SynLeaF: A Dual-Stage Multimodal Fusion Framework for Synthetic Lethality Prediction Across Pan- and Single-Cancer Contexts

Zheming Xing[1], Siyuan Zhou[1], Ruinan Wang[1], Rui Han[1], Shiming Zhang[1], Shiqu Chen[1], Yurui Huang[1], Jiahao Ma[3], Yifan Chen[4], Xuan Wang[1], Yadong Wang[2], Junyi Li[1,2*]

[1*]School of Computer Science and Technology, Harbin Institute of Technology (Shenzhen), Shenzhen, Guang Dong 518055, China.
[2]Key Laboratory of Biological Bigdata, Ministry of Education, Harbin Institute of Technology, Harbin, Heilongjiang 150001, China.
[3]School of Biomedical Sciences, The University of Hong Kong, Hong Kong SAR, China.
[4]Departments of Mathematics and Computer Science, Hong Kong Baptist University, Hong Kong SAR, China.

*Correspondence should be addressed to lijunyi@hit.edu.cn

Abstract

Accurate prediction of synthetic lethality (SL) is important for guiding the development of cancer drugs and therapies. SL prediction faces significant challenges in effectively fusing heterogeneous multi-source data. Existing multimodal methods often suffer from "modality laziness" due to disparate convergence speeds, which hinders the exploitation of complementary information and causes most SL prediction models to perform poorly for both pan-cancer and single-cancer SL pair predictions. In this study, we propose SynLeaF, a dual-stage multimodal fusion framework for SL prediction in pan-cancer and single-cancer contexts. The framework employs a VAE-based cross-encoder with a Product of Experts mechanism to fuse four omics data types (gene expression, mutation, methylation, and CNV), simultaneously utilizing a relational graph convolutional network to capture structured gene representations from biomedical knowledge graphs. To mitigate modality laziness, SynLeaF introduces a dual-stage training mechanism that employs a feature-level knowledge distillation. In extensive experiments across eight specific cancer types and a pan-cancer dataset, SynLeaF achieved superior

1

performance in 17 of the 19 scenarios. Ablation studies and gradient analyses further validate the critical contributions of the proposed fusion and distillation mechanisms for model robustness and generalization. To facilitate community use, a web server is available at https://synleaf.bioinformatics-lilab.cn.

1 Introduction

Synthetic Lethality (SL) characterizes a specific genetic relationship wherein the deficiency of an individual gene remains viable for the cell, whereas concurrent impairment or inactivation of a gene pair results in cell death [1]. As a promising targeted anticancer therapy, SL can eliminate malignant cells while preserving healthy tissues [2], and expands druggable targets for genes that are difficult to target directly [3]. A classic clinical success is the PARP inhibitors, approved by the FDA in 2014 for ovarian cancer with BRCA1/2 mutations [4, 5]. Recently, computationally designed aptamers were shown to induce SL by blocking the RAD51–BRCA2 interaction [6]. ADAR1 was also identified as a new SL target in BRCA-mutant tumors, where its inhibition activates innate immunity via autocrine interferon poisoning [7]. These advances continue to expand druggable targets and accelerate their translation into therapies [8].

Despite this potential, identifying clinically relevant SL pairs remains a challenge [9]. Wet-lab screens, including yeast, RNAi, and CRISPR, are accurate but costly, time-consuming, and suffer from significant off-target effects [10, 11]. The sheer volume of possible gene pairs makes exhaustive screening infeasible [12]. To overcome the limitations of wet-lab methods, computational methods, which can be grouped into statistics-based, network-based, traditional machine learning, and deep learning methods, have emerged as effective complements [3].

Statistics and network approaches for synthetic lethal prediction rely on hypotheses and domain knowledge and often ignore other data types such as sequences or functional attributes [13]. Traditional machine learning methods integrate multi-source data but still depend on feature engineering, which can introduce noise [14, 15]. To address these limitations, deep learning methods have been introduced to automatically learn complex representations. Deep learning methods learn complex representations automatically. Early network representation models such as GRSMF and SL2MF use matrix factorization [16, 17]. GRSMF introduces graph-regularized self-representation to alleviate data sparsity, and SL2MF uses logistic matrix factorization with different weights for known and unknown pairs. However, matrix factorization methods are essentially a form of shallow embedding and may not fully exploit the structural information and node features of the network [18].

The application of Graph Neural Networks (GNN) has significantly enhanced the accuracy of predicting synthetic lethality. DDGCN [19] uses a dual-dropout strategy to address sparsity and overfitting, and GCATSL [20] applies dual attention at the node and feature levels. However, these GNN-based methods mainly rely on

homogeneous networks, which only contain gene nodes and possess limited expressive power. The incorporation of Knowledge Graphs (KG) has further broadened the use of deep learning techniques for synthetic lethality prediction. Basically, a KG represents a heterogeneous network that contains multiple types of entities (such as genes, pathways, diseases, etc.) and relations, that can provide richer biological background information [21]. KG4SL [22] automatically generates gene features using a knowledge graph convolutional network. PiLSL [23] adopts a gene pair-based method, extracting enclosing subgraphs from the knowledge graph to capture pairwise interactions between gene pairs. SLGNN [24] models relation combinations as factors and improves the interpretability of the model through node- and factor-level attention mechanisms. KR4SL [25] introduces a reasoning method based on paths, utilizing relational digraphs to extract structural semantic information within the knowledge graph. Meanwhile, MPASL [26] proposed a hybrid interaction framework involving gene-entity and entity-entity relationships to improve gene representations from various viewpoints.

Although knowledge graphs provide rich structured biological knowledge, KGs solely from a single modality often cannot fully capture the complex mechanisms of synthetic lethality. Consequently, fusing data from multiple sources has emerged as a crucial approach to enhance the accuracy of synthetic lethality prediction. PTGNN [27] uses Convolutional Neural Network (CNN) features from protein sequences and a graph reconstruction pre-training task on Protein-Protein Interaction and Gene Ontology (GO) graphs. PiLSL [23] integrates explicit omics features with KG embeddings. TARSL [28] applies non-negative matrix tri-factorization with triple attention, and Struct2SL [29] combines protein sequences, protein 3D structures, and PPI networks. Although these methods improve prediction accuracy by introducing multi-source information, they face a common and critical limitation, as in pure KG methods, the neglect of *context-specificity*.

The realization of generalized synthetic lethal effects is often impeded by specific tumor-associated factors, including cellular heterogeneity, metabolic status, and the complexities of the tumor microenvironment [30]. Many synthetic lethal interactions are observed in only a few specific cancers. An investigation utilizing CRISPR-Cas9 screening to pinpoint synthetic lethal gene pairs across three cell lines revealed a minimal intersection: merely 10% of the interactions were shared between any two lines, while no single pair was consistently observed across all three [31, 32]. Adequately addressing tumor heterogeneity can mitigate certain challenges associated with the translation of synthetic lethality strategies from cellular models to *in vivo* systems and, ultimately, clinical applications [33]. Context-specific synthetic lethal effects have garnered significant interest within the medicinal chemistry community, as investigating cancer-specific synthetic lethality offers novel avenues for pharmaceutical development. Although several studies have attempted to address the specificity problem, they still exhibit significant shortcomings. While ELISL [34] pioneered the integration of context-free protein sequence associations with context-specific omics data, its reliance on shallow models, such as random forests, limits its ability to capture the non-linear relationships in high-dimensional data. Based on cancer-specific positive

3

synthetic lethality datasets, SLGNNCT [35] divides the knowledge graph into different subgraphs and performs cancer-specific synthetic lethality prediction on the small knowledge graphs separately via factor attention modeling, such as SLGNN. However, the resulting knowledge subgraphs have very few nodes, and the small graph size leads to poor generalization performance in deep learning models.

To address these obstacles, we propose SynLeaF, a dual-stage multimodal fusion framework for SL prediction across pan- and single-cancer contexts. The method combines a Cross-VAE with Product-of-Experts early fusion for four omics modalities and an RGCN-based KG encoder, and introduces an adaptive two-stage fusion/training paradigm (Uni-Modal Teacher, UMT; Uni-Modal Ensemble, UME) to mitigate modality laziness [36–39]. In extensive experiments covering eight specific cancer types and a pan-cancer dataset, SynLeaF demonstrated excellent generalization capability and robustness, surpassing existing state-of-the-art techniques across the majority of core evaluation metrics.

2 Results

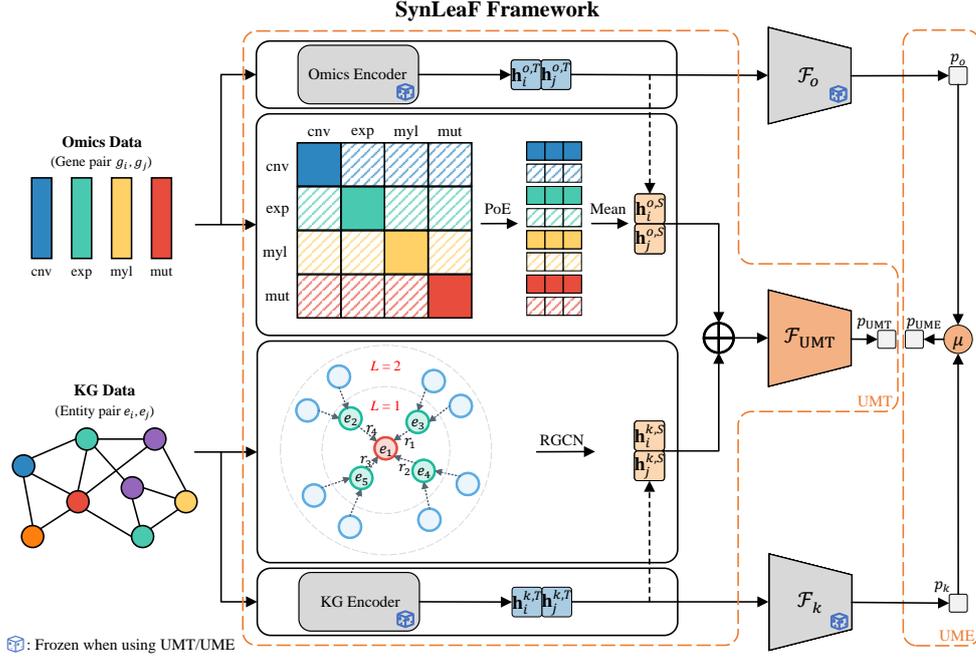### 2.1 SynLeaF: A Dual-Stage Multimodal Fusion Framework

As shown in Figure 1, SynLeaF is a dual-stage multimodal fusion framework, designed for the prediction of cancer synthetic lethality by integrating heterogeneous data from multiple sources. The model accepts two streams of input: first, four types of omics data, including gene expression, mutation, DNA methylation, and copy number variation (CNV) [1]; and second, a biomedical KG that contains various entities such as genes, pathways, and diseases, along with their relationships. The omics data are encoded through a VAE-based early fusion module, while the KG is processed by an RGCN to extract structured representations.

The high-dimensional nature and significant modal heterogeneity of omics data hinder the effective fusion of multi-source information. Drawing inspiration from the framework established by Wang et al. [40], we employed a Variational Autoencoder (VAE) architecture [36] and innovatively designed a Cross-VAE-based early fusion module. As shown in the Omics Encoder module of Figure 1, unlike traditional feature concatenation, our method constructs an $N \times N$ encoder matrix, which skillfully balances intra-modal feature self-learning (via diagonal autoencoders) with inter-modal interactive inference (via off-diagonal cross-encoders). By introducing the Product of Experts (PoE) mechanism [37] to aggregate multi-source information, SynLeaF can generate unified and robust gene representations in the latent space. Notably, this architecture naturally addresses the problem of missing data through its cross-inference mechanism. Even if some omics data are unavailable, the model can still reconstruct the missing features based on other modalities. This greatly improves the model's generalization ability when dealing with fragmented clinical data.

