A Two-Phase Contrastive Learning Framework for Multi-Modal Ovarian Cancer Prognosis with Explainable AI

Lakshmi C R[1]*, D.V. Venkata Vara Prasad[1], Y.V. Lokeswari[1], and V. Ramesh[2]

[1] Dept. of CSE, Sri Sivasubramaniya Nadar College of Engineering, Kalavakkam, India
lakshmicr2420099@ssn.edu.in*, dvvprasad@ssn.edu.in, lokeswariyv@ssn.edu.in

[2] Dept. of CSA, SCSVMV University, Kanchipuram, India
vramesh@kanchiuniv.ac.in

Abstract. Ovarian cancer remains one of the most lethal gynecologic malignancies due to late diagnosis and frequent recurrence after treatment. Accurate prediction of both survival and recurrence risk is essential for personalised clinical decision-making. This research work proposes a Multi-Modal Supervised Contrastive Learning framework (MMSCL) that integrates clinical variables, multi-omics data, and CT imaging from 92 patients in the TCGA-OV cohort. The model is trained in two phases. First, supervised contrastive pre-training structures the embedding space using a dual-label pairing strategy based on survival and recurrence outcomes. Second, multi-task fine-tuning simultaneously predicts survival status, recurrence risk, and time-to-recurrence. Dimensionality reduction using PCA and mutual-information-based feature selection reduces noise in the small cohort. Under five-fold cross-validation, MMSCL achieves a mean combined AUC of 0.794±0.040. On an independent hold-out set, the framework obtains a survival AUC of 0.8750 and a Harrell's C-Index of 0.7091. SHAP analysis highlights tumour stage and molecular subtype features as the most influential predictors. The proposed framework demonstrates the potential of multimodal contrastive learning for interpretable cancer prognosis in small clinical cohorts.

Keywords: Ovarian cancer, survival analysis, recurrence prediction, multi-omics, supervised contrastive learning, multi-modal fusion, contrastive learning

1 Introduction

Ovarian cancer is one of the deadliest gynecologic malignancies, with most cases diagnosed at advanced stages (Stage III or IV). Late detection limits treatment options and leads to poor survival outcomes. The standard treatment - cytoreductive surgery followed by platinum-based chemotherapy - offers limited protection against relapse; nearly 70% of patients experience recurrence within three years of diagnosis [1]. This high recurrence rate, along with strong molecular

R. Vasanth Kumar Mehta et al. (eds.), *Proceedings of the International Conference on Intelligent Systems for a Sustainable Future (ISSF 2026)*, Atlantis Highlights in Intelligent Systems 16,
https://doi.org/10.2991/978-94-6239-693-7_29

heterogeneity across patients, creates a need for computational models that can jointly predict recurrence risk and survival probability.

High-throughput sequencing enables multi-omics profiling across genomics, transcriptomics, proteomics, and epigenomics, each capturing different aspects of tumour behaviour. However, building reliable prognostic models remains challenging due to high dimensionality, varying feature scales, small cohort sizes, and experimental noise, which often lead to overfitting [2, 3]. Dimensionality reduction using Principal Component Analysis (PCA) and feature selection using Mutual Information (MI) are therefore essential to retain relevant signals while reducing noise.

To address these challenges, this work proposes a supervised contrastive learning framework for ovarian cancer prognosis that integrates clinical data, multi-omics protein expression, and CT imaging. A dual-objective pair generation strategy, based on both recurrence and survival labels, enables the model to learn clinically meaningful patient similarities [11]. The framework achieves a survival AUC of 0.8750 and a Harrell's C-Index of 0.7091 on the held-out set, outperforming conventional baselines.

The main contributions are as follows. A two-phase pipeline first structures the embedding space through contrastive pre-training, followed by multi-task fine-tuning. A dual-label pairing strategy enhances supervision compared to single-endpoint approaches. PCA and MI reduce feature dimensionality to 13 key clinical variables, improving generalisation on the 92-patient cohort. SHAP-based interpretability links predictions to known biomarkers. An ablation study evaluates the contribution of each component, and a three-tier risk stratification module translates predictions into clinically actionable treatment groups.

2 Related Works and Literature Survey

Recent studies have explored multi-omics integration, contrastive learning, graph neural networks, and deep learning for cancer prognosis and biomarker discovery.

Singh et al. [1] proposed a multi-level ensemble combining CNN and handcrafted features from histopathology images, achieving high accuracy but relying only on imaging data. Zhang et al. [2] introduced a multi-omics framework using autoencoders and attention for feature fusion, though it is computationally expensive and does not address survival prediction. Wang et al. [3] proposed SMMGCL for spatial multi-omics integration, but it focuses on cell-level analysis and requires spatial transcriptomics data. Yang et al. [4] developed a hierarchical contrastive clustering model that improves clustering performance but does not handle prognostic tasks.

