled log-edge-weight prior to learned compatibility:

$$\boldsymbol{h}_{ij}^{(0)} = m_j \boldsymbol{W}_0 \left[\tilde{x}_{ij}^{(t)} ; x_{ij}^{(g)}\right] + \boldsymbol{e}_j$$
$$e_{jk}^{(l,h)} = LeakyReLU\left(\boldsymbol{a}_h^\top \left[\boldsymbol{W}_h^{(l)} \boldsymbol{h}_{ij}^{(l)} \| \boldsymbol{W}_h^{(l)} \boldsymbol{h}_{ik}^{(l)}\right]\right) + \kappa log\left(\hat{w}_{jk} + \epsilon\right), \alpha_{jk}^{(l,h)} = \underset{k\in\mathcal{N}(j)}{softmaxe}_{jk}^{(l,h)} \quad (4)$$
$$\boldsymbol{h}_{ij}^{(l+1)} = \|_{h=1}^{H} \sigma\left(\sum_{k\in\mathcal{N}(j)} \alpha_{jk}^{(l,h)} \boldsymbol{W}_h^{(l)} \boldsymbol{h}_{ik}^{(l)}\right)$$
$$\boldsymbol{g}_i = \sum_{j\in\mathcal{V}_G} \vartheta_{ij} \boldsymbol{h}_{ij}^{(L)}, \boldsymbol{r}_i = \left[\boldsymbol{z}_i \| \boldsymbol{g}_i\right] \quad (5)$$

Here $\kappa = 0$ recovers standard GAT and $\kappa = 1$ with a constant learned term gives random-walk propagation on $\hat{\boldsymbol{W}}$ (Proposition 2); ϑ_{ij} are attention-pooling weights. The branch is trained with link reconstruction and Laplacian smoothness, and with a response cross-entropy plus Cox partial likelihood [38] for the risk score $\eta_i = \boldsymbol{w}_\eta^\top \boldsymbol{r}_i$:

$$\mathcal{L}_{graph} = -\frac{1}{|\mathcal{E}_{pos}|}\sum_{(j,k)\in\mathcal{E}_{pos}} log\,\sigma\left(\boldsymbol{h}_j^\top \boldsymbol{h}_k\right) - \frac{1}{|\mathcal{E}_{neg}|}\sum_{(j,k)\in\mathcal{E}_{neg}} log\left(1-\sigma\left(\boldsymbol{h}_j^\top \boldsymbol{h}_k\right)\right) + \mu tr\left(\boldsymbol{H}^\top \boldsymbol{L}\boldsymbol{H}\right)$$
$$\mathcal{L}_{pred} = -\frac{1}{N}\sum_i log\,\hat{y}_{i,y_i} - \lambda_{cox}\frac{1}{N_\delta}\sum_{i:\delta_i=1}\left[\eta_i - log\sum_{k\in\mathcal{R}(t_i)} exp\left(\eta_k\right)\right] \quad (6)$$

4.5 LP-Stability Signature Discovery

LP-Stability (Laplacian-perturbation stability) scores gene j by min–max-normalised training-set evidence: attention centrality A_j (mean attention received), connectivity B_j (normalised weighted degree), pathway contribution P_j (mean enrichment significance $-log_{10} q_\ell$ of its pathways [39, 40]) and clinical relevance C_j (absolute univariate Cox Wald statistic). Over B bootstrap resamples of patients [41] with a fraction ξ of edges dropped, the base score $u_j^{(b)}$ is recomputed and gene stability is one minus its coefficient of variation [12]:

$$u_j = \lambda_A A_j + \lambda_B B_j + \lambda_P P_j + \lambda_C C_j, \lambda_A + \lambda_B + \lambda_P + \lambda_C = 1$$
$$S_j = max\left\{0, 1 - \frac{sd_b\left(u_j^{(b)}\right)}{\bar{u}_j + \epsilon}\right\}, GS_j = (1-\lambda_S)\bar{u}_j + \lambda_S S_j, \mathcal{T}_K = Top_K\{GS_j\} \quad (7)$$

By construction $GS_j \in [0,1]$ (Proposition 3). Signature reproducibility across resamples is measured by the Kuncheva index [13], and the gates are regularised for sparsity and consistency across two edge-dropout views:

$$KI = \frac{2}{B(B-1)}\sum_{a<b}\frac{\left|\mathcal{T}_K^{(a)} \cap \mathcal{T}_K^{(b)}\right| - K^2/p}{K - K^2/p}, \mathcal{L}_{sig} = \nu_1 \|\boldsymbol{m}\|_1 + \nu_2 \mathbb{E}\left\|\hat{\boldsymbol{u}}^{(1)} - \hat{\boldsymbol{u}}^{(2)}\right\|_2^2 \quad (8)$$

4.6 Patient-Conditioned Radiotherapy Branch

A 3D CNN stem converts CT, structure masks and dose $[\boldsymbol{C}_i, \boldsymbol{M}_i, \boldsymbol{D}_i]$ into n tokens processed by a Transformer encoder [42, 43] to model long-range target–organ relations. The pooled feature $\boldsymbol{r}_i^{RT}$ is modulated by the fused patient vector through FiLM [44]; a class-balanced focal loss [45] and a masked voxel-wise dose term are used jointly:

$$\overset{\sim}{\boldsymbol{r}}_i^{RT} = \left(1 + \boldsymbol{k}_1(\boldsymbol{z}_i)\right) \odot \boldsymbol{r}_i^{RT} + \boldsymbol{k}_2(\boldsymbol{z}_i), \hat{\boldsymbol{e}}_i = softmax\left(\boldsymbol{W}_e \overset{\sim}{\boldsymbol{r}}_i^{RT} + \boldsymbol{b}_e\right), e_i = 1 - \hat{e}_{i,0}$$
$$\mathcal{L}_{RT} = -\frac{1}{N}\sum_i \sum_{c=0}^{K_e} \varpi_c \left(1 - \hat{e}_{ic}\right)^{\gamma_f} y_{ic}^e log\hat{e}_{ic} + \lambda_D \frac{1}{N}\sum_i \frac{1}{|\Omega_i|}\sum_{v\in\Omega_i}\left|D_i(v) - \hat{D}_i(v)\right| \quad (9)$$

Dose agreement is measured with the gamma index [46] for distance and dose tolerances $\Delta d, \Delta D$, and the gamma pass rate (GPR):

$$\Gamma(\boldsymbol{r}_e) = \min_{\boldsymbol{r}_r}\sqrt{\frac{\|\boldsymbol{r}_r - \boldsymbol{r}_e\|^2}{\Delta d^2} + \frac{\left(D_r(\boldsymbol{r}_r) - D_e(\boldsymbol{r}_e)\right)^2}{\Delta D^2}}$$
$$GPR = \frac{|\{\boldsymbol{r}_e : \Gamma(\boldsymbol{r}_e) \le 1\}|}{|\{\boldsymbol{r}_e\}|} \quad (10)$$

4.7 Risk-Aware Offline Reinforcement Learning

Planning is a finite-horizon MDP $(\mathcal{S}, \mathcal{A}, P, r, c, \gamma, T)$ (Fig. 2b). The state combines the patient representation, error posterior and time-varying observations $\boldsymbol{o}_t$ (response, toxicity grade). The reward trades benefit B_t against error risk and departure from guideline-concordant actions; toxicity enters a separate cost. With return $G = \sum_t \gamma^t r_t$, we maximise a mean–CVaR mixture [47] under a budget d_{max}, handled by a Lagrange multiplier λ_c updated by projected dual ascent:

$$s_t = \left[\boldsymbol{r}_i; \hat{\boldsymbol{e}}_{i,t}; \boldsymbol{o}_t\right], r_t = \eta B_t - \lambda_{err}\hat{e}_t - \lambda_{reg}1\left[a_t \notin \mathcal{A}_{guide}(s_t)\right], c_t = Tox(s_t, a_t)$$
$$\pi^\star = \underset{\pi}{argmax}(1-\zeta)\mathbb{E}_\pi[G] + \zeta CVaR_\varsigma(G) s.t. \mathbb{E}_\pi[\textstyle\sum_t \gamma^t c_t] \le d_{max} \quad (11)$$
$$CVaR_\varsigma(G) = \max_\nu \left(\nu - \frac{1}{\varsigma}\mathbb{E}[max(0, \nu - G)]\right), \lambda_c \leftarrow max\left(0, \lambda_c + \eta_\lambda\left(\hat{J}_c - d_{max}\right)\right)$$