In multimodal learning, features are generally categorized into two distinct classes. The former comprises unimodal attributes acquirable via isolated training, whereas the latter consists of paired characteristics that necessitate cross-modal interaction

---

[1]Due to data source limitations, the pan-cancer dataset uses only three types of omics and does not include DNA methylation data.

Fig. 1 Overview of the SynLeaF Framework. SynLeaF constitutes a dual-stage multimodal integration architecture designed for synthetic lethality prediction, taking cancer-specific omics profiles and biomedical knowledge graphs as inputs. The **Omics Encoder** uses a cross-encoder architecture based on a Variational Autoencoder (VAE) and performs early fusion on four types of omics data: copy number variation (cnv), gene expression (exp), DNA methylation (myl), and mutation (mut), through a Product of Experts (PoE) mechanism. The **Knowledge Graph Encoder** utilizes a Relational Graph Convolutional Network (RGCN) to extract structural features associated with genes within the biological network. With these encoders, SynLeaF first independently pre-trains two unimodal models and then constructs two base estimators respectively under two complementary fusion strategies in further training. For both the UMT and UME strategies, the parameters of the pre-trained unimodal encoders are strictly frozen. Specifically, under the **UMT** strategy, SynLeaF treats the pre-trained unimodal encoders as teachers, guiding the training of a multimodal student model through feature-level knowledge distillation. Under the **UME** strategy, SynLeaF directly integrates the prediction results (p_o and p_k) from the two pre-trained unimodal models. Here, p_o and p_k denote the predicted probabilities from the Omics and Knowledge Graph models, respectively. Finally, SynLeaF adaptively selects the optimal strategy between UMT and UME according to observed validation efficacy.

for extraction. Optimally, the objective is for a multimodal model to capture paired features via cross-modal mechanisms, while ensuring that it also learns sufficient unimodal features. However, Du et al. [39] found that traditional multimodal late-fusion training methods suffer from the problem of "Modality Laziness". This means that encoders trained jointly on multiple modalities perform worse on unimodal feature learning than encoders trained unimodally, and this phenomenon is particularly evident in tasks where unimodal priors are meaningful. To mitigate this problem, Du et al. [39] proposed two complementary multimodal fusion strategies, designated as

Uni-Modal Teacher (UMT) and Uni-Modal Ensemble (UME), and achieved good performance in multimodal audio-visual classification tasks. The UMT strategy uses pre-trained unimodal encoders as "teachers" to guide a multimodal "student" model to learn the teachers' feature representations through feature-level Knowledge Distillation. The UME strategy, on the other hand, directly integrates the prediction results of the two pre-trained unimodal models.

Inspired by the UMT/UME framework, this paper proposes a dual-stage training strategy, which is adapted and extended for the characteristics of the omics-graph bimodal setting to ensure the slow modality is fully trained. During the initial stage, we conduct independent pre-training for both the omics encoder and the knowledge graph encoder to ensure that each unimodal encoder can fully learn the feature representations within its modality. During the second stage, considering the significant differences in inter-modal interactions across different cancer types and data splitting strategies, we designed an adaptive selection mechanism. After training, the model automatically selects the best strategy between UMT and UME based on the Area Under the ROC Curve (AUC) metric obtained from the validation dataset. UMT usually performs better when the unimodal features from both omics and the knowledge graph are strong, and the cross-modal interaction provides additional information. However, when one modality is clearly dominant or the cross-modal interaction introduces noise, the simple ensemble strategy of UME is more effective.

2.2 SynLeaF Surpasses Current State-of-the-Art Techniques in Both Pan-Cancer and Cancer-Specific Settings

To evaluate the effectiveness and robustness of SynLeaF, we conducted a comprehensive comparison with four state-of-the-art methods on a collection of datasets, which includes eight cancer-specific datasets and one large pan-cancer dataset, under various splitting settings. The eight specific cancer types are: breast cancer (BRCA), cervical cancer (CESC), colon cancer (COAD), kidney renal clear cell carcinoma (KIRC), acute myeloid leukemia (LAML), lung adenocarcinoma (LUAD), ovarian cancer (OV), and skin cutaneous melanoma (SKCM). These settings include CV1 (Random Split), CV2 (Semi-New Gene Split), and CV3 (New Gene Split) [23]. The four state-of-the-art methods compared were SLGNN, ELISL, PTGNN, and MPASL.

To guarantee a fair evaluation, we standardized the data loading part for all baseline methods to make sure they all used the same datasets, including the SL gene pairs, protein sequences, omics features, and knowledge graph data. The implementation details for baselines and all training configurations are described in the Supplementary Material (see *Experimental Design*). Empirical outcomes indicate that SynLeaF exhibits a distinct performance edge. On the core comparison metrics across a total of 19 scenarios, SynLeaF achieved state-of-the-art (SOTA) results in 17 of these instances.

Table 1 details the comparative outcomes across cancer-specific datasets, where we observe that SynLeaF has a significant lead in most cancer types. Notably, regarding the SKCM dataset, when evaluated against the second-best model, SynLeaF achieved huge improvements of 17.71% and 6.89% under the CV1 and CV2 settings, respectively. Since the ELISL method relies on cancer-specific clinical omics data and cell

Table 1 Experiment results on cancer-specific datasets. The standard deviations of the readings are reported in parentheses. The best performing method is highlighted in **bold** and the second best is <u>underlined</u>. The last column indicates the improvement made by SynLeaF.

Cancer	Split	Metric	SLGNN	ELISL	PTGNN	MPASL	SynLeaF	↑(%)
BRCA	CV1	AUC	0.8847(0.0130)	0.9041(0.0104)	<u>0.9453(0.0093)</u>	0.8187(0.0173)	**0.9654(0.0038)**	2.13
		AUPR	0.8949(0.0169)	0.9196(0.0079)	<u>0.9600(0.0084)</u>	0.8508(0.0118)	**0.9743(0.0020)**	1.49
	CV2	AUC	0.7147(0.1036)	0.7615(0.0470)	<u>0.8153(0.0911)</u>	0.6660(0.0701)	**0.8474(0.0502)**	3.94
		AUPR	0.7358(0.1114)	0.7894(0.0664)	<u>0.8316(0.0973)</u>	0.6888(0.0763)	**0.8707(0.0563)**	4.70
CESC	CV1	AUC	0.6251(0.0795)	0.7136(0.0797)	<u>0.7765(0.0480)</u>	0.7321(0.0946)	**0.8136(0.0505)**	4.78
		AUPR	0.6543(0.0846)	0.7466(0.0484)	<u>0.7915(0.0621)</u>	0.7299(0.0847)	**0.8180(0.0650)**	3.35
	CV2	AUC	0.5957(0.0489)	0.5280(0.0374)	<u>0.6834(0.0948)</u>	0.6297(0.0750)	**0.6845(0.0444)**	0.16
		AUPR	0.5834(0.0443)	0.5753(0.0384)	<u>0.6946(0.0967)</u>	0.6412(0.0272)	**0.7190(0.0530)**	3.51
COAD	CV1	AUC	0.5933(0.0291)	<u>0.6894(0.0150)</u>	0.6212(0.0498)	0.6278(0.0172)	**0.7162(0.0263)**	3.89
		AUPR	0.5979(0.0416)	<u>0.6645(0.0382)</u>	0.6071(0.0592)	0.6535(0.0257)	**0.7088(0.0287)**	6.67
	CV2	AUC	0.5404(0.0237)	<u>0.6206(0.0391)</u>	0.5157(0.0364)	0.5702(0.0385)	**0.6317(0.0244)**	1.79
		AUPR	0.5397(0.0280)	<u>0.6064(0.0316)</u>	0.5181(0.0279)	0.6031(0.0397)	**0.6120(0.0316)**	0.92
KIRC	CV1	AUC	0.6459(0.1381)	<u>0.6754(0.0886)</u>	**0.7161(0.0957)**	0.6739(0.0601)	0.6940(0.0990)	-
		AUPR	0.6668(0.1067)	0.7238(0.0806)	**0.7327(0.0908)**	0.7059(0.0835)	<u>0.7162(0.0855)</u>	-
	CV2	AUC	0.5417(0.1334)	**0.6599(0.1128)**	<u>0.6430(0.1319)</u>	0.6317(0.1601)	0.5822(0.0573)	-
		AUPR	0.5647(0.1054)	**0.6646(0.1106)**	<u>0.6441(0.1123)</u>	0.6191(0.1975)	0.5978(0.0770)	-
LAML	CV1	AUC	0.5793(0.0224)	0.6267(0.0061)	<u>0.6960(0.0303)</u>	0.6306(0.0256)	**0.6980(0.0156)**	0.29
		AUPR	0.5914(0.0247)	0.6310(0.0177)	<u>0.6925(0.0533)</u>	0.6605(0.0151)	**0.7002(0.0239)**	1.11
	CV2	AUC	0.5467(0.0258)	0.5810(0.0218)	<u>0.6188(0.0366)</u>	0.5977(0.0468)	**0.6296(0.0195)**	1.75
		AUPR	0.5648(0.0200)	0.5930(0.0222)	<u>0.6220(0.0518)</u>	0.6297(0.0326)	**0.6378(0.0205)**	1.29
LUAD	CV1	AUC	0.8254(0.0229)	0.8513(0.0295)	<u>0.8865(0.0254)</u>	0.7945(0.0584)	**0.9000(0.0259)**	1.52
		AUPR	0.8372(0.0132)	0.8571(0.0336)	<u>0.8753(0.0356)</u>	0.7951(0.0471)	**0.8873(0.0273)**	1.37
	CV2	AUC	0.5858(0.1520)	0.7465(0.0920)	<u>0.7678(0.0982)</u>	0.6336(0.1427)	**0.8161(0.0854)**	6.29
		AUPR	0.6365(0.1477)	0.7407(0.1208)	<u>0.7676(0.1140)</u>	0.6636(0.1326)	**0.7924(0.1110)**	3.23
OV	CV1	AUC	0.9201(0.0209)	0.7790(0.0683)	<u>0.9824(0.0116)</u>	0.8555(0.0341)	**0.9827(0.0144)**	0.03
		AUPR	0.9426(0.0166)	0.7812(0.0867)	<u>0.9590(0.0384)</u>	0.8325(0.0413)	**0.9855(0.0103)**	2.76
	CV2	AUC	0.5894(0.0828)	0.7087(0.0300)	<u>0.7416(0.0813)</u>	0.6846(0.0601)	**0.7990(0.0850)**	7.74
		AUPR	0.6136(0.0598)	0.6812(0.0202)	<u>0.6694(0.0854)</u>	0.6867(0.0620)	**0.7252(0.1149)**	5.61
SKCM	CV1	AUC	0.6692(0.1494)	<u>0.6871(0.0248)</u>	0.6469(0.0788)	0.5849(0.1209)	**0.8088(0.0307)**	17.71
		AUPR	0.7106(0.1710)	<u>0.7173(0.0451)</u>	0.6815(0.0737)	0.6745(0.0609)	**0.8471(0.0436)**	18.10
	CV2	AUC	0.6777(0.0862)	0.6302(0.1424)	<u>0.7138(0.2257)</u>	0.5760(0.1715)	**0.7630(0.1576)**	6.89
		AUPR	0.7059(0.0661)	0.6216(0.1371)	<u>0.7251(0.2111)</u>	0.6150(0.1551)	**0.7876(0.1566)**	8.62