Chen et al. [11] introduced MoSCHG, a supervised contrastive model enabling effective cross-modal embedding alignment. Kim et al. [14] analysed proteomics for ovarian cancer prognosis without multi-modal integration. Xiao et al. [6] incorporated prior knowledge into graph neural networks, improving

generalisation on small datasets. Mohamed and Ezugwu [22] demonstrated improved cancer prediction using multi-omics with deep learning, while Boehm et al. [24] integrated radiomics, histopathology, and genomics with strong performance but high data requirements.

Liu et al. [8] proposed representation learning using heterogeneous gene regulatory networks. Wen et al. [7] introduced SCDMSC for multi-view subspace clustering. Xie et al. [9] developed a multi-task graph contrastive framework for cancer driver gene identification, and Liu and Chen [10] proposed collaborative attention contrast learning for subtype identification. Huang et al. [12] demonstrated the benefit of integrating biological knowledge with graph convolutional networks. Zhang et al. [13], Yao and Tian [15], and Amjad et al. [16] focused on multi-omics-based drug response prediction, while Periyasamy [17] highlighted AI-driven multi-omics integration for drug discovery. Jhee et al. [18] demonstrated the effectiveness of transformer-based gene scoring models.

Elangovan and Nagarajan [19] proposed a multi-omics model for early cancer prediction, and Akter and Akhter [20] provided an early machine learning baseline for ovarian cancer diagnosis. Li et al. [21] introduced scGCL for single-cell multi-omics clustering. Sathyamoorthi et al. [26] integrated omic and non-omic data for survival prediction, while Liu et al. [25] applied multi-omics for immunotherapy response prediction. Khosla et al. [23] established the theoretical foundation of supervised contrastive learning.

Overall, existing methods lack a unified framework that integrates supervised contrastive learning with clinical, multi-omics, and imaging data for multi-task ovarian cancer prognosis on small real-world datasets. The proposed MMSCL framework addresses this gap.

3 Proposed Methodology

3.1 Overall System Architecture

The MMSCL framework processes three independent input modalities through dedicated encoder branches before combining them into a shared patient representation, as illustrated in Fig. 1.

The **clinical encoder** ingests the 13 MI-selected standardised features through a two-layer network producing a 64-dimensional embedding:

$$\mathbb{R}^{13} \rightarrow \text{FC}(64) \xrightarrow{\text{BN-ReLU-Drop}(0.15)} \text{FC}(64) \tag{1}$$

The **omics encoder** receives PCA-compressed features through:

$$\mathbb{R}^{k} \rightarrow \text{FC}(128) \xrightarrow{\text{BN-ReLU-Drop}(0.15)} \text{FC}(64) \tag{2}$$

The **image encoder** is a lightweight two-block CNN: the first block applies 16 filters (3×3) with BN, ReLU, and 2×2 pooling; the second applies 32 filters with

the same scheme. Adaptive average pooling reduces the map to 4×4, followed by two linear layers:

$$512 \rightarrow 128 \rightarrow 64 \tag{3}$$

with BN, ReLU, and Dropout(0.15), yielding the final 64-dimensional image embedding. Dropout is set to 0.15 across all encoders rather than the conventional 0.3, which over-regularised on this 92-patient cohort. During **Phase 1** (contrastive pre-training), each encoder output passes through a projection head (FC(64) $\rightarrow$ BN-ReLU $\rightarrow$ FC(32)) mapping embeddings onto the unit hypersphere. Positive pairs share both survival and recurrence labels; all others are negatives. During **Phase 2** (multi-task fine-tuning), the projection heads are discarded, the three 64-d embeddings are concatenated to 192-d, and a **fusion MLP** (FC(128) $\rightarrow$ BN-ReLU-Drop(0.15) $\rightarrow$ FC(64) $\rightarrow$ BN-ReLU) produces the final 64-d patient embedding decoded by three task heads.

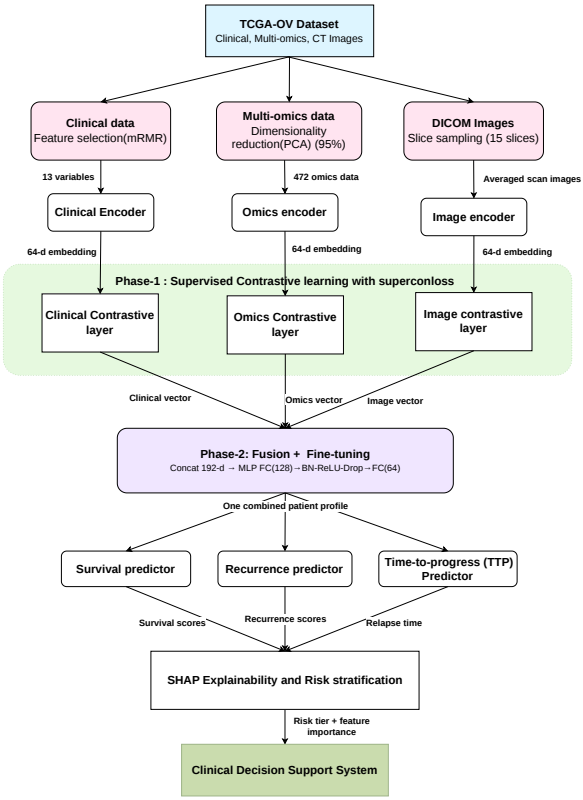

Fig. 1. Overall MMSCL architecture showing Phase 1 contrastive pre-training and Phase 2 multi-task fine-tuning.