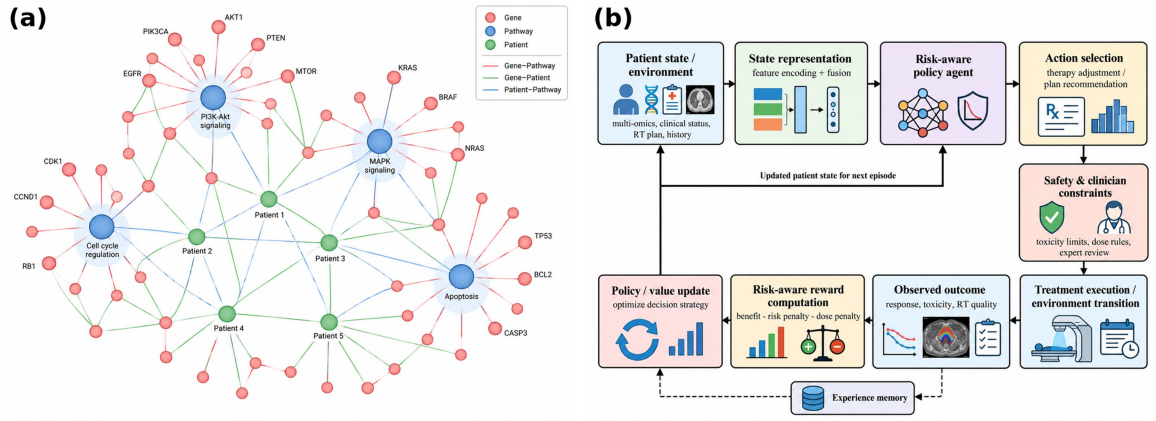


Figure 2: Schematic Gene–pathway Graph With A Patient-Specific Node Signal (a) And Decision Loop Of The Risk-Aware Agent (b).

The critic outputs N_q return quantiles $Z_n(s,a)$ [31]; actions maximise the risk-adjusted value below, with $Z_{(n)}$ the sorted quantiles. Training minimises a quantile-Huber loss with a conservative regulariser [29], and rewards are penalised by ensemble disagreement of the dynamics model [30]:

$$Q^{risk}(s,a) = \frac{1-\zeta}{N_q}\sum_{n=1}^{N_q} Z_n(s,a) + \frac{\zeta}{\lceil \varsigma N_q \rceil}\sum_{n=1}^{\lceil \varsigma N_q \rceil} Z_{(n)}(s,a), \pi(s) = \underset{a}{argmax} Q^{risk}(s,a)$$
$$\mathcal{L}_{RL} = \frac{1}{N_q}\sum_{n,n'} \rho_{\tau_n}^{\chi}(\delta_{nn'}) + \lambda_{cql}\mathbb{E}_{s\sim\mathcal{D}}\left[log\sum_a exp\,\bar{Q}(s,a) - \mathbb{E}_{a\sim\hat{\mu}}\bar{Q}(s,a)\right], \tilde{r}_t = r_t' - \lambda_u u(s_t, a_t) \quad (12)$$

Here $\delta_{nn'}$ are temporal-difference errors with $r_t' = r_t - \lambda_c c_t$, $\bar{Q}$ is the mean quantile value, ρ_τ^χ the quantile-Huber function, and u the ensemble standard deviation of predicted next states. Policies are evaluated off-policy with the doubly robust estimator [32], where $\varrho_{0:t}$ are importance ratios under an estimated behaviour policy $\hat{\mu}$; explanations use integrated gradients [34] and an attention-entropy penalty $\mathcal{L}_{XAI}$:

$$\hat{V}_{DR}(\pi) = \frac{1}{N}\sum_{i=1}^{N}\sum_{t=0}^{T}\gamma^t\left[\varrho_{0:t}^{(i)}\left(r_t^{(i)} - \hat{Q}\left(s_t^{(i)}, a_t^{(i)}\right)\right) + \varrho_{0:t-1}^{(i)}\hat{V}\left(s_t^{(i)}\right)\right] \quad (13)$$

4.8 Joint Objective And Algorithms

$$\mathcal{L}_{total} = \mathcal{L}_{pred} + \lambda_1\mathcal{L}_{graph} + \lambda_2\mathcal{L}_{sig} + \lambda_3\mathcal{L}_{RT} + \lambda_4\mathcal{L}_{XAI} + \lambda_5\mathcal{L}_{RL} \quad (14)$$

Training is staged: encoders and fusion are pre-trained; graph and signature terms are added; the radiotherapy branch is trained with the representation frozen; the RL critic is trained with

stop-gradient on the representation (λ_5 acts only on critic parameters). AdamW [48] with early stopping is used. Algorithm 1–Algorithm 3 summarise the procedure.

Algorithm 1: Representation Learning With Biologically Biased Graph Attention.

Input: Training data $\mathcal{D}_{tr}$, interaction sources, $(\beta_{\bullet}, \tau, \kappa, L, H, d)$

Output: Encoders, fusion, GAT parameters, gates $\boldsymbol{m}$

1: Standardise all splits with training statistics Eq. (1); build $\hat{\boldsymbol{W}}$ and $\boldsymbol{L}$ Eq. (3)

2: **for** epoch $e = 1, \dots, E$ **and** mini-batch $\mathcal{B} \subset \mathcal{D}_{tr}$ **do**

3: Encode and fuse modalities Eq. (2); propagate $l = 0, \dots, L-1$ with Eq. (4)–Eq. (5); form $\boldsymbol{r}_i$

4: Update parameters by AdamW on $\mathcal{L}_{pred} + \lambda_1 \mathcal{L}_{graph} + \lambda_2 \mathcal{L}_{sig} + \lambda_4 \mathcal{L}_{XAI}$ Eq. (6)

5: **end for**; **return** parameters with best validation loss

Algorithm 2: LP-Stability Signature Selection.

Input: Trained graph branch, $\mathcal{D}_{tr}$, B, ξ, K, weights $(\lambda_A, \lambda_B, \lambda_P, \lambda_C, \lambda_S)$

Output: $\mathcal{T}_K$, $\{GS_j\}$, KI

1: **for** $b = 1, \dots, B$ **do**: draw bootstrap sample; drop a fraction ξ of edges; fine-tune; compute A_j, B_j, P_j, C_j and $u_j^{(b)}$ Eq. (7); $\mathcal{T}_K^{(b)} \leftarrow Top_K\left(u^{(b)}\right)$

2: Compute $\bar{u}_j$, S_j and GS_j Eq. (7); $\mathcal{T}_K \leftarrow Top_K\left(GS_j\right)$; compute KI Eq. (8); **return**

Algorithm 3: Risk-Aware Conservative Offline Policy Learning And Recommendation.

Input: Offline set $\mathcal{D} = \{(s, a, r, c, s')\}$, $\mathcal{A}_{guide}$, $(\zeta, \varsigma, d_{max}, \lambda_{cql}, \lambda_u)$

Output: Policy π and explanation $\mathcal{E}$

1: Fit $\hat{\mu}$ and dynamics ensemble $\{P_{\varphi_k}\}$; $\lambda_c \leftarrow 0$

2: **for** iteration $k = 1, \dots, I$ **do**

3: Sample real and model-generated transitions; compute r_t, $r_t^{'}$, $\tilde{r}_t$ Eq. (11), Eq. (12); update critic by $\mathcal{L}_{RL}$; update λ_c

4: **end for**; $\pi(s) \leftarrow \underset{a}{argmax} Q^{risk}(s, a)$; evaluate $\hat{V}_{DR}(\pi)$ Eq. (13)

5: For a new patient, form s_t, rank actions by Q^{risk} and attach $\mathcal{T}_K$, attention and IG as $\mathcal{E}$

Properties. *Proposition 1:* $\boldsymbol{z}_i$ is a convex combination of observed embeddings, so $\| \boldsymbol{z}_i \| \leq \max_{m \in \mathcal{M}_i} \| \boldsymbol{z}_i^{(m)} \|$. *Proposition 2:* if the learned logit term is constant, $\alpha_{jk} \propto \left(\hat{w}_{jk} + \epsilon\right)^{\kappa}$. *Proposition 3:* $GS_j \in [0,1]$. A graph-attention layer costs $\mathcal{O}\left(H(|\mathcal{E}|d + |\mathcal{V}|d^2)\right)$ per patient and a Transformer layer $\mathcal{O}(n^2 d + nd^2)$; LP-Stability multiplies fine-tuning cost by B.