Table 2 Comparison of experimental results on the pan-cancer dataset

Cancer	Split	Metric	SLGNN	PTGNN	MPASL	SynLeaF	↑(%)
pan	CV1	AUC	<u>0.9550(0.0021)</u>	0.9315(0.0011)	0.9336(0.0037)	**0.9652(0.0012)**	1.07
		AUPR	<u>0.9616(0.0018)</u>	0.9386(0.0024)	0.9425(0.0033)	**0.9669(0.0008)**	0.55
		F1	<u>0.8964(0.0023)</u>	0.8894(0.0014)	0.8692(0.0041)	**0.9099(0.0018)**	1.51
	CV2	AUC	<u>0.7736(0.0305)</u>	0.7684(0.0657)	0.4900(0.0234)	**0.8624(0.0236)**	11.48
		AUPR	0.8050(0.0255)	<u>0.8104(0.0448)</u>	0.6118(0.0232)	**0.8754(0.0226)**	8.02
		F1	0.6006(0.0394)	<u>0.7189(0.0433)</u>	0.3858(0.1325)	**0.7955(0.0246)**	10.66
	CV3	AUC	<u>0.5757(0.0367)</u>	0.5128(0.0576)	0.5379(0.0469)	**0.7407(0.0271)**	28.66
		AUPR	<u>0.6050(0.0414)</u>	0.5213(0.0494)	0.5416(0.0413)	**0.7611(0.0417)**	25.80
		F1	0.1163(0.0944)	<u>0.6673(0.0012)</u>	0.0000(0.0000)	**0.7153(0.0156)**	7.19

line omics data, it cannot be applied to the pan-cancer synthetic lethality prediction task. Therefore, it was not included in the comparison. The experimental data confirm that our proposed method achieved the highest performance metrics under all splitting strategies on the pan-cancer dataset. Particularly in the CV3 (zero-shot) setting, which simulates the prediction of unknown genes, SynLeaF still achieved an AUC of 0.7407, representing an improvement of up to 28.66% over the second-place SLGNN. PTGNN showed strong competitiveness in the single-cancer experiments, ranking second after SynLeaF on most metrics. However, this advantage has its limitations. By looking at Table 1 and Table 2 together, we can see that although PTGNN can fit the single distribution of a specific cancer, its performance dropped significantly when faced with the pan-cancer scenario, which has more complex data and greater distribution differences. In contrast, SynLeaF demonstrated an all-around adaptability, as

it maintained the best performance across both the smaller single-cancer datasets and the larger pan-cancer dataset.

Although SynLeaF performed excellently on the vast majority of datasets, the framework failed to secure the leading position on the KIRC dataset. We conducted an in-depth analysis of this phenomenon and found that the core reason is the extremely small sample size (only about 120 samples). This caused a serious distribution shift between the validation and test sets, which in turn led to a misjudgment by the adaptive selection strategy. This statistical deviation caused the validation-set-based adaptive selection strategy to be conservative. Specifically, the sub-modules of SynLeaF actually have the potential to reach SOTA on KIRC, especially in the CV1 setting. On the testing partition, the UMT branch achieved an AUC of 0.7218 (± 0.0923) and an AUPR of 0.7246 (± 0.0737), metrics that align closely with the efficacy exhibited by the top-ranked PTGNN. This result indicates that in small-sample and highly heterogeneous cancer datasets, relying solely on validation set metrics for model selection can be risky. The result on KIRC reveals that how to overcome distribution shift and evaluation bias in extremely data-sparse scenarios remains a common challenge for the entire field of computational biology.

2.3 The Dual-Stage Adaptive Fusion Strategy Effectively Addresses Modality Dependency and Heterogeneity Challenges

To investigate the contribution of multimodal data to synthetic lethality prediction and to validate the effectiveness of the SynLeaF dual-stage fusion strategy, we executed an in-depth ablation analysis focusing on the Only Omics, Only KG, and the full version of SynLeaF.

2.3.1 Multimodal Integration Shows an "Envelope Effect" and Robustness Superior to Unimodal Benchmarks

As shown in Figure 2, SynLeaF forms a clear "envelope effect" over the unimodal models, indicating stable gains from multimodal fusion. Notably, modal advantages shift drastically depending on the data split. For SKCM in CV1, KG provides the main predictive signal (AUC $\approx$ 0.78) compared to Omics (AUC $\approx$ 0.64). However, in CV2 (unseen genes), a modal reversal occurs: KG drops sharply to 0.60 due to sparse graph connections, while Omics rises to 0.74. Despite this fluctuation, SynLeaF adaptively shifts its focus to Omics, maintaining a robust AUC of 0.76.

It is worth noting that this modal advantage shows a different pattern from the pan-cancer perspective. In the CV1 setting on the pan-cancer dataset, both Omics and KG showed very high performance. But in the more challenging zero-shot condition of the CV3 setting, the situation was reversed. The performance of Only Omics dropped significantly to 0.6259, while Only KG still maintained a robust AUC of 0.7391. This reveals that in complex scenarios that span multiple cancer types, simple omics features are easily affected by heterogeneity noise. The global knowledge graph, on the other hand, provides a stronger inductive bias through the biological network topology, thus enabling structural inference on completely unknown genes.
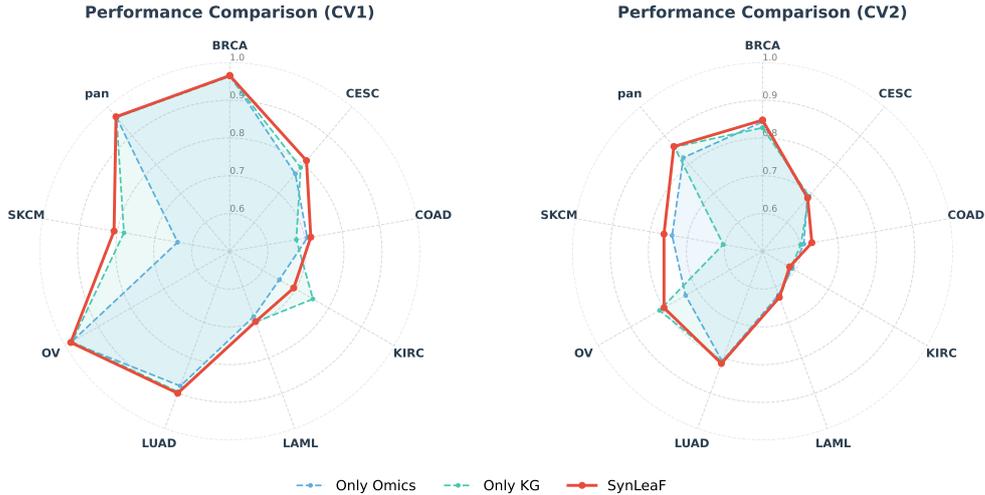
Fig. 2 Radar chart comparing the performance of SynLeaF and unimodal baseline variants. This chart shows the AUC performance comparison of SynLeaF against the two unimodal baseline variants, Only Omics and Only KG, on the pan-cancer and eight cancer-specific datasets, under the two data splitting strategies of CV1 and CV2. The performance curve of SynLeaF forms an "envelope effect" over the unimodal models on almost all datasets, demonstrating the consistent advantage of multimodal fusion.

In summary, the local case of SKCM and the global case of pan-cancer together prove that no single modality can excel under all splits, and SynLeaF's multimodal fusion mechanism is an essential approach to handle this complexity.

2.3.2 The Adaptive Selection Mechanism Captures the Differential Modality Dependency of Different Cancers

Although multimodal fusion is generally effective, we observed a key phenomenon that different cancer datasets show very different modality dependency under different splitting strategies. In the synthetic lethality prediction task, deep cross-modal interaction is very important in some scenarios, while in other scenarios, forced interaction can instead introduce noise. Therefore, SynLeaF introduces an adaptive selection mechanism to address the significant differences in inter-modal interactions under different data splitting strategies. To accommodate the heterogeneity within data distributions, the model identifies the most effective fusion path by evaluating outcomes on the validation set. We plotted bar charts to evaluate the effectiveness of the modal fusion strategies. Due to space limitations, this section only discusses two representative datasets, CESC CV1 and COAD CV1, as shown in Figure 3. Detailed outcomes of the full experiments are presented in Supplementary Figure S3.

In the CESC (CV1) dataset, unimodal Omics and KG yielded AUCs of 0.7679 and 0.7899, respectively. A simple late fusion (UME) achieved only 0.7959, struggling to capture deep interactions. However, consistent with the first scenario in Du et al. [39], SynLeaF's UMT strategy achieved a significant advantage (AUC=0.8136). By
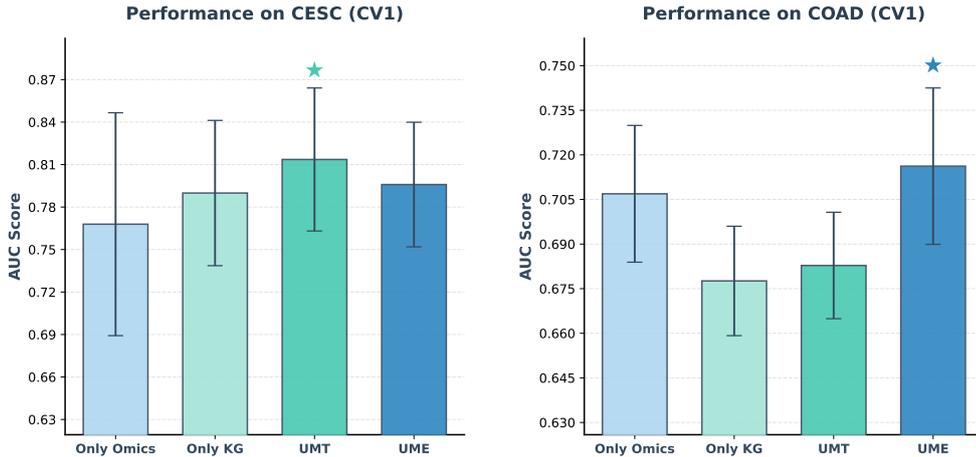
Fig. 3 **Performance comparison of SynLeaF baseline variants on two cancer datasets.**
This figure shows the AUC scores for the unimodal baseline variants (Only Omics, Only KG) and
the baseline variants respectively employing two multimodal fusion strategies (UMT, UME) on the
CESC and COAD cancer datasets under the CV1 split. The height of the bars corresponds to the
mean AUC obtained via 5-fold cross-validation, while the standard deviations are denoted by the
error bars. The star indicates the optimal fusion strategy, which is finally adopted in the SynLeaF
adaptive selection mechanism on that dataset.

utilizing feature-level distillation, UMT effectively forces the network to learn from
both modalities and retain key cross-modal interactions. Conversely, COAD (CV1)
presents a "strong omics (0.7069) and weak graph (0.6776)" scenario. Here, forcing
feature alignment via UMT introduced noise, dropping the AUC to 0.6828 (lower than
Only Omics). As proposed by Du et al. [39] for cases with insignificant paired interac-
tions, the UME strategy selected by SynLeaF performed best (AUC=0.7162). UME
straightforwardly aggregates unimodal results, effectively avoiding modality laziness
or negative transfer caused by forced cross-modal interactions.