3.2 Dataset Description

A processed multi-omics dataset from the TCGA-OV cohort was used, containing ovarian serous cystadenocarcinoma patients. After aggregating 2,409 raw records across 573 features by patient identifier, 93 unique patients were identified. One patient (TCGA-61-1727) lacked DICOM CT images and was excluded, yielding 92 patients. The dataset includes 472 protein expression features and three prediction targets: binary survival event, binary recurrence status, and continuous time-to-recurrence.

3.3 Pre-Processing Pipeline

Numerical attributes were imputed via median substitution; categorical features by mode replacement and label encoding. Continuous features were standardised to zero mean and unit variance (`StandardScaler`). Survival outcomes were derived from `DeadatFollowUp` (OS event) and `NewTumorEvent` (recurrence). Time-to-recurrence from `TTP` was converted from days to months.

3.4 Clinical Feature Selection via Mutual Information

Feature selection was performed on the raw unscaled matrix before normalisation to avoid distorting MI estimates. MI between each clinical feature X and survival label Y is given in Eq. (4):

$$\mathrm{MI}(X;\, Y) = \sum_{x} \sum_{y} p(x, y)\, \log \left[\frac{p(x, y)}{p(x)\, p(y)} \right] \tag{4}$$

The top **13** features were retained to satisfy the small-N guideline ($N/5 = 73/5 \approx 14$), preserving predominantly tumour stage and molecular subtype variables [6].

3.5 Dimensionality Reduction via PCA

PCA was applied to the scaled omics matrix to retain the fewest components explaining at least 95% of total variance (Eq. 5):

$$k = \min \left\{ j \;\middle|\; \frac{\sum_{i=1}^{j} \lambda_i}{\sum_{i=1}^{n} \lambda_i} \geq 0.95 \right\} \tag{5}$$

This reduced the omics space by approximately 89% while preserving dominant biological variation [2].

3.6 DICOM Image Pre-Processing

Up to **15 axial slices** per patient were sampled uniformly, linearly normalised to $[0, 255]$, resized to 128×128 via bilinear interpolation, and converted to RGB. The per-patient representation is the depth-axis average (Eq. 6):

$$\mathbf{I}_i = \frac{1}{S} \sum_{s=1}^{S} \mathrm{norm}(\mathrm{resize}(\mathrm{DICOM}_{is})) \tag{6}$$

3.7 Supervised Contrastive Loss

Contrastive learning principle. Contrastive learning structures the representation space by pulling similar samples together and pushing dissimilar ones apart. In the supervised variant [23], class labels define positive pairs. For each anchor i, the encoder maximises cosine similarity with positives and minimises it with negatives on the unit hypersphere, producing a structured latent space where clinically similar patients cluster together.

Dual-label positive-pair generation. MMSCL extends standard supervised contrastive learning with a *dual-label* criterion [11]: two patients form a positive pair only when they share both the **survival label** and the **recurrence label**, creating four outcome groups ({alive+no-recurrence, alive+recurrence, deceased+no-recurrence, deceased+recurrence}). This four-way discrimination prevents the encoder from conflating clinically distinct outcomes that a single label would merge.

The contrastive loss (Eq. 7) is applied to all three modality projections and summed:

$$\mathcal{L}_{\text{SCL}} = \sum_{i=1}^{N} \frac{-1}{|P(i)|} \sum_{p \in P(i)} \log \frac{\exp(\mathbf{z}_i \cdot \mathbf{z}_p / \tau)}{\sum_{a \in A(i)} \exp(\mathbf{z}_i \cdot \mathbf{z}_a / \tau)} \tag{7}$$

where $P(i)$ are positives sharing both labels with anchor i, $A(i)$ are all other batch samples, and $\tau = 0.07$. Gradient clipping at norm 1.0 stabilises training. Phase 1 runs for 20 epochs; contrastive loss falls from ≈ 15.9 to ≈ 11.2, confirming stable alignment.