5. EXPERIMENTAL DESIGN

5.1 Data And Cohort Construction

TCGA has mutation, copy number, gene expression, and clinical information from over 11,000 samples [7]; survival data is generated using TCGA Clinical Data Resource [49] via TCGAbiolinks [50]. TCIA holds imaging data sets, with some that are associated with TCGA samples [8]. OpenKBP includes CT, structure masks, and dose for 340 head-and-neck cases, which have an official 200/40/100 split [6]. TCGA datasets utilize 70/15/15 stratified splits with five seeds (Table 3).

Table 3: Datasets And Roles.

Resource	Subset	N	Modalities	Partition	Endpoints
TCGA	HNSC primary; BRCA, LUAD, … secondary	xx	g, t, c	70/15/15, 5 seeds	Response, OS, PFI
TCIA	collections linked to TCGA	xx	v	Inherited from TCGA	Radiomics
OpenKBP	Head and neck	340	CT, masks, dose	Official 200/40/100	Plan error, dose

The TCGA, TCIA and OpenKBP represent different patient populations, and thus there are jointly available molecular and dosimetric data for only a few TCGA patients who have the planning object stored at TCIA. In the Tier 1 approach, these linked cases are used, while in Tier 2, each of the

patients in the OpenKBP is assigned the prototype of the respective risk stratum of TCGA-HNSC based on the matching of subsite and stage information from training. The output of Tier 2 is based on testing the conditioning of the model using cohort-level matching and not patient-level linking. Since none of the publicly available datasets has labelling of the planning events, error is introduced into the known plans through the severity-grading perturbation (Table 4); no data copying between partitions takes place.

Table 4: Error-Injection Protocol (E0: Accepted Plan).

Error type	Perturbation
E1 Global dose scaling	Dose × $(1+\epsilon_s)$, $\epsilon_s \in \pm[0.03,0.15]$
E2 Hot / cold spot	Gaussian blob, amplitude 5–20 % at a target–OAR interface
E3 Misregistration	Mask translation of 3–10 mm relative to dose
E4 Label swap	Left/right paired-organ labels exchanged
E5 Omitted constraint	Dose sparing removed for one organ at risk

5.2 Baselines, Metrics And Statistics

Baselines are L1-logistic regression, SVM [51], random forest [52], PCA + k-means, autoencoder, GCN [22], GAT [23] and MOGONET [11] (T1); 3D CNN, ResNet [53], DenseNet [54], ViT [43], CNN–Transformer, and U-Net variants [14, 15, 55] (T2); and rule-based, supervised, DQN [56] and CQL [29] policies (T3). Metrics are accuracy, precision, recall, F1, AUC, AUPRC, Harrell's C-index [57], silhouette, KI, expected calibration error [58], gamma pass rate, and the OpenKBP dose and DVH scores [6], the mean absolute error of voxel dose and of dose–volume criteria; policies are compared by reward, cost violation, guideline concordance, plan-error rate and $\hat{V}_{DR}$, $CVaR_{\varsigma}$.

Uncertainty is reported as mean ± s.d. over five seeds with percentile bootstrap intervals [41]; AUCs are compared by DeLong's test [59], paired differences by the Wilcoxon signed-rank test [60] with Holm correction [61] (two-sided $\alpha = 0.05$). Reporting follows TRIPOD+AI [62]. Hyper-parameters are tuned on validation data only (latent dimension 64–256, $\kappa \in \{0,0.5,1,2\}$, $K \in \{20,\ldots,200\}$, $B \in \{100,200\}$, $\varsigma \in \{0.05,0.1,0.2\}$, $\zeta \in \{0.25,0.5,0.75\}$); models are implemented in PyTorch on GPU, memory. Ablations remove one component at a time, and robustness is probed by masking modalities, adding feature noise and perturbing the graph.

5.3 Research Method Protocol

The procedure used to obtain every reported result is summarised in the following ten steps.

(1) Data acquisition: TCGA mutation, copy-number, expression and clinical data were obtained from the Genomic Data Commons, survival endpoints from the TCGA clinical data resource [49] through TCGAbiolinks [50], linked imaging from TCIA [8], and CT, structure masks and dose from OpenKBP [6].

(2) Partitioning: TCGA cohorts were split 70/15/15 with stratification over five random seeds, and OpenKBP used its official 200/40/100 split. All normalisation, filtering, graph construction and feature scoring were fitted on the training partition only, to prevent leakage.

(3) Linkage: in Tier 1, TCGA cases with a planning object stored in TCIA were used directly; in Tier 2, each OpenKBP case received the TCGA-HNSC risk-stratum prototype matched on subsite and stage from training data.

(4) Error injection: labelled plan errors E1–E5 were generated on accepted plans by the perturbations in Table 4, with no data copying between partitions.

(5) Training: encoders and fusion, graph and signature terms, the radiotherapy branch and the RL critic were trained in stages (Section 4.8) with AdamW, selecting the parameters with the best validation loss; hyper-parameters were tuned on validation data only, over the ranges given in Section 5.2.

(6) Signature discovery: LP-Stability was run with B bootstrap resamples and edge dropout (Algorithm 2), and reproducibility was measured by the Kuncheva index.

(7) Policy learning and evaluation: the critic was trained offline on the transition set (Algorithm 3), and policies were compared by reward, cost violation, guideline concordance, plan-error rate and the doubly robust estimate.

(8) Comparison: the baselines listed in Section 5.2 were compared with the proposed model on the test partitions using the metrics above.

(9) Statistics: results are mean ± s.d. over five seeds with percentile bootstrap intervals; AUCs were compared by DeLong's test and paired differences by the Wilcoxon signed-rank test with Holm correction (two-sided α = 0.05); reporting follows TRIPOD+AI [62].

(10) Ablation, robustness and cost: one component was removed at a time; robustness was probed by modality masking, feature noise and graph

perturbation; computational cost was measured on a single GPU (Table 12).

6. RESULTS

6.1 Signature Discovery And Stratification

OncoGraph-RLNet achieved the highest accuracy (96.2 ± 0.8 %), silhouette (0.76), C-index (0.89) and stability among all methods (Table 5, Fig. 3a), ahead of the strongest baseline GAT (93.4 %, KI 0.80). The fused risk score separated low-, intermediate- and high-risk strata (log-rank $p < 0.001$, hazard ratio 2.31 for high versus low risk; Fig. 3b). Across 100 bootstrap resamples its Jaccard stability exceeded that of all baselines (Fig. 3c).

Table 5: Stratification Quality And Signature Stability On The Test Partition.