2.3.3 Parameter Sensitivity Analysis Confirms the Complementary Stability of UMT and UME Strategies

To evaluate the contribution of the feature-level distillation module and determine the
optimal hyperparameter configuration, a sensitivity analysis was performed regarding
the distillation weight λ_{distill} in equation (10) for the UMT module, within the range
of $[0, 1, 10, 20, 50, 100]$ (as shown in Figure 4). Here, $\lambda_{\text{distill}} = 0$ is equivalent to Naïve
early fusion without distillation regularization.

The experimental results show that for most datasets, introducing moderate distil-
lation regularization ($\lambda_{\text{distill}} = 1$) significantly outperforms the no-distillation baseline
($\lambda_{\text{distill}} = 0$). For example, in CESC (CV1), the AUC increased from 0.7995 at
$\lambda_{\text{distill}} = 0$ to 0.8136 at $\lambda_{\text{distill}} = 1$. This indicates that lightweight feature alignment
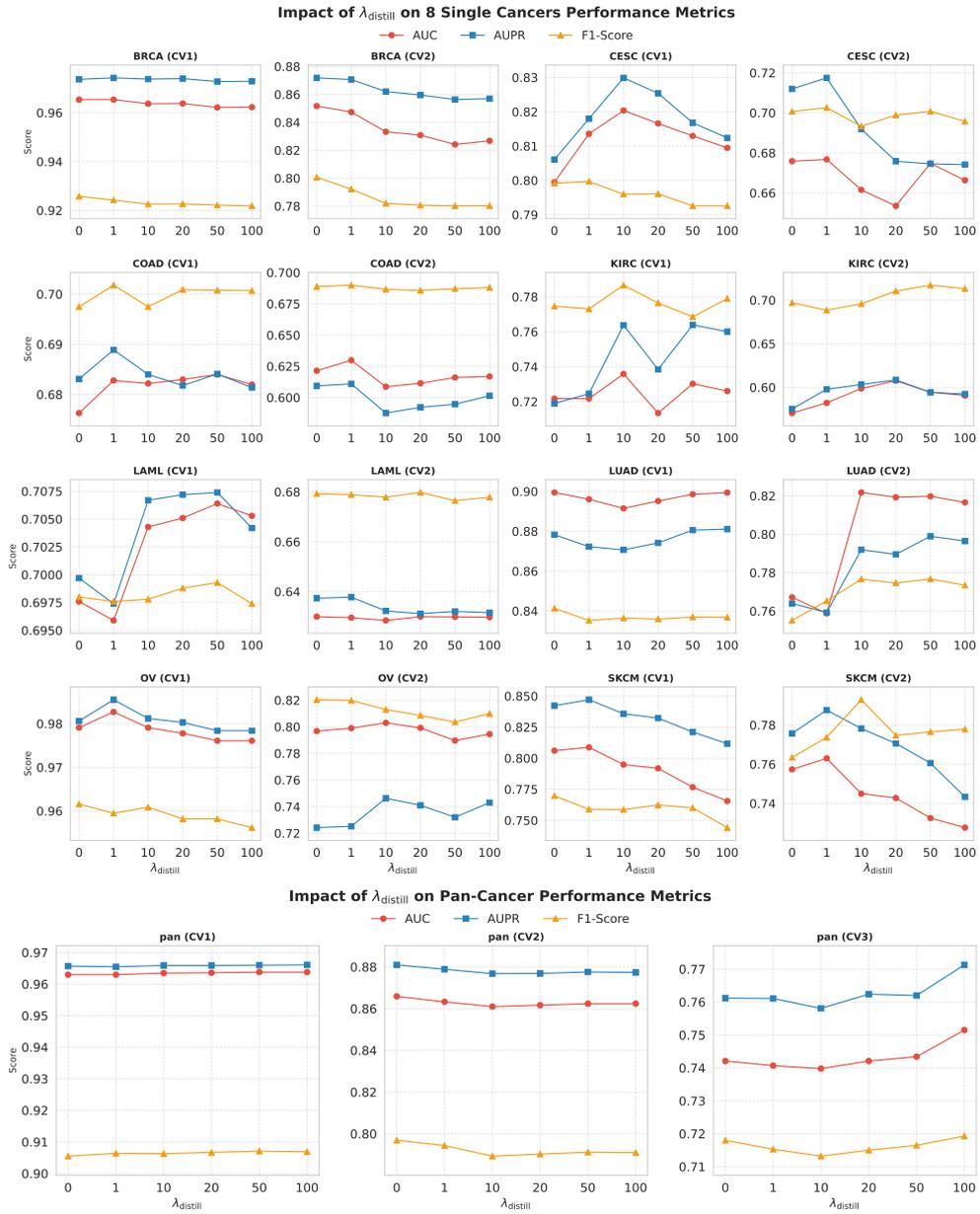
Fig. 4 Sensitivity analysis of the λ_{distill} parameter in the UMT module. This figure shows how the performance (AUC, AUPR, F1-Score) of the UMT fusion strategy changes with the knowledge distillation weight λ_{distill} on all single-cancer and pan-cancer datasets. A value of $\lambda_{\text{distill}} = 0$ corresponds to Naïve multimodal training without distillation regularization.

can effectively mitigate modality laziness while avoiding the risk of excessive regularization. Based on the majority principle and considerations for model generality, we uniformly set $\lambda_{\mathrm{distill}}$ to 1 in the final SynLeaF model.

However, this fixed-parameter strategy inevitably faces challenges in some highly heterogeneous datasets. We observed that LAML (CV1) and LUAD (CV2) showed a special decrease-then-increase trend. A weak distillation ($\lambda_{\mathrm{distill}} = 1$) actually interfered with the model's feature learning, leading to performance lower than the baseline with $\lambda_{\mathrm{distill}} = 0$. For example, in LUAD (CV2), the AUC dropped from 0.7672 to 0.7589 when $\lambda_{\mathrm{distill}} = 1$. Although the data trend shows that the model could overcome this obstacle and achieve better performance if the distillation weight were further increased (for example, $\lambda_{\mathrm{distill}} \geq 10$), under the unified setting of $\lambda_{\mathrm{distill}} = 1$, the UMT module did indeed reach a performance low. But during the validation phase, SynLeaF's adaptive strategy successfully identified the performance decline of UMT and selected UME as the final inference model. This allowed the final model to still maintain a competitive performance (the final AUC for LUAD CV2 was 0.8161). This result precisely validates the necessity of the adaptive selection mechanism and the high fault-tolerance of SynLeaF's dual-stage architecture. It allows the model to use a single set of general hyperparameters for most scenarios, while relying on the adaptive switching mechanism to provide a robust fallback for the few scenarios that are sensitive to parameters.

2.4 Gradient Dynamics Analysis Reveals the Mechanism for Mitigating Modality Laziness

To better understand how the UMT strategy effectively mitigates the modality laziness problem in multimodal learning, The CESC (CV1, Fold2) dataset, which shows a typical UMT advantage, was employed as a representative case to study the gradient dynamics across the training phase (as shown in Figure 5).

2.4.1 The UMT Strategy Significantly Improves Both Overall and Unimodal Performance

Figure 5(a) shows the comparison of overall test performance. In the early stage of training (about the first 50 epochs), the Naïve no-distillation baseline converges faster because it lacks additional regularization. However, the cost of this rapid convergence is overfitting. After reaching its peak, the performance fluctuations followed by a decline are observed in the baseline model. In contrast, although the UMT strategy rises more slowly at the beginning, it shows a continuous and stable learning ability, and continues to climb after surpassing the baseline around epoch 60, eventually stabilizing at a high level above 0.86. The light green shaded area visually illustrates the substantial performance advantage achieved by the UMT strategy during the advanced training epochs.

Figure 5(b) further reveals that this overall performance improvement comes from improvements at the unimodal level. We observed two key phenomena. First, for the Omics modality, the Naïve no-distillation baseline ($\lambda_{\mathrm{distill}} = 0$) experienced a significant performance collapse in the later stages of training. The AUC dropped
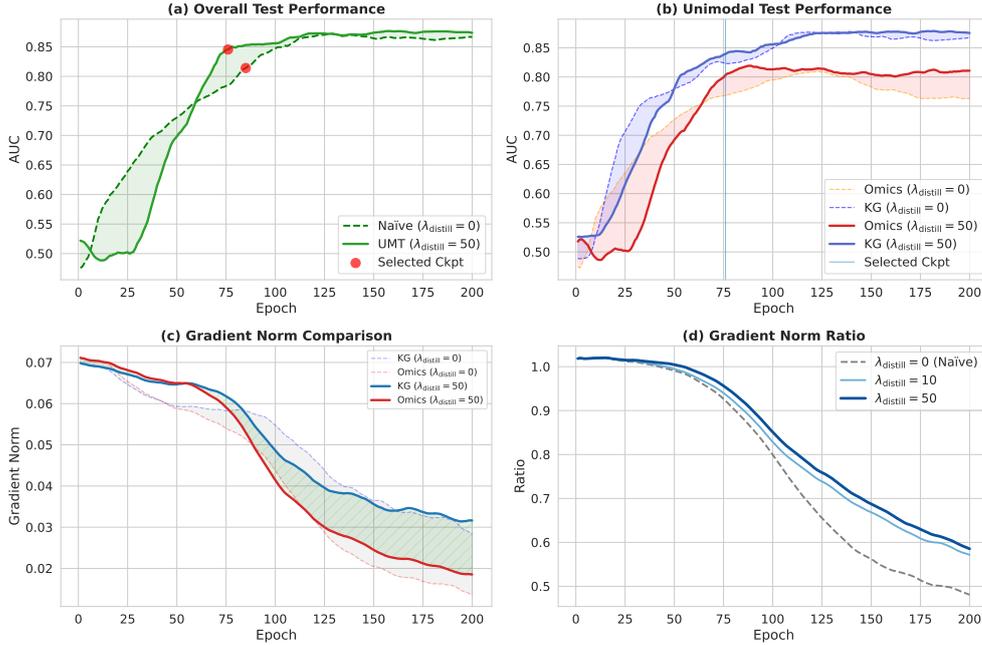
Fig. 5 Gradient dynamics analysis on the CESC (CV1, Fold2) dataset. (a) Overall test performance comparison: Test AUC curves for UMT (solid line) and the Naïve no-distillation baseline (dashed line). The red dots mark the checkpoints selected based on the validation set. **(b)** improvements in unimodal scenarios: change in test AUC for the Omics and KG modalities under the UMT and Naïve no-distillation strategies (denoted as Omics/KG ($\lambda_{\text{distill}} = 50$) and Omics/KG ($\lambda_{\text{distill}} = 0$), respectively). The vertical line marks the UMT checkpoint position. **(c)** Gradient norm comparison: change in the gradient norms of the two modality encoders over training epochs under different distillation weights λ_{distill}. **(d)** Gradient norm ratio: trend of the gradient balance between the two modalities. Here, the gradient norm ratio is defined as the L_2 norm of the gradients with respect to the Omics encoder's parameters divided by that of the KG encoder's parameters.

sharply from its peak to about 0.75, which is a typical phenomenon of overfitting. In contrast, the UMT strategy ($\lambda_{\text{distill}} = 50$) not only increased steadily but also successfully avoided performance degradation in the later stages, always staying above 0.80. Second, for the KG modality, the UMT strategy also maintained a small but consistent advantage compared to the baseline, indicating that the knowledge distillation mechanism had a positive regularization effect on both modalities.