3.8 Multi-Task Optimisation

The Phase 2 total loss is (Eq. 8):

$$\mathcal{L}_{\text{total}} = 0.7\,\mathcal{L}_{\text{survival}} + 1.2\,\mathcal{L}_{\text{recurrence}} + 0.3\,\mathcal{L}_{\text{TTP}} + 0.8\,\mathcal{L}_{\text{Cox}} \tag{8}$$

Recurrence receives the highest weight (1.2×) given its direct clinical impact on treatment escalation. Survival is weighted at 0.7× to avoid dominating gradient updates; TTP at 0.3× for a meaningful regression signal; and the auxiliary Cox term at 0.8× to align risk ranking with Harrell's C-index. The Cox partial likelihood (Breslow approximation, Eq. 9) is computed when a batch contains at least two events:

$$\mathcal{L}_{\text{Cox}} = -\frac{1}{|\mathcal{E}|} \sum_{i \in \mathcal{E}} \left[\log r_i - \log \sum_{j:\, t_j \geq t_i} \exp(r_j) \right] \tag{9}$$

Label smoothing ($\varepsilon = 0.08$) and class-weighted binary cross-entropy address logit saturation and class imbalance respectively. AdamW uses weight decay 1×10^{-4} and learning rate 3×10^{-4}, decayed via cosine annealing to 10^{-6} over $T_{\max} = 150$ epochs.

3.9 Prediction Heads and Test-Time Augmentation

The 64-d fusion embedding feeds three independent linear heads ($\mathbb{R}^{64} \to \mathbb{R}^1$). Sigmoid is applied at inference for the survival and recurrence classifiers; the TTP regressor uses no activation. At inference, 10 forward passes are run per patient: the first uses clean inputs; passes 2–10 add Gaussian noise ($\sigma = 0.01$) to clinical and omics vectors. Averaging the outputs reduces prediction variance by $\sqrt{10}$.

3.10 SHAP Explainability

SHAP was applied via a `KernelExplainer` with 10 k-means background samples and 100 coalition evaluations per patient. The SHAP value ϕ_i quantifies the marginal contribution of feature i to the survival prediction (Eq. 10):

$$\phi_i(\mathbf{x}) = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!\,(|F| - |S| - 1)!}{|F|!} \big[f(S \cup \{i\}) - f(S) \big] \tag{10}$$

Global importance is computed as the mean absolute SHAP value across the test set (Fig. 2).

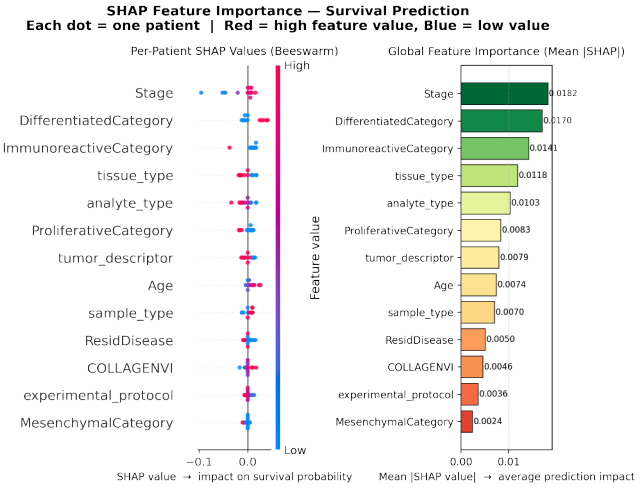

Fig. 2. SHAP feature importance for the survival prediction head. Left: raw mean $|\phi_i|$. Right: normalised % contribution.

3.11 Risk Stratification

Model outputs are converted into clinical tiers using separate thresholds for $\hat{y}_r$ and $\hat{y}_s$ rather than a composite score, since survival and recurrence are partially

independent. A patient with $\hat{y}_s = 0.97$ and $\hat{y}_r = 0.80$ would appear low-risk under a weighted average yet requires aggressive management; independent thresholds prevent this masking. Thresholds 0.60 and 0.40 correspond to the upper and lower thirds of the sigmoid range; the Medium tier captures uncertain cases. High and Medium tiers use OR logic (either alarming indicator suffices to escalate); Low requires both to be favourable (AND logic). Protocols are summarised in Table 1.

Table 1. Three-Tier Risk Stratification with Clinical Protocols

Risk	$\hat{y}_r$	$\hat{y}_s$	**Protocol**
High	≥ 0.60	≤ 0.40	Aggressive platinum + dose-dense + PARP inhibitors + HIPEC; CA-125 every 2 months
Medium	0.40–0.59	0.41–0.60	Standard Carboplatin/Paclitaxel ×6 + Bevacizumab; CT scan every 6 months
Low	< 0.40	> 0.60	Adjuvant chemo ± fertility-sparing surgery; CA-125 every 3–6 months + annual follow-up

4 Experiment and Analysis

4.1 Experimental Setup

All experiments were conducted on an Intel Core i7 CPU using PyTorch 2.x with four optimised threads. The dataset contains 92 matched TCGA-OV patients, split into 73 for five-fold cross-validation and 19 for hold-out testing. The test set was not used during training.