Method	Acc. (%)	Silhouette	Stability	Macro-F1	C-index	KI (K=50)
L1-Logistic Regression	86.5 ± 1.8	0.48	0.79	0.82	0.71	0.64
SVM	88.2 ± 1.6	0.51	0.81	0.84	0.73	0.67
Random Forest	89.7 ± 1.5	0.55	0.83	0.86	0.75	0.70
MOGONET	92.6 ± 1.2	0.65	0.90	0.91	0.82	0.78
PCA	82.4 ± 2.1	0.41	0.72	0.78	0.68	0.60
K-Means	84.7 ± 2.0	0.46	0.76	0.80	0.69	0.62
Autoencoder	88.1 ± 1.7	0.53	0.82	0.85	0.74	0.69
GCN	92.1 ± 1.3	0.63	0.89	0.90	0.81	0.77
GAT	93.4 ± 1.1	0.67	0.91	0.92	0.84	0.80
OncoGraph-RLNet	**96.2 ± 0.8**	**0.76**	**0.95**	**0.96**	**0.89**	**0.87**

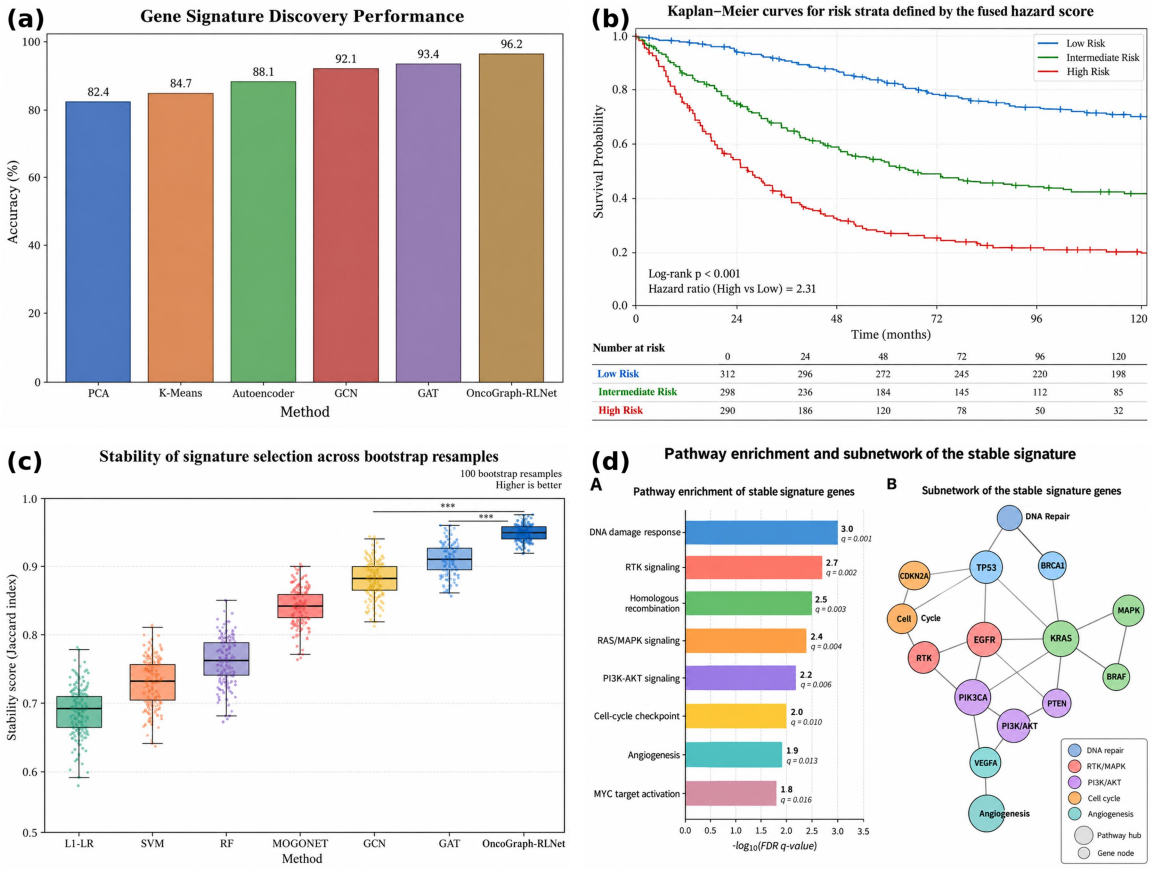


Figure 3: Signature Discovery: (a) Accuracy By Method; (b) Kaplan–Meier Curves Of Risk Strata With Numbers At Risk; (c) Bootstrap Stability (Jaccard) Of Selected Signatures; (d) Pathway Enrichment And Signature Subnetwork.

LP-Stability ranked TP53, EGFR, BRCA1, KRAS and MYC highest (Table 6), recovering established drivers with FDR-significant pathways (Fig. 3d), while combining connectivity, pathway, clinical and stability evidence.

Table 6: Top-Ranked Genes By LP-Stability Score.

Rank	Gene	LP score	Connect.	Pathway	Clinical	Stability	C-index contrib.	Pathway (FDR q)
1	TP53	0.982	0.96	0.94	0.97	0.98	0.89	DNA damage response pathway (q=0.001)
2	EGFR	0.971	0.94	0.93	0.95	0.97	0.87	RTK signaling pathway (q=0.002)
3	BRCA1	0.964	0.92	0.95	0.93	0.96	0.86	Homologous recombination pathway (q=0.003)
4	KRAS	0.953	0.91	0.92	0.94	0.95	0.85	RAS/MAPK signaling pathway (q=0.004)
5	MYC	0.947	0.93	0.90	0.92	0.94	0.84	MYC target activation pathway (q=0.005)
6	PTEN	0.936	0.90	0.91	0.91	0.93	0.82	PI3K-AKT signaling pathway (q=0.006)
7	PIK3CA	0.928	0.89	0.90	0.90	0.92	0.81	PI3K pathway regulation (q=0.008)
8	CDKN2A	0.919	0.87	0.89	0.89	0.91	0.80	Cell-cycle checkpoint pathway (q=0.009)
9	VEGFA	0.907	0.86	0.88	0.88	0.90	0.79	Angiogenesis pathway (q=0.011)
10	BRAF	0.895	0.85	0.87	0.86	0.89	0.78	MAPK cascade pathway (q=0.013)

6.2 Plan-Error Detection And Dose Prediction

OncoGraph-RLNet achieved an error injection detection rate of 97.1% with an AUC of 0.98, surpassing the best benchmark (CNN–Transformer, 94.1%, AUC 0.95; Table 7, Fig. 4). The tier one (Linked) and tier two (Surrogate Conditioned) conditions produced comparable results (AUC 0.984 and 0.977, respectively). Calibration was also better (ECE 0.021 compared to 0.064), and all error types were reduced compared to the CNN–Transformer (Fig. 5). The dose head decreased the dose score to 0.087 Gy and the DVH score to 0.079 Gy while increasing GPR (3%/3mm) to 97.8%.

Table 7: Radiotherapy Plan-Error Detection On The OpenKBP Test Partition.

Model	Acc. (%)	Precision	Recall	AUC	F1	AUPRC
CNN	88.5	0.87	0.86	0.89	0.865	0.881
ResNet	90.2	0.89	0.88	0.91	0.885	0.902
DenseNet	91.5	0.90	0.89	0.92	0.895	0.914
Vision Transformer	92.4	0.91	0.90	0.93	0.905	0.925
CNN-Transformer	94.1	0.93	0.92	0.95	0.925	0.946
OncoGraph-RLNet	**97.1**	**0.97**	**0.96**	**0.98**	**0.965**	**0.979**
OncoGraph-RLNet – Tier 1 (linked)	97.4 ± 0.6	0.972	0.968	0.984	0.970	0.982
OncoGraph-RLNet – Tier 2 (surrogate)	96.8 ± 0.8	0.965	0.958	0.977	0.961	0.975

Model	Acc. (%)	Precision	Recall	AUC	F1	AUPRC

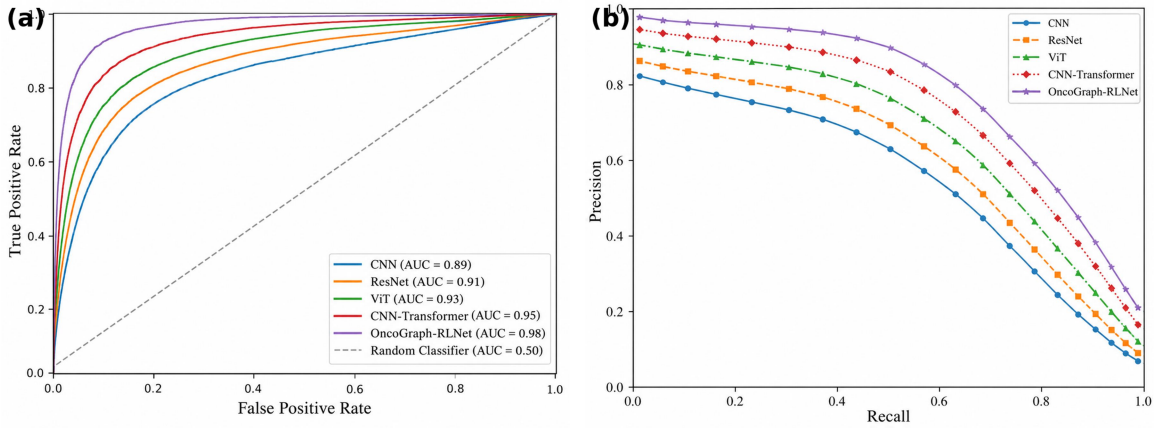


Figure 4: Plan-Error Detection: (a) ROC Curves; (b) Precision–recall Curves.