Notably, the checkpoints denoted by red dots within Figure 5(a) are not the global optimal points of their respective curves on the test set, and there is still room for improvement after these points. This observation once again confirms the previous discussion about KIRC, showing that the distribution shift between the validation and test sets is a common challenge in biomedical data analysis [41]. Selecting models based on validation set metrics may fail to capture the true optimal state on the test set.

13

2.4.2 Gradient Norm Analysis: Knowledge Distillation Enhances the Learning Signal of the Weaker Modality

To explain the above phenomena from an optimization dynamics perspective, we analyzed the changes in the gradient norms of the two modal encoders during the training process (Figure 5(c)), following the analytical protocols established in recent multimodal learning studies [42].

In the Naïve no-distillation baseline ($\lambda_{\text{distill}} = 0$), we observed a significant asymmetry in gradient decay. In the early stage of training, the gradient norms of both modalities were at a relatively high level of about 0.07. However, as the training progressed, the gradient norm of the Omics modality dropped sharply, eventually falling below about 0.02. On the other hand, the decrease for the KG modality was relatively gentle, finally staying at a level of about 0.03. This difference means that in the later stages of training, the Omics encoder almost stopped learning effectively. The model mainly relied on the KG modality for prediction, which is direct evidence of modality laziness.

After introducing the knowledge distillation regularization ($\lambda_{\text{distill}} = 50$), the situation changed significantly. The light green shaded area clearly shows how the UMT strategy successfully increased the gradient norm of the Omics modality. Although the gradient for Omics was still decreasing, the rate of decrease slowed down significantly. This means that the distillation loss provided an additional supervision signal for the Omics encoder, forcing it to continue learning to align with the feature representations derived from the pre-trained teacher model and thus avoiding early gradient vanishing. Notably, under both strategies, the gradients for the KG modality in the later training stages almost overlapped. This indicates that the main role of UMT is to selectively enhance the weaker modality, rather than interfering with the normal learning of the stronger modality.

Figure 5(d) more directly quantifies the degree of learning balance between the two modalities through the gradient ratio (Omics/KG). A ratio closer to 1.0 indicates that the learning dynamics of the two modalities are more balanced, while a decreasing ratio means that the relative contribution of Omics is weakening. In the Naïve no-distillation baseline ($\lambda_{\text{distill}} = 0$), the gradient ratio starts from a perfectly balanced 1.0 and then rapidly decreases, dropping to 0.50 in the later training stages. This means that by the end of training, the gradient contribution of the Omics modality was only half that of the KG modality, causing a serious learning imbalance. As the distillation weight λ_{distill} increases, this imbalance is gradually mitigated. A value of $\lambda_{\text{distill}} = 10$ keeps the ratio at a higher level, and the effect of $\lambda_{\text{distill}} = 50$ is the most significant. Even at the end of training, the ratio remains at about 0.58, showing a clear improvement in balance compared to the baseline's value of below 0.50.

The gradient dynamics analysis above provides direct mechanical evidence for the effectiveness of the UMT strategy. Through the knowledge distillation framework, the pre-trained Uni-modal Teacher model provides continuous feature-level supervision for the student encoders in the joint training. This effectively enhances the gradient signal of the weaker modality and reduces the learning gap with the stronger modality. We also honestly point out that the UMT strategy did not completely eliminate the gradient gap between modalities. Even with $\lambda_{\text{distill}} = 50$, the final gradient ratio was

still 0.58, which is still far from the ideal 1.0. This suggests that modality laziness is a deep-rooted optimization problem that is difficult to solve completely with a single technique. However, the experimental results show that even mitigating rather than completely eliminating the problem is enough to bring substantial improvements in predictive performance, which validates the rationale of our method's design.

2.5 SynLeaF Web Server and Case Study Applications

To facilitate the exploration of synthetic lethality interactions, we developed a web-based query platform[2]. The backend uses an ensemble of five CV2 models per cancer type and averages their outputs to improve robustness. The interface visualizes predictions across eight cancers and the pan-cancer context, alongside ground-truth labels from existing datasets for direct comparison.

Table 3 Prediction scores retrieved from the SynLeaF web server for two case studies.

Gene Pairs	BRCA	CESC	COAD	KIRC	LAML	LUAD	OV	SKCM	pan
RAD51–BRCA1	**0.9928**	0.3817	0.6562	0.6038	0.5830	0.4169	0.3849	0.5850	0.9748
RAD51–BRCA2	**0.9213**	0.4462	0.4572	0.6160	0.5406	0.5259	0.3624	0.6082	0.9596
ADAR-BRCA1	**0.9307**	0.4876	0.5443	0.3883	0.5413	0.4984	0.2611	0.5221	0.8667
ADAR-BRCA2	**0.7750**	0.4697	0.3759	0.3465	0.5174	0.4815	0.1832	0.5595	0.7112

Milordini et al. [6] demonstrated the synthetic lethality of targeting the RAD51–BRCA2 interaction, particularly in pancreatic cancer. Although these pairs are labeled as positive in our pan-cancer dataset, SynLeaF provides detailed context-specific insights. As shown in Table 3, the prediction scores in the BRCA column are exceptionally high, significantly surpassing those in other cancer types. This strong signal suggests that breast cancer may share the same RAD51-mediated synthetic lethality mechanism observed in pancreatic cancer. Chabanon et al. [7] identified ADAR1 as a key synthetic lethal target in BRCA-mutant cancers. Notably, despite the absence of ADAR-BRCA* pairs from our dataset, SynLeaF successfully predicted this latent relationship, crucially exhibiting high efficacy in the relevant cancer context: the prediction score in the BRCA column reached 0.9307 for ADAR-BRCA1, which is remarkably higher than in other unrelated cancer types. This result demonstrates SynLeaF's capability to generalize and discover novel, clinically relevant synthetic lethality pairs in specific cancer types.

3 Discussion

Synthetic lethality is a key mechanism in precision oncology, yet identifying robust SL pairs remains challenging [9]. Our results show that SynLeaF improves prediction performance across pan- and single-cancer settings and adapts to modality dependency differences. A key finding is that balancing modality interaction and avoiding modality laziness are crucial for multimodal SL prediction [39]. SynLeaF combines unimodal

---

[2]The web server is accessible at: https://synleaf.bioinformatics-lilab.cn

pre-training with two complementary fusion strategies: UMT for deep interaction and UME for conservative ensembling, providing robustness across diverse cancer contexts and alleviating negative transfer [43].

Although SynLeaF has achieved a breakthrough in prediction accuracy, the current work still faces limitations regarding the practical requirements for clinical translation. First, our model lacks interpretability regarding multimodal interactions. Future work will be devoted to clarifying how continuous omics features enhance or suppress KG topology to provide a complete biological evidence chain. Second, the 1:1 class balancing strategy deviates from the highly imbalanced biological reality. Furthermore, distribution shifts between validation and test sets in small-sample cancers hinder optimal model selection. Future research could reframe SL prediction as a recommendation task to handle long-tail distributions and develop more robust evaluation strategies. Finally, while Cross-VAE handles missing intra-omics data, SynLeaF still requires both omics and KG modalities. In clinical practice, a patient may completely lack sequencing data, or some new genes may not be included in existing knowledge graphs. Future studies could explore cross-modal completion techniques based on generative modeling to enable flexible prediction when only a single fundamental modality is available.

4 Conclusion

This study introduces SynLeaF, a dual-stage multimodal fusion framework for synthetic lethality prediction across pan- and single-cancer contexts. Through VAE-enhanced omics encoding and a knowledge distillation strategy, SynLeaF effectively overcomes the challenges of modality laziness and heterogeneity in multi-source data fusion. Extensive experiments show that the proposed framework exhibits state-of-the-art capability within prediction scenarios targeting for both known and new genes. SynLeaF not only provides a powerful computational tool for discovering synthetic lethality targets, but its approach to handling dynamic multimodal dependencies also offers a new perspective for general link prediction tasks in the biomedical field.

5 Methods

5.1 Data Acquisition and Preprocessing

We constructed a strictly filtered and multi-source integrated pan-cancer and cancer-specific dataset. The data covers SL gene pairs, multi-omics data, biomedical knowledge graphs, and protein sequence information. We integrated data from authoritative databases such as SynLethDB 2.0 [3], ELISL [34], TCGA [44], and UniProt [45], and we designed a standardized preprocessing workflow to eliminate noise and heterogeneity.

5.1.1 Data Sources

Synthetic Lethality Data. The synthetic lethality label data comes from two main sources. For the pan-cancer prediction task, we use the SynLethDB 2.0 database, which

provides general synthetic lethality gene pairs across cancer types. For the cancer-specific prediction task, we adopted the dataset organized by the ELISL study. ELISL integrates high-confidence SL gene pairs from previous studies such as DiscoverSL [14], ISLE [46], and EXP2SL [15]. It covers eight specific cancer types, and the number of synthetic lethality pairs before preprocessing is shown in Supplementary Table S3.

Multi-Omics Data. The omics data comes from the TCGA database curated by the cBioPortal platform [47]. We collected four key omics features for the above eight cancer types: Copy Number Variation, which represents the relative linear copy number variation values of genes; Gene Expression, which uses mRNA z-score data processed by RNA-Seq V2 RSEM normalization; DNA Methylation, which reflects the epigenetic regulation status of genes (HM27 or HM450); and Mutation data, which records the mutation status of each gene in every sample, where we only consider the mutation counts of genes in each sample. For the pan-cancer dataset, due to data limitations, the dataset is restricted to three data modalities consisting of gene expression, mutation, and copy number variation.

Knowledge Graph. We adopted the biomedical knowledge graph (SLKG 2.0) provided by SynLethDB 2.0. This knowledge graph contains 37,341 entities of 11 types (including genes, Gene Ontology, pathways, drugs, and diseases) and 1,405,652 relationships of 27 types, forming a rich biomedical knowledge network.

Protein Sequences. To support comparative experiments, we downloaded reviewed human protein sequence data from the UniProt database (as of October 26, 2024), which contains a total of 20,428 sequences where each gene corresponds to a representative protein sequence consisting of 21 types of amino acids.

5.1.2 Data Preprocessing

We executed rigorous protocols for gene filtering and data cleaning to guarantee the uniformity and superior quality of the multimodal data.

Gene Filtering and Alignment. We performed multi-condition intersection filtering on the genes. Genes included in the study must meet the following conditions simultaneously: they must have corresponding reviewed protein sequences in UniProt, be recorded in at least one type of omics data, and have at least one accessible neighbor node in the knowledge graph. Through this strict filtering mechanism, we filtered out genes with incomplete information and ensured that every gene used for model training has a complete multimodal feature representation. The number of synthetic lethality pairs after filtering is shown in Table 4.

SL Data Construction and Balancing. To address the noise and imbalance issues in the original SL data, we performed the following processing. First, in the pan-cancer dataset, we treated pairs explicitly labeled as Non-SL and Synthetic Rescue as negative samples. Second, for conflicting gene pairs that appeared in both the positive and negative sample sets, we adopted a conservative strategy and uniformly classified them as negative samples. Additionally, we corrected errors in the ELISL data where some self-loops were incorrectly identified as positive samples. Consequently, we reassigned them to the negative category. Considering the inherent sparsity of genuine synthetic lethality associations, the quantity of negative samples within the dataset typically vastly surpasses the count of positive ones. To avoid the model biasing

Table 4 Statistics of synthetic lethality pair counts after data processing

Cancer Type	# Total Genes	# Pos.	# Neg.
BRCA	17965	1349	990
CESC	17977	144	4738
COAD	17961	1560	70982
KIRC	17963	60	2514
LAML	17944	1147	18912
LUAD	17954	582	5460
OV	17958	253	556
SKCM	17969	101	16157
pan	17690	33746	3509

Pos. denotes the count of positive samples; # Neg. denotes the count of negative samples.

towards the majority class, we adopted a 1:1 positive-negative sample balance strategy. When there were too many negative samples, we used random undersampling; when known negative samples were insufficient in specific cancer types, we randomly generated unlabeled gene pairs from the filtered gene pool as supplementary negative samples, and we ensured that these generated pairs did not overlap with known positive samples.