Key settings include embedding size 64; contrastive pre-training for 20 epochs (Adam, $\eta = 10^{-3}$, weight decay 5×10^{-4}); fine-tuning up to 150 epochs with early stopping (AdamW, $\eta = 3 \times 10^{-4}$, weight decay 1×10^{-4}); batch size 8; $\tau = 0.07$; $\varepsilon = 0.08$; 13 clinical features; up to 15 CT slices; and Dropout 0.15. Early stopping is based on a 5-epoch moving average of validation AUC.

4.2 Evaluation Metrics

Model performance was evaluated using the Area Under the Receiver Operating Characteristic Curve (AUC) for classification tasks and Harrell's Concordance Index (C-Index) for survival analysis.

The AUC measures the model's ability to distinguish between positive and negative classes and is defined in Eq. (11):

$$\text{AUC} = \int_0^1 TPR(FPR^{-1}(x))\,dx \tag{11}$$

where TPR and FPR denote the true positive rate and false positive rate.

The Harrell's C-Index evaluates the ranking quality of predicted risk scores in survival analysis and is defined in Eq. (12):

$$C = \frac{1}{|\mathcal{P}|} \sum_{(i,j)\in\mathcal{P}} \mathbf{1}(r_i > r_j) \tag{12}$$

where $\mathcal{P}$ represents all comparable patient pairs and r_i denotes the predicted risk score.

Higher values of AUC and C-Index indicate better predictive performance.

4.3 Training Convergence

The supervised contrastive model demonstrated stable convergence across all five folds. During Phase 1, all five folds sustained exactly four valid batches per epoch; the summed three-modality contrastive loss decreased from $\approx$15.9 at epoch 1 to $\approx$11.2 at epoch 20, confirming progressive embedding alignment. Batches containing only patients with the same label combination contributed zero loss and were automatically skipped. During Phase 2, the total multi-task training loss decreased from approximately 1.75 at epoch 1. Early stopping triggered between epochs 36 and 65 depending on fold. Training and validation loss curves for the best fold (Fold 1, stop epoch 65) are shown in Fig. 3.

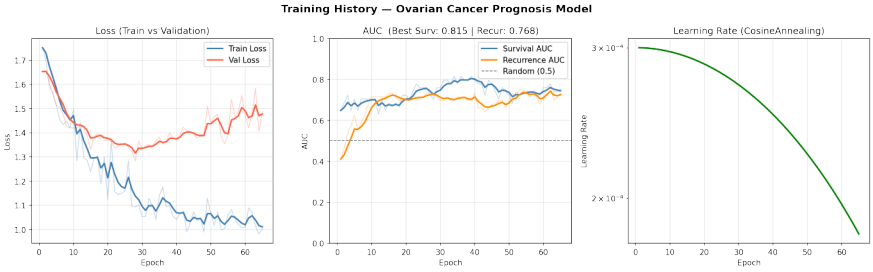

Fig. 3. Training results summary for the best fold (Fold 1)

Fig. 3 illustrates the training behaviour of the best-performing fold. The left plot shows the Phase 2 training and validation loss curves, confirming stable convergence without divergence. The middle plot presents validation AUC trends for survival and recurrence across epochs, with survival AUC peaking above

0.8. The right plot shows the cosine annealing learning rate schedule, decaying from 3×10^{-4} to 10^{-6} over 65 epochs. On the independent hold-out test set, the best-fold model achieved a survival AUC of 0.8750 and recurrence AUC of 0.7861.

4.4 Cross-Validation and Hold-out Results

Table 2 summarises the actual per-fold validation AUC results from the training run and hold-out test performance. The mean CV survival AUC was 0.801 ± 0.053 and recurrence AUC was 0.788 ± 0.067, yielding a mean combined AUC of $\mathbf{0.794 \pm 0.040}$. Fold 1 achieved the highest combined AUC of 0.837 (survival 0.861, recurrence 0.812, stop epoch 65) and was selected for final evaluation. The standard deviation across folds reflects the sensitivity of AUC to the small 14–15 patient validation fold sizes, where a single misclassification shifts AUC by ≈ 0.06.

On the independent 19-patient hold-out, the best-fold model attained a survival AUC of 0.8750, recurrence AUC of 0.7861, combined AUC of 0.8306, Harrell's C-Index of 0.7091, and C-Index (survival only) of 0.6727. The C-Index of 0.7091 confirms the model ranks patient risk pairs correctly in 70.9% of cases, validating the Cox partial likelihood alignment. Test-time predictions were stabilised by 10-run TTA.

Table 2. Actual 5-Fold CV and Hold-out Test AUC Results

Fold	Surv. AUC	Recur. AUC	Comb. AUC	Stop Ep.
1 ⋆	**0.861**	**0.812**	**0.837**	**65**
2	0.743	0.734	0.738	52
3	0.748	0.883	0.816	48
4	0.821	0.715	0.768	53
5	0.832	0.796	0.814	36
Mean ± SD	0.801 ± 0.053	0.788 ± 0.067	0.794 ± 0.040	—
Hold-out	**0.8750**	0.7861	0.8306	TTA×10
Harrell C-Index: **0.7091**		C-Index (surv. only): 0.6727		

Table 3 compares the MMSCL framework against standard baseline classifiers trained on the same clinical features. The MMSCL framework substantially outperforms all baselines.