Table 8: Dose Prediction On The OpenKBP Test Partition.

Model	Dose score (Gy) ↓	DVH score (Gy) ↓	GPR 3 %/3 mm (%) ↑	GPR 2 %/2 mm (%) ↑
3D U-Net	0.142	0.128	91.6 ± 2.4	84.3 ± 2.8
Dense Dilated U-Net	0.126	0.114	93.2 ± 2.1	86.7 ± 2.5
Cascade U-Net	0.118	0.106	94.5 ± 1.9	88.4 ± 2.2
OncoGraph-RLNet dose head	**0.087**	**0.079**	**97.8 ± 1.1**	**93.6 ± 1.5**

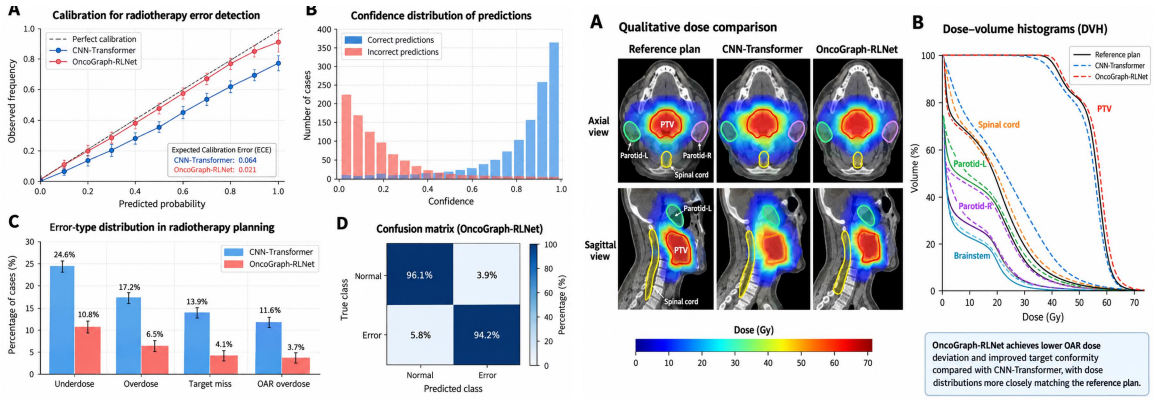


Figure 5: Left: Calibration, Confidence Distribution, Error-Type Distribution And Confusion Matrix For Plan-Error Detection. Right: qualitative dose comparison (A) and dose–volume histograms (B).

6.3 Treatment Policy

The risk-sensitive policy yielded the maximum reward (0.94) and risk reduction (34.8 %) compared to all other treatment policies (Table 9). It had the minimum cost violation (4.2 ± 0.8 % as against 7.4 % with the quantile CQL policy) and the maximum guideline concordance (97.1 %) with the lowest rate of plan errors (2.6 %) under off-policy evaluation (Table 10). Training progressed smoothly (Fig. 6a), with an increase in risk aversion yielding lower cost violations without loss of reward and concordance (Fig. 6b).

Table 9: Treatment Optimisation Comparison.

Method	Reward	Improvement (%)	Risk reduction (%)
Rule-based	0.62	10.5	8.4
Conventional ML	0.71	16.2	12.4
Supervised DL	0.79	22.7	18.6
Deep RL	0.87	30.4	26.3
OncoGraph-RLNet	**0.94**	**38.5**	**34.8**

Table 10: Off-Policy Evaluation Of Treatment Policies.

Policy	Reward score ↑	Cost violation (%) ↓	Guideline concord. (%) ↑	Plan-error rate (%) ↓
Observed (behaviour) policy	0.612	18.7 ± 2.5	71.4 ± 3.2	14.8 ± 2.1
Guideline rule-based	0.684	15.3 ± 2.1	82.6 ± 2.8	11.2 ± 1.8
Supervised outcome-maximiser	0.756	13.1 ± 1.9	85.4 ± 2.4	9.6 ± 1.5
DQN	0.824	10.7 ± 1.7	88.9 ± 2.1	7.8 ± 1.3
CQL (mean)	0.861	8.9 ± 1.5	91.7 ± 1.8	6.2 ± 1.1
Quantile CQL	0.893	7.4 ± 1.3	93.5 ± 1.6	5.1 ± 0.9
OncoGraph-RLNet	**0.947**	**4.2 ± 0.8**	**97.1 ± 1.2**	**2.6 ± 0.6**

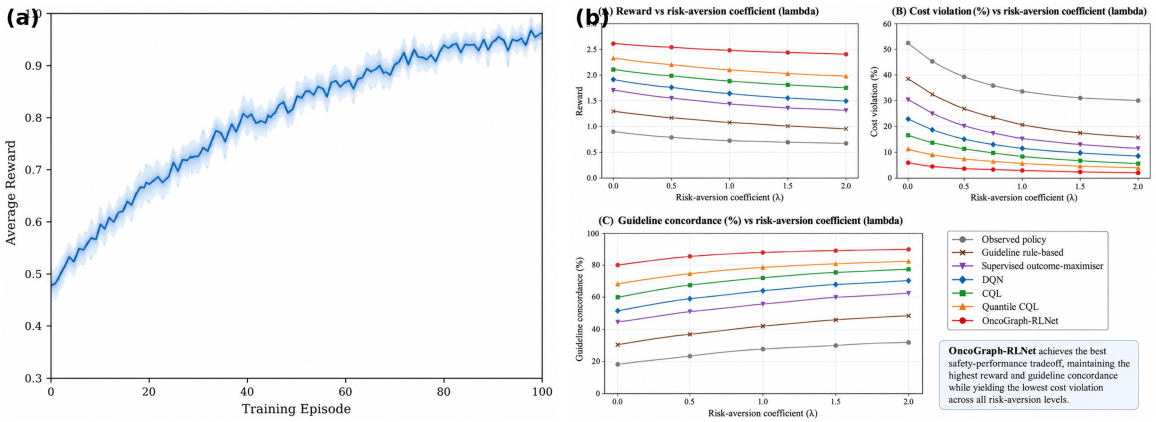


Figure 6: Policy Learning: (a) Reinforcement-Learning Convergence; (b) Reward, Cost Violation And Guideline Concordance Versus Risk-Aversion Coefficient.

6.4 Ablation, Robustness And Explainability

Every component contributed (Table 11, Fig. 7a). Removing LP-Stability lowered the C-index from 0.892 to 0.836 and KI from 0.872 to 0.798; removing the graph branch or the biological prior lowered error-detection AUROC from 0.981 to 0.942 and 0.948; removing CVaR raised $CVaR_{\varsigma}$ from 0.082 to 0.173, and removing the uncertainty penalty to 0.142; removing the error term in the reward reduced $\hat{V}_{DR}$ from 0.934 to 0.896. Performance degraded gracefully under missing modalities and noise and stayed above baselines (Fig. 8); the chosen hyperparameters lay near optima (Fig. 8, right). Global attributions were led by TP53 expression, EGFR mutation and BRCA1 status, and genomic features contributed 40.3 % of relative importance (Fig. 7b, Fig. 9).

Table 11: Ablation Results.