Omics Data Normalization. The sample counts for different features across various cancer types in the original omics data are shown in Figure 6. We performed sample-level alignment on the omics data for each cancer to ensure that the four omics features for each case sample are one-to-one matched.

To address the sparsity and distribution characteristics of the data, the processing workflow is as follows. Due to the sparsity of omics data, we filled all missing values and non-numeric values with 0 to represent no abnormality or a default state; for gene expression data, we truncated values smaller than -10 to -10 because we observed that extremely small values could cause numerical instability; regarding mutation profiles, we calculated the mutation frequency for every individual gene and applied a logarithmic transformation to map it to the interval $[0, 1]$ using the following formula:

$$x' = \begin{cases} \frac{\ln(1+x)}{\ln(1+M)}, & \text{if } M > 0 \\ 0, & \text{if } M = 0 \end{cases} \tag{1}$$

where x is the original mutation count and M is the maximum mutation count in that cancer type. Furthermore, given that synthetic lethality effects often involve interactions between normal genes and abnormal genes [48], we did not exclude genes with low variation based on the case sample distribution but instead retained all genes that met the omics alignment requirements. We retained only unique records for duplicate gene entries appearing in single-omics data.

Knowledge Graph Refinement. To construct cancer-specific knowledge graphs, we performed subgraph extraction on the original graph based on cancer types. For each specific cancer, we only retained its corresponding "Disease" node and the edges
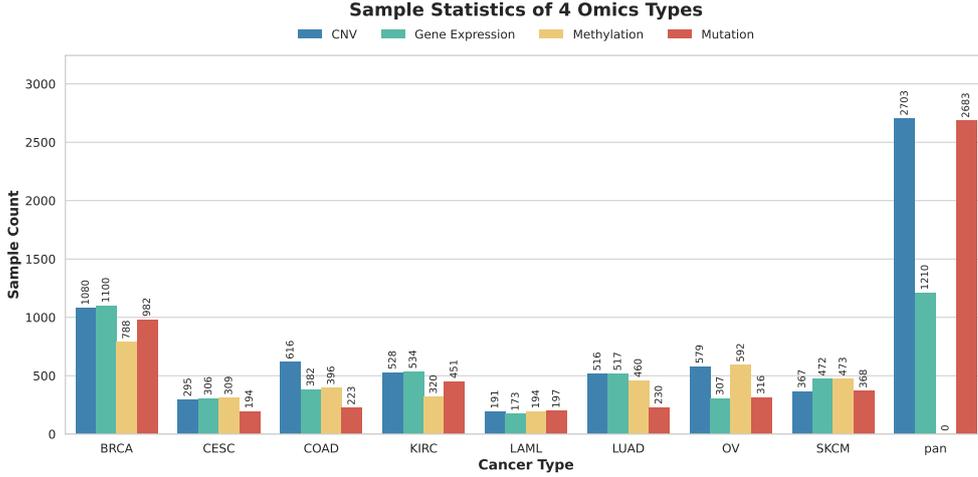
Fig. 6 Statistics of sample counts for original omics data across cancer types. This figure displays the number of raw available samples before preprocessing for four omics data types (Copy Number Variation (CNV), Gene Expression, DNA Methylation, and Mutation) across the eight specific cancer types and the pan-cancer dataset used in this study. Noting that the pan-cancer (pan) dataset does not include DNA methylation data due to data source limitations.

directly connected to it, removing other irrelevant cancer nodes to reduce noise interference. Additionally, for the unidirectional relationships in the graph, we generated corresponding reversed edges to transform the directed graph into a heterogeneous graph containing bidirectional information.

5.1.3 Dataset Splitting Strategies

In order to thoroughly assess the robustness and generalization capacity of our framework, we referred to the work of PiLSL [23] and adopted three cross-validation (CV) splitting strategies. The partitioning of all datasets adhered to a fixed proportion of 7:1:2 for the training, validation, and testing subsets, respectively.

- **CV1 (Random Split)**: All samples (gene pairs) are randomly shuffled and split. Under this setting, genes present in the testing phase might also exist within the training set. This primarily evaluates the inductive potential of the model concerning novel combinations of known genes.
- **CV2 (Semi-New Gene Split)**: The gene set is partitioned to guarantee that for every gene pair situated in the validation or testing subsets, precisely a single gene is found in the training data, while the other gene is completely new to the training set. This simulates the real-world scenario of finding known targets for new genes, and it evaluates the model's generalization ability for semi-new gene pairs.
- **CV3 (New Gene Split)**: This is the strictest splitting method, ensuring a complete absence of test-set genes within the training set. This tests the model's inferential capability in a completely unexplored gene space (zero-shot setting).

19

It is worth noting that due to the small sample size of some cancer-specific datasets, performing CV3 splitting may result in extremely few samples in the test set, rendering it statistically insignificant. Therefore, in subsequent experiments, we only report the results of CV1 and CV2 for single-cancer prediction tasks, while we use all three splitting strategies for comprehensive evaluation in the pan-cancer prediction task.

5.2 The SynLeaF Framework

This paper introduces the SynLeaF architecture, a deep learning system grounded in dual-stage multimodal fusion strategies. This framework aims to accurately predict SL interactions by integrating omics features from genomics and structured knowledge from biomedical KGs. The architecture of the model is illustrated in Figure 1.

5.2.1 Problem Definition

We define the synthetic lethality prediction task as a binary classification problem. Given a specific instance of a gene pair (g_i, g_j), the objective is to predict whether an SL relationship exists between them, which corresponds to outputting a label $y \in \{0, 1\}$. Our method follows a unified Siamese-like network architecture [49] for representation learning of gene pairs, regardless of whether it is in a single-modal or multimodal setting. Specifically, for each gene in the pair, we first utilize its multimodal features to learn its embedding representation through a corresponding modality-specific encoder φ_m (where $m \in \{o, k\}$ represents the omics or knowledge graph modality):

$$\mathbf{h}_i^m = \varphi_m(\mathbf{f}_i^m), \quad \mathbf{h}_j^m = \varphi_m(\mathbf{f}_j^m) \tag{2}$$

where $\mathbf{f}_i^m$ denotes the feature input corresponding to the m-th modality for the gene g_i.

The obtained embedding vectors of the two genes are subsequently used as input features to output the final prediction logits through a shared classifier $\mathcal{F}_m$:

$$\hat{y}_{i,j}^m = \mathcal{F}_m([\mathbf{h}_i^m; \mathbf{h}_j^m]) \tag{3}$$

where $[\cdot; \cdot]$ denotes the concatenation operation.

We aim to optimize the model by minimizing the Binary Cross-Entropy (BCE) loss, which is formulated as:

$$\mathcal{L}_{\text{BCE}} = -\frac{1}{|\mathcal{D}|} \sum_{(g_i, g_j) \in \mathcal{D}} \left(y_{i,j} \log(\varsigma(\hat{y}_{i,j}^m)) + (1 - y_{i,j}) \log(1 - \varsigma(\hat{y}_{i,j}^m)) \right) \tag{4}$$

where $\mathcal{D}$ represents the set of all gene pairs within the training set, $y_{i,j}$ denotes the ground truth annotation, and $\varsigma(\cdot)$ serves as the sigmoid activation function for transforming logits into probability scores.

We denote the predicted probability after sigmoid transformation as $p_{i,j}^m = \varsigma(\hat{y}_{i,j}^m)$, which represents the probability that the gene pair (g_i, g_j) is predicted to have a synthetic lethality relationship under modality m.

5.2.2 Omics Encoder with VAE Early Fusion

The encoding of the omics modality adopts an early fusion module (OmicsEncoder, φ_o). We constructed an $N \times N$ VAE encoder matrix (where $N = 4$ denotes the total count of omics categories) and leveraged the Product of Experts (PoE) [37] mechanism to execute early fusion on the omics features. Each VAE encoder is implemented by a Multi-Layer Perceptron (MLP), which maps input features to the mean $\boldsymbol{\mu}$ and log-variance $\log \boldsymbol{\sigma}^2$ in the latent space.

Let the N types of omics features for gene g_i be denoted as $\mathbf{F}_i = \{\mathbf{f}_i^{(1)}, \ldots, \mathbf{f}_i^{(N)}\}$. As shown in the Omics Encoder module of Figure 1, the encoder matrix consists of self-encoders (Self-VAE) on the diagonal and cross-encoders (Cross-VAE) off the diagonal. For the k-th omics data $\mathbf{f}_i^{(k)}$ of gene i, the self-encoder $\text{VAE}_{k,k}$ maps it to the parameters of the latent distribution, $\boldsymbol{\mu}_i^{(k,\text{self})}$ and $\log(\boldsymbol{\sigma}_i^{(k,\text{self})})^2$. For the j-th ($j \neq k$) omics data $\mathbf{f}_i^{(j)}$ of gene i, the cross-encoder $\text{VAE}_{k,j}$ attempts to infer the latent distribution of the k-th omics type, with parameters denoted as $\boldsymbol{\mu}_i^{(k,\text{cross},j)}$ and $\log(\boldsymbol{\sigma}_i^{(k,\text{cross},j)})^2$.

To aggregate features from different perspectives and address the deficiency of missing modalities for certain genes, we utilize the PoE mechanism to calculate a joint posterior distribution for each target omics type k of gene i, where the parameters $(\boldsymbol{\mu}_i^{(k,\text{PoE})}, \boldsymbol{\sigma}_i^{(k,\text{PoE})})$ are given by the following formulas:

$$\frac{1}{(\boldsymbol{\sigma}_i^{(k,\text{PoE})})^2} = \sum_{j \neq k} \frac{1}{(\boldsymbol{\sigma}_i^{(k,\text{cross},j)})^2}, \quad \boldsymbol{\mu}_i^{(k,\text{PoE})} = (\boldsymbol{\sigma}_i^{(k,\text{PoE})})^2 \sum_{j \neq k} \frac{\boldsymbol{\mu}_i^{(k,\text{cross},j)}}{(\boldsymbol{\sigma}_i^{(k,\text{cross},j)})^2} \tag{5}$$

Then, we sample from the posterior distributions of the self-encoder path and the PoE path using the reparameterization trick to obtain the respective sets of latent variables $\mathbf{Z}_i^{\text{self}} = \{\mathbf{z}_i^{(k,\text{self})}\}_{k=1}^N$ and $\mathbf{Z}_i^{\text{PoE}} = \{\mathbf{z}_i^{(k,\text{PoE})}\}_{k=1}^N$:

$$\mathbf{z}_i^{(k,\text{path})} = \boldsymbol{\mu}_i^{(k,\text{path})} + \boldsymbol{\sigma}_i^{(k,\text{path})} \odot \boldsymbol{\epsilon}, \quad \text{path} \in \{\text{self}, \text{PoE}\} \tag{6}$$

where $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{I})$ denotes noise sampled from a standard normal distribution.