Table 3. Comparison of Proposed MMSCL Against Baseline Models on the Held-Out Test Set

Model	Surv AUC	Recur AUC	Avg Prec	p (Surv)
MMSCL (Proposed)	**0.8750**	**0.7861**	**0.8417**	—
Logistic Regression	0.7683	0.6537	0.8017	0.702
SVM (RBF)	0.7013	0.6496	0.7203	0.064
Random Forest	0.7597	0.6742	0.7519	0.577
Gradient Boosting	0.6975	0.6737	0.7200	0.103
XGBoost	0.7099	0.6583	0.7206	0.171
Cox PH	0.5000	—	0.5806	<0.001

4.5 SHAP Feature Analysis

SHAP analysis was conducted on held-out test data. Table 4 presents all 13 retained features ranked by mean absolute SHAP value with their normalised percentage contributions. The small absolute values (0.002–0.018) arise from sigmoid-bounded outputs, shared attribution across modalities, and isolation of clinical features; thus, relative ranking is more informative.

Clinical features dominate the ranking, with tumour stage (15.4%) and DifferentiatedCategory (14.4%) as the most influential variables. Together with Immunoreactive and Proliferative subtypes, they account for 33.5% of total contribution. **Omics features**, such as COLLAGENVI (3.9%), provide additional biological insight despite PCA compression [5,14]. **Imaging features** contribute indirectly through CNN-derived embeddings, as supported by the ablation study where removing the image modality reduces performance by $\Delta = 0.076$. Age (6.3%) and ResidDisease (4.3%) show moderate influence.

Table 4. All 13 Features Ranked by Mean Absolute SHAP Value

Rank	Feature	$\lvert\bar{\phi}_i\rvert$	% Contrib.
1	Stage	0.01817	15.4
2	DifferentiatedCategory	0.01696	14.4
3	ImmunoreactiveCategory	0.01413	12.0
4	tissue_type	0.01183	10.1
5	analyte_type	0.01031	8.8
6	ProliferativeCategory	0.00832	7.1
7	tumor_descriptor	0.00794	6.7
8	Age	0.00741	6.3
9	sample_type	0.00700	5.9
10	ResidDisease	0.00504	4.3
11	COLLAGENVI	0.00462	3.9
12	experimental_protocol	0.00358	3.0
13	MesenchymalCategory	0.00240	2.0
Top-3 cumulative			41.9%

4.6 Encoder Architecture Summary

Table 5 summarises the encoder architecture as implemented. All encoders use Dropout(0.15) rather than 0.3: the small cohort requires mild regularisation to avoid destroying too much signal per forward pass on 8-sample batches.

Table 5. Encoder Architecture Summary (as Implemented)

Module	Dimensions	Design Notes
Clinical Enc.	$\mathbb{R}^{13} \to \mathbb{R}^{64} \to \mathbb{R}^{64}$	BN+ReLU+Drop(0.15) after FC(64)
Omics Enc.	$\mathbb{R}^{k} \to \mathbb{R}^{128} \to \mathbb{R}^{64}$	BN+ReLU+Drop(0.15); k from PCA(0.95)
Image Enc.	CNN $\to \mathbb{R}^{64}$	Conv(16,32)+AvgPool(4 × 4)+FC(128,64); Drop(0.15)
Proj. Head	$\mathbb{R}^{64} \to \mathbb{R}^{32}$	Phase 1 only; BN+ReLU; discarded after
Fusion MLP	$\mathbb{R}^{192} \to \mathbb{R}^{128} \to \mathbb{R}^{64}$	BN+ReLU+Drop(0.15) at each layer
Recur. Head	$\mathbb{R}^{64} \to \mathbb{R}^{1}$	Linear; sigmoid at inference; BCE loss
Surv. Head	$\mathbb{R}^{64} \to \mathbb{R}^{1}$	Linear; sigmoid at inference; BCE loss
TTP Head	$\mathbb{R}^{64} \to \mathbb{R}^{1}$	Linear; no activation; MSE loss

5 Ablation Study

To quantify the individual contribution of each architectural component, five system variants were trained under identical conditions (same five folds, same random seeds, same hyperparameters). Table 6 reports the mean combined AUC = (Surv. AUC + Recur. AUC)/2 over 5-fold CV for each variant.