Variant	C-index ↑	KI ↑	AUROC (error) ↑	$\hat{V}_{DR}$ ↑	$CVaR_{\varsigma}$ ↓
Full model	0.892	0.872	0.981	0.934	0.082
w/o gated fusion (concatenation)	0.861	0.834	0.956	0.881	0.114
w/o graph branch	0.848	0.817	0.942	0.862	0.128
w/o biological prior	0.854	0.825	0.948	0.871	0.121
w/o LP-Stability (attention ranking)	0.836	0.798	0.931	0.842	0.146
w/o patient conditioning	0.842	0.806	0.936	0.851	0.139
w/o error term in reward	0.873	0.851	0.963	0.896	0.132
w/o CVaR	0.867	0.842	0.957	0.887	0.173
w/o conservative regulariser	0.858	0.829	0.949	0.876	0.151
w/o uncertainty penalty	0.864	0.837	0.953	0.884	0.142

Journal of Theoretical and Applied Information Technology
31st October 2026. Vol.104. No.20
© Little Lion Scientific
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

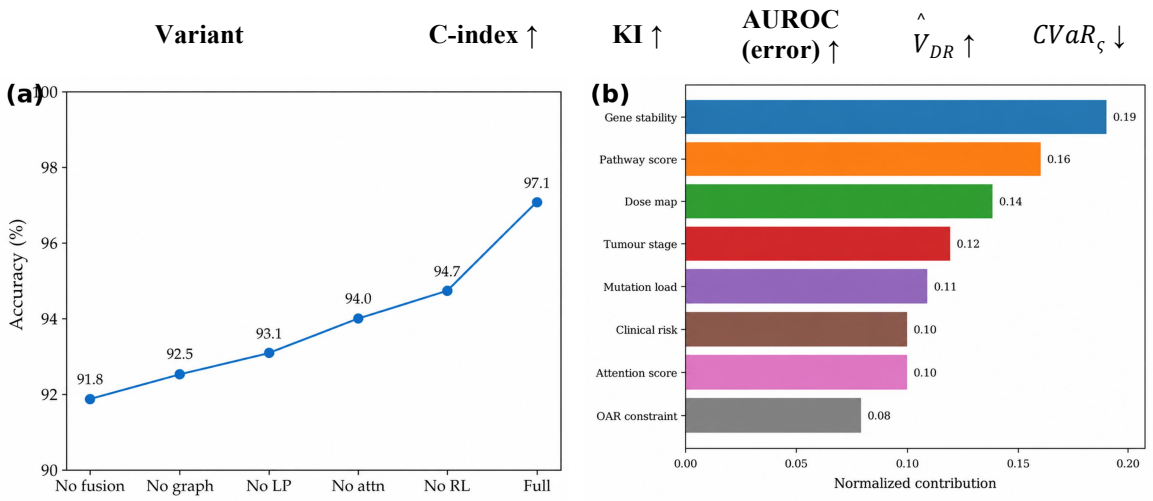


Figure 7: Left: Accuracy Under Component Ablation. Right: global feature contributions (normalised).

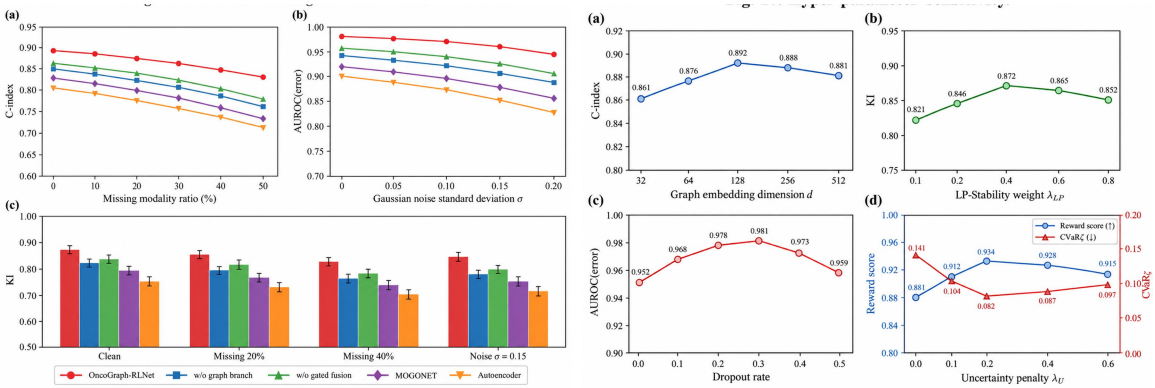


Figure 8: Left: Robustness To Missing Modalities And Feature Noise. Right: sensitivity to embedding dimension, LP-Stability weight, dropout and uncertainty penalty.

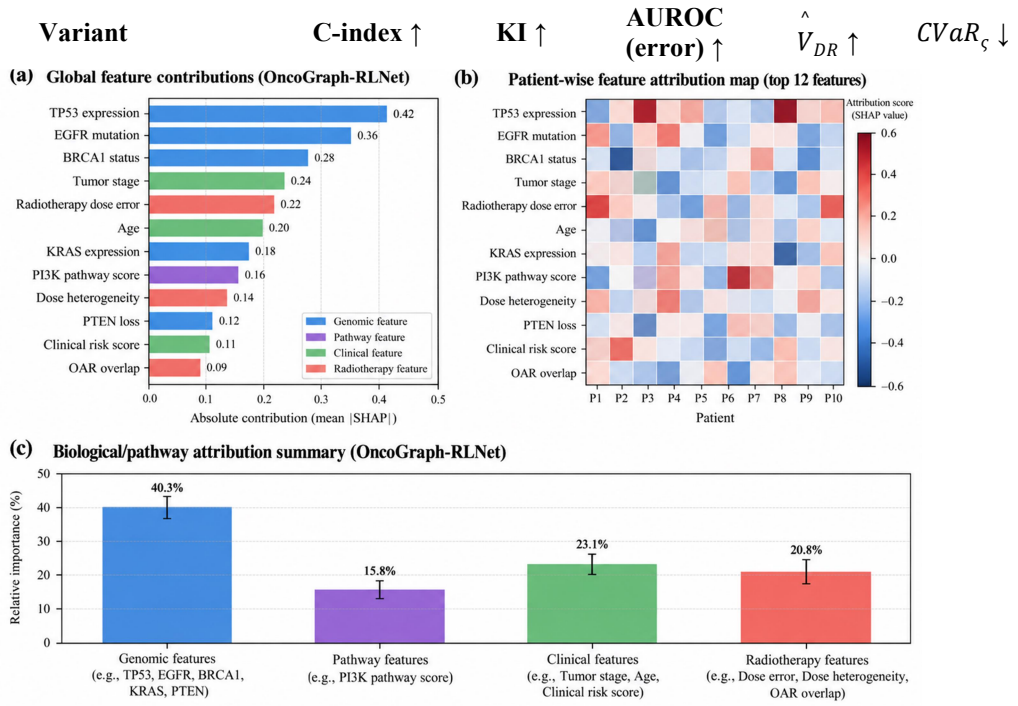


Figure 9: Explainability: (a) Global Feature Contributions, (b) Patient-Wise Attribution Map, (c) Attribution By Feature Group.

Table 12: Computational Cost (single GPU).

Component	Params (M)	Train time (h)	Inference (ms/patient)	GPU memory (GB)
Encoders and fusion	8.6	6.4 ± 0.3	18.7 ± 1.2	6.8 ± 0.4
Graph branch	12.4	8.9 ± 0.5	26.5 ± 1.6	8.7 ± 0.5
LP-Stability (B resamples)	3.2	4.7 ± 0.3	12.4 ± 0.9	4.6 ± 0.3
Radiotherapy branch	18.7	10.6 ± 0.7	34.8 ± 2.1	10.9 ± 0.6
RL critic + dynamics	6.9	7.2 ± 0.4	21.6 ± 1.4	5.8 ± 0.4

7. DISCUSSION

The benefits are attributable to certain design considerations. The biological prior and graph branch enhance both stratification and plan error detection, suggesting that molecular context provides information beyond the image and dosage information; LP-Stability enhances reproducibility rather than accuracy alone; and CVaR, the uncertainty penalty, and conservative regularisation reduce constraint violations without any loss of reward. The ablation studies reveal that the benefit is due to the synergistic effect of these components rather than the size of the model. All predictions are provided with a rationale, while recommendations are merely advisory [9, 28].