Finally, the latent variable sets from the two paths are subjected to mean pooling respectively and then concatenated. They are then projected via a fully connected layer to yield the final omics embedding $\mathbf{h}_i^o$:

$$\mathbf{h}_i^o = \text{FC}\left(\left[\frac{1}{N}\sum_{k=1}^N \mathbf{z}_i^{(k,\text{self})}; \frac{1}{N}\sum_{k=1}^N \mathbf{z}_i^{(k,\text{PoE})}\right]\right) \tag{7}$$

The training process adopts a Variational Information Bottleneck (VIB) loss function [50], which includes reconstruction error and a KL divergence regularization term:

$$\mathcal{L}_{\text{omics}} = \mathcal{L}_{\text{BCE}} + \lambda_{\text{self}} \sum_k D_{KL}(q_{\text{self}}^{(k)} || p) + \lambda_{\text{cross}} \sum_k D_{KL}(q_{\text{PoE}}^{(k)} || p) \tag{8}$$

where p is the standard normal prior distribution $\mathcal{N}(0, \mathbf{I})$. The weights of the two KL divergences, λ_{self} and λ_{cross}, are set to 0.1 and 0.5, respectively.

5.2.3 Knowledge Graph Encoder with RGCN

For the KG modality, we adopt RGCN [38] as the encoder (KGEncoder, φ_k) to capture the topological structures and semantic relationships of genes in the biological network.

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{R})$ be a knowledge graph, comprising the collections of entities $\mathcal{V}$, edges $\mathcal{E}$, and relations $\mathcal{R}$. Regarding a specific target gene g_i (corresponding to graph node i), we first employ an L-hop subgraph sampling strategy to extract its local neighborhood subgraph to reduce computational complexity. The initial node features are obtained through an Embedding Layer. The RGCN layer updates node representations by aggregating neighbor information under different relationship types $r \in \mathcal{R}$:

$$\mathbf{h}_i^{(l+1)} = \mathrm{ReLU}\left(\sum_{r\in\mathcal{R}}\sum_{j\in\mathcal{N}_i^r}\frac{1}{c_{i,r}}\mathbf{W}_r^{(l)}\mathbf{h}_j^{(l)} + \mathbf{W}_0^{(l)}\mathbf{h}_i^{(l)}\right) \tag{9}$$

where $\mathcal{N}_i^r$ constitutes the neighbor set for node i associated with relationship r, while $c_{i,r}$ serves as a normalization factor equal to $|\mathcal{N}_i^r|$.

After multi-layer RGCN aggregation, we extract the final representation $\mathbf{h}_i^k$ of the central gene node. Optimization of the knowledge graph branch is likewise achieved through the minimization of the binary cross-entropy loss, denoted as $\mathcal{L}_{\mathrm{kg}} = \mathcal{L}_{\mathrm{BCE}}$.

5.2.4 Dual-Stage Multimodal Fusion Strategy

The multimodal fusion adopts a dual-stage training strategy, which includes two complementary schemes named UMT and UME [39]. In the first stage, we independently train the omics prediction model (Only Omics) and the knowledge graph prediction model (Only KG). Let the omics teacher encoder obtained from the first stage training be φ_o^T, and the knowledge graph teacher encoder be φ_k^T. This stage ensures that each single-modal encoder can fully explore the feature representations within its modality without interference from other modalities. In the second stage, we designed two complementary fusion strategies and dynamically identified the optimal scheme according to the results observed in the validation subset.

(1) **Uni-Modal Teacher (UMT)**: This strategy employs a Knowledge Distillation framework to address the issue of modality laziness and guarantee that the multimodal architecture comprehensively captures the unimodal feature representations specific to each modality. We freeze the single-modal encoders pre-trained in the first stage as the "Teacher" and initialize a new multimodal model as the "Student", which includes an omics student encoder φ_o^S and a knowledge graph student encoder φ_k^S. While learning the classification task, the student model is also required to simulate the intermediate feature representations of the teacher model by minimizing the distillation loss (Mean Squared Error, MSE). The overall loss function is formulated as follows:

$$\mathcal{L}_{\mathrm{UMT}} = \mathcal{L}_{\mathrm{BCE}} + \lambda_{\mathrm{distill}}\sum_{m\in\{o,k\}}\|\mathbf{h}_i^{m,S} - \mathbf{h}_i^{m,T}\|^2 + \mathcal{L}_{\mathrm{KL}} \tag{10}$$

Here, $\mathbf{h}_i^{o,T}$ and $\mathbf{h}_i^{k,T}$ are the omics and graph embeddings output by the frozen teacher encoders for gene i, respectively. While $\mathbf{h}_i^{o,S}$ and $\mathbf{h}_i^{k,S}$ are the corresponding outputs of the student encoders.

The classifier $\mathcal{F}_{\text{UMT}}$ receives the genomic and knowledge graph features $[\mathbf{h}_i^{o,S}; \mathbf{h}_i^{k,S}]$ extracted by the student encoders to make predictions. The hyperparameter λ_{distill} is set to 1, and this feature-level distillation forces each branch of the student model to maintain strong feature extraction capabilities, which effectively alleviates modality laziness.

(2) **Uni-Modal Ensemble (UME)**: As a simpler late fusion baseline, UME directly integrates the prediction results of the two pre-trained models from the first stage. The final predicted probability p_{UME} is the average of the output probabilities from the two single-modal models:

$$p_{\text{UME}} = \frac{1}{2} \left(\varsigma(\hat{y}_o) + \varsigma(\hat{y}_k) \right) \tag{11}$$

UME requires no additional training and completely avoids gradient interference issues during joint training, making it more robust in cases where there are significant differences between modalities.

Considering the differences in the effects of cross-modal interactions across different datasets, we adopted a data-driven adaptive selection strategy. After training is completed, we evaluate the AUC metrics of UMT and UME on the validation set separately and select the better-performing method as the final model:

$$p_{i,j} = \begin{cases} p_{i,j}^{\text{UMT}}, & \text{if } \text{AUC}_{\text{UMT}}^{\text{val}} > \text{AUC}_{\text{UME}}^{\text{val}} \\ p_{i,j}^{\text{UME}}, & \text{otherwise} \end{cases} \tag{12}$$

UMT typically performs better when both omics and knowledge graph single-modal features are strong and their interaction provides additional information. Conversely, when one modality is significantly dominant or when cross-modal interaction introduces noise, the simple ensemble strategy of UME proves to be more effective.

It is worth noting that compared to traditional complex fusion methods, our method is simple to implement and easy to tune. UMT requires tuning only one additional hyperparameter (the distillation loss weight λ_{distill}), while UME does not even require extra training. This simplicity not only improves the practicality of the method but also enhances its portability to new datasets.

Code availability

The data and code can be accessed at the following GitHub repository: https://github.com/Jmpax404/SynLeaF.

Acknowledgements

Author contributions

J.L. conceived and designed the project and supervised the work. Z.X. developed the methods, performed bioinformatics analysis and drafted the manuscript. S.Z., R.W., and R.H. prepared the data and performed the benchmarks. S.Z., S.C., Y.H., and J.M. contributed to the benchmarks. Y.C., X.W., and Y.W. participated in project design and coordination.

Competing interests

The authors declare no competing interests.

References

[1] Huang, A., Garraway, L.A., Ashworth, A., Weber, B.: Synthetic lethality as an engine for cancer drug target discovery. Nature Reviews Drug Discovery **19**(1), 23–38 (2020) https://doi.org/10.1038/s41573-019-0046-z

[2] Ashworth, A., Lord, C.J.: Synthetic lethal therapies for cancer: what's next after PARP inhibitors? Nature Reviews Clinical Oncology **15**(9), 564–576 (2018) https://doi.org/10.1038/s41571-018-0055-6

[3] Wang, J., Zhang, Q., Han, J., Zhao, Y., Zhao, C., Yan, B., Dai, C., Wu, L., Wen, Y., Zhang, Y., Leng, D., Wang, Z., Yang, X., He, S., Bo, X.: Computational methods, databases and tools for synthetic lethality prediction. Briefings in Bioinformatics **23**(3), 106 (2022) https://doi.org/10.1093/bib/bbac106

[4] Topatana, W., Juengpanich, S., Li, S., Cao, J., Hu, J., Lee, J., Suliyanto, K., Ma, D., Zhang, B., Chen, M., Cai, X.: Advances in synthetic lethality for cancer therapy: cellular mechanism and clinical translation. Journal of Hematology & Oncology **13**(1), 118 (2020) https://doi.org/10.1186/s13045-020-00956-5

[5] Lord, C.J., Ashworth, A.: PARP inhibitors: Synthetic lethality in the clinic. Science **355**(6330), 1152–1158 (2017) https://doi.org/10.1126/science.aam7344

[6] Milordini, G., Zacco, E., Masi, M., Armaos, A., Di Palma, F., Oneto, M., Gilodi, M., Rupert, J., Broglia, L., Varignani, G., Scotto, M., Marotta, R., Girotto, S., Cavalli, A., Tartaglia, G.G.: Computationally-designed aptamers targeting rad51-brca2 interaction impair homologous recombination and induce synthetic lethality. Nature Communications (2025) https://doi.org/10.1038/s41467-025-66694-9

[7] Chabanon, R.M., Shcherbakova, L., Lacroix-Triki, M., Aglave, M., Zeghondy, J., Kriaa, V., Gougé, A., Garrido, M., Edmond, E., Bigot, L., Krastev, D.B., Brough, R., Pettitt, S.J., Thomas-Bonafos, T., Samstein, R., Massard, C., Deloger, M., Tutt, A.N., Barlesi, F., Loriot, Y., Delaloge, S., Tawk, M., Degerny, C., Lin, Y.-L., Pistilli, B., Pasero, P., Lord, C.J., Postel-Vinay, S.: Autocrine interferon poisoning

mediates adar1-dependent synthetic lethality in brca1/2-mutant cancers. Nature Communications **16**, 6972 (2025) https://doi.org/10.1038/s41467-025-62309-5

[8] Gonçalves, E., Ryan, C.J., Adams, D.J.: Synthetic lethality in cancer drug discovery: challenges and opportunities. Nature Reviews Drug Discovery **25**, 22–38 (2026) https://doi.org/10.1038/s41573-025-01273-7

[9] O'Neil, N.J., Bailey, M.L., Hieter, P.: Synthetic lethality and cancer. Nature Reviews Genetics **18**(10), 613–623 (2017) https://doi.org/10.1038/nrg.2017.47

[10] Hao, Z., Wu, D., Fang, Y., Wu, M., Cai, R., Li, X.: Prediction of Synthetic Lethal Interactions in Human Cancers Using Multi-View Graph Auto-Encoder. IEEE Journal of Biomedical and Health Informatics **25**(10), 4041–4051 (2021) https://doi.org/10.1109/JBHI.2021.3079302

[11] Fath, M.K., Najafiyan, B., Morovatshoar, R., Khorsandi, M., Dashtizadeh, A., Kiani, A., Farzam, F., Kazemi, K.S., Afjadi, M.N.: Potential promising of synthetic lethality in cancer research and treatment. Naunyn-Schmiedeberg's Archives of Pharmacology **398**(2), 1403–1431 (2025) https://doi.org/10.1007/s00210-024-03444-6

[12] Horlbeck, M.A., Xu, A., Wang, M., Bennett, N.K., Park, C.Y., Bogdanoff, D., Adamson, B., Chow, E.D., Kampmann, M., Peterson, T.R., Nakamura, K., Fischbach, M.A., Weissman, J.S., Gilbert, L.A.: Mapping the Genetic Landscape of Human Cells. Cell **174**(4), 953–96722 (2018) https://doi.org/10.1016/j.cell.2018.06.010