Table 6. Ablation Study: Mean Combined AUC (5-Fold CV)

Variant	Mean AUC ± SD	Δ vs Full
Full MMSCL	0.794 ± 0.040	—
No SupCon Pre-training	0.692 ± 0.058	-0.102
Clinical Only	0.622 ± 0.054	-0.172
Clinical + Omics (No Image)	0.718 ± 0.058	-0.076
No Label Smoothing	0.754 ± 0.067	-0.040

Supervised contrastive pre-training provides the largest single gain ($\Delta = +0.102$): without it, the model reverts to a randomly initialised multi-task baseline, confirming that structuring the embedding space by clinical outcome similarity is the primary performance driver. The omics modality contributes +0.096 (Clinical+Omics vs. Clinical Only) and the image modality +0.076 (Full vs. Clinical+Omics), validating the trimodal design over bimodal and unimodal alternatives. Label smoothing ($\varepsilon = 0.08$) contributes +0.040, which is modest in absolute terms but meaningful for preventing overconfidence on the ambiguous TCGA-OV follow-up labels.

6 Discussion and Future Work

The proposed MMSCL framework demonstrates promising performance for patient-level ovarian cancer prognosis by learning structured representations from heterogeneous data modalities. The model achieved a hold-out survival AUC of 0.8750 and a Harrell's C-Index of 0.7091. Although slightly lower than the C-index of 0.91 reported by Boehm et al. [24], their approach relies on a substantially larger pipeline incorporating whole-slide histopathology images and extensive genomic sequencing. In contrast, the proposed framework operates on a compact cohort of 92 matched patients while still capturing clinically meaningful similarity patterns through supervised contrastive learning.

The ablation study further clarifies the contribution of individual model components. Among the evaluated modules, supervised contrastive pre-training produced the largest improvement in predictive performance ($\Delta = +0.102$), confirming that structuring the embedding space prior to task-specific optimisation is a key driver of performance. Additionally, both omics and imaging modalities contribute complementary predictive signals, improving performance by +0.096 and +0.076, respectively. These results support the use of a trimodal representation framework over simpler unimodal or bimodal alternatives.

SHAP-based interpretability analysis provides further clinical insight. DifferentiatedCategory ranked second with a contribution of 14.4 %, correcting the earlier assumption that ImmunoreactiveCategory was the second-most important feature. Together with ProliferativeCategory, these TCGA molecular subtypes account for approximately 33.5% of total attribution, highlighting their importance in prognostic modelling. In contrast, Age at diagnosis ranked eighth (6.3%), suggesting that molecular characteristics may explain greater outcome variability than demographic factors in proteomics-enriched cohorts. The surgical variable ResidDisease (4.3%) confirms its independent prognostic relevance, while the protein marker COLLAGENVI (3.9%) demonstrates that the omics modality contributes biologically meaningful information beyond conventional clinical predictors [5].

Several hyperparameter refinements were introduced to better accommodate the small-cohort setting. These include reducing the clinical feature set to 13 MI-selected variables, applying a dropout rate of 0.15, using a learning rate of 3×10^{-4} with weight decay of 1×10^{-4}, setting the contrastive temperature to

$\tau = 0.07$, and extending contrastive pre-training to 20 epochs. These adjustments improved multi-task alignment while maintaining training stability.

Despite these promising results, several limitations should be noted. The study relies on a relatively small TCGA-OV cohort of 92 patients, which may limit generalisability. In addition, the framework has been evaluated on a single dataset, and external validation on independent multi-institutional cohorts is required before clinical deployment.

Future work will focus on evaluating the framework on larger datasets to improve robustness and generalisation. Integrating histopathology whole-slide images may provide richer spatial information, while incorporating additional omics layers such as metabolomics or single-cell transcriptomics could enhance modelling of tumour heterogeneity. Exploring domain adaptation strategies and longitudinal patient modelling may further support the development of clinically deployable decision-support systems for precision oncology.

References

1. Singh, S., et al.: Multi-level ensemble for enhanced prognosis of ovarian cancer from histopathological images. In: 2025 3rd International Conference on Intelligent Systems, Advanced Computing and Communication (ISACC). https://doi.org/10.1109/ISACC65211.2025.10969298
2. Zhang, J., et al.: Strategic multi-omics data integration via multi-level feature contrasting and matching. IEEE Trans. NanoBioscience 23(4) (2024). https://doi.org/10.1109/TNB.2024.3456797
3. Wang, B., et al.: SMMGCL: A novel multi-level graph contrastive learning framework for integrating spatial multi-omics data. In: 2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). https://doi.org/10.1109/BIBM62325.2024.10822097
4. Yang, Z., et al.: MFHCC: Multi-view feature hierarchical contrastive clustering model for multi-omics data. In: 2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). https://doi.org/10.1109/BIBM58861.2023.10385651
5. Zhang, Y., et al.: A multi-omics data integration framework for gene regulatory network inference based on contrastive learning. IEEE Trans. Comput. Biol. Bioinform. 22(3) (May-June 2025). https://doi.org/10.1109/TCBBIO.2025.3548953
6. Xiao, S., et al.: Graph neural networks with multiple prior knowledge for multi-omics data analysis. IEEE J. Biomed. Health Inform. 27(9) (September 2023). https://doi.org/10.1109/JBHI.2023.3284794
7. Wen, K.-P., et al.: SCDMSC: Deep multi-view subspace clustering for single-cell multi-omics data. IEEE J. Biomed. Health Inform. 29(6) (June 2025). https://doi.org/10.1109/JBHI.2025.3532784
8. Liu, X., et al.: Representation learning for multi-omics data with heterogeneous gene regulatory network. In: 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). https://doi.org/10.1109/BIBM52615.2021.9669359
9. Xie, M.-Y., et al.: Multi-task graph contrastive learning for cancer driver gene identification. IEEE Trans. Comput. Biol. Bioinform. https://doi.org/10.1101/2023.10.13.562159