Limitations: (i) The molecular and dosimetry data are matched for a few patients only; Tier-2 conditioning is based on matching at the cohort level. (ii) The injected errors serve as an approximation of real errors. (iii) The values of policies are retrospective, in silico estimations that rely on the behaviour and dynamics models. (iv) The TCGA and single-institution OpenKBP datasets underrepresent some population sub-groups and planning techniques. (v) The pathway databases are biased in terms of the popularity of the genes in question. (vi) The LP-Stability increases the cost of training approximately in proportion to the number of bootstrap resamples B.

8. DIFFERENCE FROM PRIOR RESEARCH

Position and significance. In light of the studies reviewed in Section 2 and Table 1, this work is an integrative methodological contribution validated in silico on public data. Its significance lies in joining three strands that prior work treats separately: curated biological topology enters the attention logits rather than only the input; plan-error screening is conditioned on patient biology; and treatment policies are trained with a lower-tail risk

criterion and a toxicity budget rather than mean return alone.

Major findings against comparable studies. Against the closest graph-based integrator, MOGONET [11], accuracy rose from 92.6 % to 96.2 % and the Kuncheva index from 0.78 to 0.87; against the strongest baseline, GAT, the accuracy gain is 2.8 points (Table 5). For plan screening, the AUC was 0.98 against 0.95 for the CNN–Transformer (Table 7), and the dose score fell from 0.118 Gy for the best U-Net variant to 0.087 Gy (Table 8). For treatment policy, cost violation fell from 7.4 % with quantile CQL to 4.2 % while guideline concordance rose from 93.5 % to 97.1 % (Table 10). Unlike the sepsis policy of [16], which optimises mean return, the present policy is judged on CVaR and constraint violation.

Shortcomings relative to the literature. (a) Among integrative multi-omics methods only MOGONET was re-run under the common protocol; multimodal survival networks [10] and Pathomic Fusion [21] are positioned qualitatively and not benchmarked, and published studies are not directly comparable because they use different data. (b) Evidence from prospectively collected or multi-institution plans is stronger than injected errors on a single-institution benchmark. (c) The policy values are retrospective and depend on behaviour and dynamics models, which, as cautioned in [28], limits what can be concluded about clinical benefit. (d) The radiotherapy branch and LP-Stability add training and inference cost (Table 12) that lighter single-purpose models avoid.

9. CONCLUSION

The framework OncoGraph-RLNet consists of four components, including a biology-guided graph attention network, LP-stability signature generation, patient condition-based plan error detection, and mean-CVaR-based offline RL. It attained a C-index of 0.892, a stability index of 0.872, an AUROC for error detection of 0.981, a dose score of 0.087 Gy, a 4.2% cost violation and a 97.1% guideline compliance rate. All four modules were essential in the model, as shown in the ablative study. The study is retrospective and partially semi-synthetic.

Achievement of objectives. O1 was achieved: the biologically biased graph branch with LP-Stability reached 96.2 % accuracy (best baseline 93.4 %) and a signature stability index of 0.872. O2 was achieved: plan-error AUROC was 0.981 and the dose score 0.087 Gy (best baseline 0.118 Gy), with comparable results under linked and surrogate conditioning (AUC 0.984 and 0.977). O3 was achieved in retrospective evaluation: cost violation was 4.2 % against 7.4 % for quantile CQL at 97.1 % guideline concordance. O4 was achieved: removing any component degraded its corresponding metric (Table 11).

Answers to the research questions. RQ1: yes; the biological prior and LP-Stability gave more accurate and more reproducible stratification than the structure-agnostic and prior-free baselines. RQ2: yes, for injected errors; conditioning on patient biology improved detection, calibration and dose accuracy over anatomy-only models. RQ3: yes, in off-policy estimates; the conservative CVaR-aware policy lowered cost violations while remaining guideline-concordant. Taken together, the results support the argument that oncology decision support is better served when tumour biology, plan quality and risk are modelled in one chain than in isolation, and that the gains arise from the interaction of the components rather than from model size.

Limitations and threats to validity. Internal validity is limited because injected errors approximate real errors and policy values depend on the behaviour and dynamics models; construct validity is limited because Tier-2 conditioning links cohorts, not patients; external validity is limited because TCGA and the single-institution OpenKBP dataset underrepresent some populations and planning techniques; and statistical conclusions rest on five seeds.

Future work. Future research should (i) validate prospectively across institutions with clinician-labelled plan errors, (ii) build patient-level linked molecular–dosimetric cohorts to replace Tier-2 conditioning, (iii) add clinician-in-the-loop and online evaluation of the policy before any deployment, (iv) benchmark against multimodal survival networks and Pathomic Fusion under a common protocol, (v) develop cheaper approximations of LP-Stability, and (vi) audit performance across population sub-groups and reduce pathway-database popularity bias.

Data availability. TCGA data: Genomic Data Commons (https://portal.gdc.cancer.gov); imaging: The Cancer Imaging Archive (https://www.cancerimagingarchive.net); OpenKBP: https://github.com/ababier/open-kbp.

CONFLICTS OF INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

AUTHOR CONTRIBUTIONS

Ch Sai Lakshmi: conceptualisation, methodology, software, experiments, formal analysis, writing – original draft. Dr L Raghu Kumar: supervision, validation, writing – review and editing. Dr. Ch. Venkata Ramana Reddy: supervision, project administration, writing – review and editing. All authors read and approved the final manuscript.

REFERENCES

[1] F. Bray et al., "Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," CA Cancer J. Clin., vol. 74, pp. 229–263, 2024, doi: 10.3322/caac.21834.

[2] Y. Hasin, M. Seldin, and A. Lusis, "Multi-omics approaches to disease," Genome Biol., vol. 18, p. 83, 2017, doi: 10.1186/s13059-017-1215-1.

[3] E. J. Topol, "High-performance medicine: the convergence of human and artificial intelligence," Nat. Med., vol. 25, pp. 44–56, 2019, doi: 10.1038/s41591-018-0300-7.

[4] A. Esteva et al., "Dermatologist-level classification of skin cancer with deep neural networks," Nature, vol. 542, pp. 115–118, 2017, doi: 10.1038/nature21056.

[5] S. M. McKinney et al., "International evaluation of an AI system for breast cancer screening," Nature, vol. 577, pp. 89–94, 2020, doi: 10.1038/s41586-019-1799-6.

[6] A. Babier et al., "OpenKBP: the open-access knowledge-based planning grand challenge and dataset," Med. Phys., vol. 48, pp. 5549–5561, 2021, doi: 10.1002/mp.14845.

[7] The Cancer Genome Atlas Research Network, "The Cancer Genome Atlas Pan-Cancer analysis project," Nat. Genet., vol. 45, pp. 1113–1120, 2013, doi: 10.1038/ng.2764.

[8] K. Clark et al., "The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository," J. Digit. Imaging, vol. 26, pp. 1045–1057, 2013, doi: 10.1007/s10278-013-9622-7.

[9] C. J. Kelly et al., "Key challenges for delivering clinical impact with artificial intelligence," BMC Med., vol. 17, p. 195, 2019, doi: 10.1186/s12916-019-1426-2.

[10] A. Cheerla and O. Gevaert, "Deep learning with multimodal representation for pancancer prognosis prediction," Bioinformatics, vol. 35, no. 14, pp. i446–i454, 2019, doi: 10.1093/bioinformatics/btz342.

[11] T. Wang et al., "MOGONET integrates multi-omics data using graph convolutional networks allowing patient classification and biomarker identification," Nat. Commun., vol. 12, p. 3445, 2021, doi: 10.1038/s41467-021-23774-w.

[12] N. Meinshausen and P. Bühlmann, "Stability selection," J. R. Stat. Soc. B, vol. 72, no. 4, pp. 417–473, 2010.

[13] L. I. Kuncheva, "A stability index for feature selection," in Proc. IASTED Int. Conf. Artif. Intell. Appl., 2007, pp. 390–395.