[13] Nijman, S.M.B.: Synthetic lethality: general principles, utility and detection using genetic screens in human cells. FEBS Letters **585**(1), 1–6 (2011) https://doi.org/10.1016/j.febslet.2010.11.024

[14] Das, S., Deng, X., Camphausen, K., Shankavaram, U.: DiscoverSL: an R package for multi-omic data driven prediction of synthetic lethality in cancers. Bioinformatics **35**(4), 701–702 (2019) https://doi.org/10.1093/bioinformatics/bty673

[15] Wan, F., Li, S., Tian, T., Lei, Y., Zhao, D., Zeng, J.: EXP2SL: A Machine Learning Framework for Cell-Line-Specific Synthetic Lethality Prediction. Frontiers in Pharmacology **11**, 112 (2020) https://doi.org/10.3389/fphar.2020.00112

[16] Huang, J., Wu, M., Lu, F., Ou-Yang, L., Zhu, Z.: Predicting synthetic lethal interactions in human cancers using graph regularized self-representative matrix factorization. BMC Bioinformatics **20**(Suppl 19), 657 (2019) https://doi.org/10.1186/s12859-019-3197-3

[17] Liu, Y., Wu, M., Liu, C., Li, X.-L., Zheng, J.: SL2MF: Predicting Synthetic Lethality in Human Cancers via Logistic Matrix Factorization. IEEE/ACM

Transactions on Computational Biology and Bioinformatics **17**(3), 748–757 (2020) https://doi.org/10.1109/TCBB.2019.2909908

[18] Hamilton, W.L.: Graph Representation Learning. Synthesis Lectures on Artificial Intelligence and Machine Learning, vol. 14, pp. 1–159. Springer, Cham (2020)

[19] Cai, R., Chen, X., Fang, Y., Wu, M., Hao, Y.: Dual-dropout graph convolutional network for predicting synthetic lethality in human cancers. Bioinformatics **36**(16), 4458–4465 (2020) https://doi.org/10.1093/bioinformatics/btaa211

[20] Long, Y., Wu, M., Liu, Y., Zheng, J., Kwoh, C.K., Luo, J., Li, X.: Graph contextualized attention network for predicting synthetic lethality in human cancers. Bioinformatics **37**(16), 2432–2440 (2021) https://doi.org/10.1093/bioinformatics/btab110

[21] Ye, Q., Hsieh, C.-Y., Yang, Z., Kang, Y., Chen, J., Cao, D., He, S., Hou, T.: A unified drug-target interaction prediction framework based on knowledge graph and recommendation system. Nature Communications **12**(1), 6775 (2021) https://doi.org/10.1038/s41467-021-27137-3

[22] Wang, S., Xu, F., Li, Y., Wang, J., Zhang, K., Liu, Y., Wu, M., Zheng, J.: KG4SL: knowledge graph neural network for synthetic lethality prediction in human cancers. Bioinformatics **37**(Supplement_1), 418–425 (2021) https://doi.org/10.1093/bioinformatics/btab271

[23] Liu, X., Yu, J., Tao, S., Yang, B., Wang, S., Wang, L., Bai, F., Zheng, J.: PiLSL: pairwise interaction learning-based graph neural network for synthetic lethality prediction in human cancers. Bioinformatics **38**(Supplement_2), 106–112 (2022) https://doi.org/10.1093/bioinformatics/btac476

[24] Zhu, Y., Zhou, Y., Liu, Y., Wang, X., Li, J.: SLGNN: synthetic lethality prediction in human cancers based on factor-aware knowledge graph neural network. Bioinformatics **39**(2), 015 (2023) https://doi.org/10.1093/bioinformatics/btad015

[25] Zhang, K., Wu, M., Liu, Y., Feng, Y., Zheng, J.: KR4SL: knowledge graph reasoning for explainable prediction of synthetic lethality. Bioinformatics **39**(Supplement_1), 158–167 (2023) https://doi.org/10.1093/bioinformatics/btad261

[26] Zhang, G., Chen, Y., Yan, C., Wang, J., Liang, W., Luo, J., Luo, H.: MPASL: multi-perspective learning knowledge graph attention network for synthetic lethality prediction in human cancer. Frontiers in Pharmacology **15**, 1398231 (2024) https://doi.org/10.3389/fphar.2024.1398231

[27] Long, Y., Wu, M., Liu, Y., Fang, Y., Kwoh, C.K., Chen, J., Luo, J., Li, X.: Pre-training graph neural networks for link prediction in biomedical networks.

Bioinformatics **38**(8), 2254–2262 (2022) https://doi.org/10.1093/bioinformatics/btac100

[28] Li, J., Lu, X., Jiang, K., Tang, D., Ning, B., Sun, F.: TARSL: Triple-Attention Cross-Network Representation Learning to Predict Synthetic Lethality for Anti-Cancer Drug Discovery. IEEE Journal of Biomedical and Health Informatics **29**(3), 1680–1691 (2025) https://doi.org/10.1109/JBHI.2023.3306768

[29] Huang, Y., Yuan, R., Li, Y., Xing, Z., Li, J.: Struct2SL: Synthetic lethality prediction based on AlphaFold2 structure information and Multilayer Perceptron. Computational and Structural Biotechnology Journal **27**, 1570–1577 (2025) https://doi.org/10.1016/j.csbj.2025.04.012

[30] Previtali, V., Bagnolini, G., Ciamarone, A., Ferrandi, G., Rinaldi, F., Myers, S.H., Roberti, M., Cavalli, A.: New Horizons of Synthetic Lethality in Cancer: Current Development and Future Perspectives. Journal of Medicinal Chemistry **67**(14), 11488–11521 (2024) https://doi.org/10.1021/acs.jmedchem.4c00113

[31] Shen, J.P., Zhao, D., Sasik, R., Luebeck, J., Birmingham, A., Bojorquez-Gomez, A., Licon, K., Klepper, K., Pekin, D., Beckett, A.N., Sanchez, K.S., Thomas, A., Kuo, C.-C., Du, D., Roguev, A., Lewis, N.E., Chang, A.N., Kreisberg, J.F., Krogan, N., Qi, L., Ideker, T., Mali, P.: Combinatorial CRISPR–Cas9 screens for de novo mapping of genetic interactions. Nature Methods **14**(6), 573–576 (2017) https://doi.org/10.1038/nmeth.4225

[32] Tang, S., Gökbağ, B., Fan, K., Shao, S., Huo, Y., Wu, X., Cheng, L., Li, L.: Synthetic lethal gene pairs: Experimental approaches and predictive models. Frontiers in Genetics **13**, 961611 (2022) https://doi.org/10.3389/fgene.2022.961611

[33] Ryan, C.J., Bajrami, I., Lord, C.J.: Synthetic Lethality and Cancer - Penetrance as the Major Barrier. Trends in Cancer **4**(10), 671–683 (2018) https://doi.org/10.1016/j.trecan.2018.08.003

[34] Tepeli, Y.I., Seale, C., Gonçalves, J.P.: ELISL: early–late integrated synthetic lethality prediction in cancer. Bioinformatics **40**(1), 764 (2023) https://doi.org/10.1093/bioinformatics/btad764

[35] Chen, J., Pan, J., Zhu, Y., Li, J.: SLGNNCT: Synthetic lethality prediction based on knowledge graph for different cancers types. In: International Conference on Intelligent Computing, pp. 159–170 (2024)

[36] Tu, X., Cao, Z.-J., Xia, C., Mostafavi, S., Gao, G.: Cross-linked unified embedding for cross-modality representation learning. In: Koyejo, S., Mohamed, S., Agarwal, A., Belgrave, D., Cho, K., Oh, A. (eds.) Advances in Neural Information Processing Systems, vol. 35, pp. 15942–15955. Curran Associates, Inc., Red Hook, NY (2022)

[37] Kutuzova, S., Krause, O., McCloskey, D., Nielsen, M., Igel, C.: Multimodal variational autoencoders for semi-supervised learning: In defense of product-of-experts. arXiv, 2101–07240 (2021)

[38] Schlichtkrull, M., Kipf, T.N., Bloem, P., Berg, R., Titov, I., Welling, M.: Modeling relational data with graph convolutional networks. arXiv, 1703–06103 (2017)

[39] Du, C., Teng, J., Li, T., Liu, Y., Yuan, T., Wang, Y., Yuan, Y., Zhao, H.: On uni-modal feature learning in supervised multi-modal learning. Computer Vision and Pattern Recognition (2023)

[40] Wang, F.-A., Zhuang, Z., Gao, F., He, R., Zhang, S., Wang, L., Liu, J., Li, Y.: Tmo-net: an explainable pretrained multi-omics model for multi-task learning in oncology. Genome Biology **25**(1), 149 (2024)

[41] Subbaswamy, A., Saria, S.: From development to deployment: dataset shift, causality, and shift-stable models in health ai. Biostatistics **21**(2), 345–352 (2020) https://doi.org/10.1093/biostatistics/kxz041

[42] Peng, X., Wei, Y., Deng, A., Wang, D., Hu, D.: Balanced multimodal learning via on-the-fly gradient modulation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 18228–18237 (2022). https://doi.org/10.1109/CVPR52688.2022.01772

[43] Wang, Z., Dai, Z., Poczos, B., Carbonell, J.: Characterizing and avoiding negative transfer. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 11285–11294 (2019). https://doi.org/10.1109/CVPR.2019.01155

[44] Cerami, E., Gao, J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., Jacobsen, A., Byrne, C.J., Heuer, M.L., Larsson, E., Antipin, Y., Reva, B., Goldberg, A.P., Sander, C., Schultz, N.: The cbio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discovery **2**(5), 401–404 (2012)

[45] Consortium, T.U.: Uniprot: the universal protein knowledgebase in 2021. Nucleic Acids Research **49**(D1), 480–489 (2020)

[46] Lee, J.S., Das, A., Jerby-Arnon, L., Arafeh, R., Auslander, N., Davidson, M., McGarry, L., James, D., Amzallag, A., Park, S.G., Cheng, K., Robinson, W., Atias, D., Stossel, C., Buzhor, E., Stein, G., Waterfall, J.J., Meltzer, P.S., Golan, T., Hannenhalli, S., Gottlieb, E., Benes, C.H., Samuels, Y., Shanks, E., Ruppin, E.: Harnessing synthetic lethality to predict the response to cancer treatment. Nature Communications **9**(1), 2546 (2018) https://doi.org/10.1038/s41467-018-04647-1

[47] Cerami, E., Gao, J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., Jacobsen, A., Byrne, C.J., Heuer, M.L., Larsson, E., Antipin, Y., Reva, B., Goldberg, A.P., Sander, C., Schultz, N.: The cbio cancer genomics portal: An open platform for exploring multidimensional cancer genomics data. Cancer Discovery **2**(5), 401–404 (2012) https://doi.org/10.1158/2159-8290.CD-12-0095

[48] Kaelin, W.G.: The concept of synthetic lethality in the context of anticancer therapy. Nature Reviews Cancer **5**(9), 689–698 (2005) https://doi.org/10.1038/nrc1691

[49] Chicco, D.: Siamese neural networks: An overview. In: Cartwright, H. (ed.) Artificial Neural Networks, pp. 73–94. Springer, New York, NY (2021). https://doi.org/10.1007/978-1-0716-0826-5_3

[50] Alemi, A.A., Fischer, I., Dillon, J.V., Murphy, K.: Deep variational information bottleneck. arXiv, 1612–00410 (2016)