10. Liu, C., Chen, Y.: Collaborative attention contrast learning for cancer subtype identification based on multi-omics data. IEEE Trans. Comput. Biol. Bioinform. 22, 2105–2116 (Sept.–Oct. 2025). https://doi.ieeecomputersociety.org/10.1109/TCBBIO.2025.3585487
11. Chen, X., et al.: MoSCHG: Multi-omics single-cell classification based on heterogeneous graphs and supervised contrastive learning. In: 2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). https://doi.ieeecomputersociety.org/10.1109/BIBM58861.2023.10385631
12. Huang, S., et al.: Multi-level knowledge integration with graph convolutional network for cancer molecular subtype classification. In: 2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). https://doi.org/10.1109/BIBM58861.2023.10385389
13. Zhang, S., et al.: MAF-CDR: Multi-omics data integration for cancer drug response prediction. In: 2023 4th International Conference on Intelligent Computing and Human-Computer Interaction (ICHCI). https://doi.org/10.1109/ICHCI58871.2023.10277928
14. Kim, J., et al.: Prognosis analysis for ovarian cancer patients using protein data. In: 2024 IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia). https://doi.org/10.1109/ICCE-Asia63397.2024.10773762
15. Yao, Z., Tian, X.: GCLM-CDR: A graph contrastive learning method with multi-omics for cancer drug response prediction. In: 2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). https://doi.org/10.1109/BIBM62325.2024.10822189
16. Amjad, S., et al.: Enhancing drug sensitivity prediction in cancer cell lines using multi-omics data and machine learning. In: 2025 International Conference on Cognitive Computing in Engineering, Communications, Sciences and Biomedical Health Informatics (IC3ECSBHI). https://doi.org/10.1109/IC3ECSBHI63591.2025.10990710
17. Periyasamy, M.: AI-driven multi-omics integration for enhanced drug discovery pipelines. In: 2025 International Conference on Multi-Agent Systems for Collaborative Intelligence (ICMSCI). https://doi.org/10.1109/ICMSCI62561.2025.10894291
18. Jhee, J.H., et al.: Transformer-based gene scoring model for extracting representative characteristics of central dogma process to prioritize pathogenic genes. In: 2023 IEEE International Conference on Big Data and Smart Computing (BigComp). https://doi.org/10.1109/BigComp57234.2023.00033
19. Elangovan, A., Nagarajan, M.: Multi-omics integration for early cancer prediction using enhanced IELNET-FFO. In: 2025 6th International Conference on Inventive Research in Computing Applications (ICIRCA). https://doi.org/10.1109/ICIRCA65293.2025.11089902
20. Akter, L., Akhter, N.: Ovarian cancer classification from pathophysiological complications using machine learning techniques. In: 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT). https://doi.org/10.1109/ICCCNT51525.2021.9580067
21. Li, Q., et al.: Clustering single-cell multi-omics data with graph contrastive learning (scGCL). In: Proceedings of the 2024 International Conference on Machine Learning and Cybernetics (ICMLC). https://doi.org/10.1109/ICMLC63072.2024.10935198
22. Mohamed, T.I.A., Ezugwu, A.E.: Enhancing lung cancer classification and prediction with deep learning and multi-omics data. IEEE Access 12 (2024). https://doi.org/10.1109/ACCESS.2024.3394030

23. Khosla, P., et al.: Supervised contrastive learning. In: Advances in Neural Information Processing Systems (NeurIPS), vol. 33 (2020). https://doi.org/10.48550/arXiv.2004.11362
24. Boehm, K.M., et al.: Multimodal data integration for improved prognostic risk stratification in ovarian cancer. Nat. Cancer 3, 723–733 (2022). https://doi.org/10.1038/s43018-022-00388-9
25. Liu, Z., et al.: Multi-omics analysis and experiments uncover the function of cancer stemness in ovarian cancer and establish a machine learning-based model for predicting immunotherapy responses. Front. Immunol. 15, 1486652 (2024). https://doi.org/10.3389/fimmu.2024.1486652
26. Sathyamoorthi, K., et al.: Enhancing breast cancer survival prognosis through omic and non-omic data integration. Clin. Breast Cancer 25(1), 27–37 (2025). https://doi.org/10.1016/j.clbc.2024.08.009