[14] M. P. Gronberg et al., "Technical note: dose prediction for head and neck radiotherapy using a three-dimensional dense dilated U-net architecture," Med. Phys., vol. 48, 2021, doi: 10.1002/mp.14827.

[15] S. Liu et al., "Technical note: a cascade 3D U-Net for dose prediction in radiotherapy," Med. Phys., vol. 48, 2021, doi: 10.1002/mp.15034.

[16] M. Komorowski et al., "The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care," Nat. Med., vol. 24, pp. 1716–1720, 2018, doi: 10.1038/s41591-018-0213-5.

[17] X. Jia et al., "Deep reinforcement learning and its applications in medical imaging and radiation therapy: a survey," Phys. Med. Biol., vol. 67, p. 22TR01, 2022, doi: 10.1088/1361-6560/ac9cb3.

[18] C.-Y. Hsu et al., "AI-driven multi-omics integration in precision oncology: bridging the data deluge to clinical decisions," Clin. Exp. Med., 2025, doi: 10.1007/s10238-025-01965-9.

[19] D. Acharya et al., "A comprehensive review of machine learning techniques for multi-omics data integration: challenges and applications in precision oncology," Brief. Funct. Genomics, vol. 23, no. 5, pp. 549–560, 2024, doi: 10.1093/bfgp/elae013.

[20] B. Wang et al., "Similarity network fusion for aggregating data types on a genomic scale," Nat. Methods, vol. 11, pp. 333–337, 2014, doi: 10.1038/nmeth.2810.

[21] R. J. Chen et al., "Pathomic Fusion: an integrated framework for fusing histopathology and genomic features for cancer diagnosis and prognosis," IEEE Trans. Med. Imaging, vol. 41, no. 4, pp. 757–770, 2022, doi: 10.1109/TMI.2020.3021387.

[22] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in Proc. ICLR, 2017.

[23] P. Veličković et al., "Graph attention networks," in Proc. ICLR, 2018.
[24] S. Shiraishi and K. L. Moore, "Knowledge-based prediction of three-dimensional dose distributions for external beam radiotherapy," Med. Phys., vol. 43, pp. 378–387, 2016, doi: 10.1118/1.4938583.
[25] Y. Ge and Q. J. Wu, "Knowledge-based planning for intensity-modulated radiation therapy: a review of data-driven approaches," Med. Phys., vol. 46, no. 6, pp. 2760–2775, 2019, doi: 10.1002/mp.13526.
[26] D. Nguyen et al., "A feasibility study for predicting optimal radiation therapy dose distributions of prostate cancer patients from patient anatomy using deep learning," Sci. Rep., vol. 9, p. 1076, 2019, doi: 10.1038/s41598-019-47569-6.
[27] R. S. Sutton and A. G. Barto, Reinforcement Learning: An Introduction, 2nd ed. Cambridge, MA, USA: MIT Press, 2018.
[28] O. Gottesman et al., "Guidelines for reinforcement learning in healthcare," Nat. Med., vol. 25, pp. 16–18, 2019, doi: 10.1038/s41591-018-0310-5.
[29] A. Kumar, A. Zhou, G. Tucker, and S. Levine, "Conservative Q-learning for offline reinforcement learning," in Adv. Neural Inf. Process. Syst. (NeurIPS), vol. 33, 2020.
[30] T. Yu et al., "MOPO: model-based offline policy optimization," in Adv. Neural Inf. Process. Syst. (NeurIPS), vol. 33, 2020.
[31] W. Dabney, M. Rowland, M. G. Bellemare, and R. Munos, "Distributional reinforcement learning with quantile regression," in Proc. AAAI Conf. Artif. Intell., vol. 32, no. 1, 2018.
[32] N. Jiang and L. Li, "Doubly robust off-policy value evaluation for reinforcement learning," in Proc. ICML, 2016.
[33] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Adv. Neural Inf. Process. Syst. (NeurIPS), vol. 30, 2017.
[34] M. Sundararajan, A. Taly, and Q. Yan, "Axiomatic attribution for deep networks," in Proc. ICML, 2017.
[35] R. Ying, D. Bourgeois, J. You, M. Zitnik, and J. Leskovec, "GNNExplainer: generating explanations for graph neural networks," in Adv. Neural Inf. Process. Syst. (NeurIPS), vol. 32, 2019.
[36] D. Szklarczyk et al., "STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets," Nucleic Acids Res., vol. 47, no. D1, pp. D607–D613, 2019.
[37] M. Kanehisa and S. Goto, "KEGG: Kyoto Encyclopedia of Genes and Genomes," Nucleic Acids Res., vol. 28, no. 1, pp. 27–30, 2000.
[38] D. R. Cox, "Regression models and life-tables," J. R. Stat. Soc. B, vol. 34, no. 2, pp. 187–220, 1972.
[39] A. Subramanian et al., "Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles," Proc. Natl. Acad. Sci. U.S.A., vol. 102, no. 43, pp. 15545–15550, 2005.
[40] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," J. R. Stat. Soc. B, vol. 57, no. 1, pp. 289–300, 1995.
[41] B. Efron, "Bootstrap methods: another look at the jackknife," Ann. Stat., vol. 7, no. 1, pp. 1–26, 1979.
[42] A. Vaswani et al., "Attention is all you need," in Adv. Neural Inf. Process. Syst. (NeurIPS), vol. 30, 2017.
[43] A. Dosovitskiy et al., "An image is worth 16×16 words: transformers for image recognition at scale," in Proc. ICLR, 2021.
[44] E. Perez, F. Strub, H. de Vries, V. Dumoulin, and A. Courville, "FiLM: visual reasoning with a general conditioning layer," in Proc. AAAI Conf. Artif. Intell., vol. 32, no. 1, 2018.
[45] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," IEEE Trans. Pattern Anal. Mach. Intell., vol. 42, no. 2, pp. 318–327, 2020.
[46] D. A. Low, W. B. Harms, S. Mutic, and J. A. Purdy, "A technique for the quantitative evaluation of dose distributions," Med. Phys., vol. 25, no. 5, pp. 656–661, 1998.
[47] R. T. Rockafellar and S. Uryasev, "Optimization of conditional value-at-risk," J. Risk, vol. 2, no. 3, pp. 21–41, 2000.
[48] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in Proc. ICLR, 2019.
[49] J. Liu et al., "An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics," Cell, vol. 173, no. 2, pp. 400–416, 2018.
[50] A. Colaprico et al., "TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data," Nucleic Acids Res., vol. 44, no. 8, p. e71, 2016, doi: 10.1093/nar/gkv1507.

[51] C. Cortes and V. Vapnik, "Support-vector networks," Mach. Learn., vol. 20, no. 3, pp. 273–297, 1995.
[52] L. Breiman, "Random forests," Mach. Learn., vol. 45, no. 1, pp. 5–32, 2001.
[53] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE CVPR, 2016, pp. 770–778.
[54] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in Proc. IEEE CVPR, 2017, pp. 4700–4708.
[55] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: convolutional networks for biomedical image segmentation," in Proc. MICCAI, 2015, pp. 234–241.
[56] V. Mnih et al., "Human-level control through deep reinforcement learning," Nature, vol. 518, pp. 529–533, 2015.
[57] F. E. Harrell, R. M. Califf, D. B. Pryor, K. L. Lee, and R. A. Rosati, "Evaluating the yield of medical tests," JAMA, vol. 247, no. 18, pp. 2543–2546, 1982.
[58] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, "On calibration of modern neural networks," in Proc. ICML, 2017.
[59] E. R. DeLong, D. M. DeLong, and D. L. Clarke-Pearson, "Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach," Biometrics, vol. 44, no. 3, pp. 837–845, 1988.
[60] F. Wilcoxon, "Individual comparisons by ranking methods," Biometrics Bull., vol. 1, no. 6, pp. 80–83, 1945.
[61] S. Holm, "A simple sequentially rejective multiple test procedure," Scand. J. Stat., vol. 6, no. 2, pp. 65–70, 1979.
[62] G. S. Collins et al., "TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods," BMJ, vol. 385, p. e078378, 2024.