

Multi-Sensor View Fusion via Cross-Domain Recommendation for Cervical Cancer Risk Prediction

Jiaxin Zhang

Qilu Hospital of Shandong University

Yuanyuan Fang

Boston University

Anna Dai

Shandong Second Medical University

Yumin Wang

Qilu Hospital of Shandong University

Jingrui Zhang

Shandong University of Finance and Economics

Mengyuan Sun

Qilu Hospital of Shandong University

Yongchun Deng

RIZHAO STEEL HOLDING GROUP CO., LTD.

Bingchen Liu

Shandong University

Yuchen Cui

Qilu Hospital of Shandong University

Youzhong Zhang

Qilu Hospital of Shandong University

Qian Chen

Qilu Hospital of Shandong University

Sai Han

qiluyiguanhansai@163.com

Qilu Hospital of Shandong University

Research Article

Keywords: Cervical cancer, Cross-domain recommendation, Machine learning, Multi-view learning, Risk prediction

Posted Date: September 22nd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10974727/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Multi-Sensor View Fusion via Cross-Domain Recommendation for Cervical Cancer Risk Prediction

Jiaxin Zhang^{1,2}, Yuanyuan Fang³, Anna Dai⁴, Yumin Wang¹,
Jingrui Zhang⁵, Mengyuan Sun^{1,2}, Yongchun Deng⁷,
Bingchen Liu⁶, Yuchen Cui^{1,2}, Youzhong Zhang^{1,2}, Qian Chen^{1,2},
Sai Han^{1,2*}

¹Department of Obstetrics and Gynecology, Qilu Hospital of Shandong
University, Jinan, 250012, China.

²Shandong Key Laboratory of Reproductive Health and Birth Defects
Prevention and Control, Jinan, China.

³Department of Computer Science, Metropolitan College, Boston
University, Boston, MA, 02215, USA.

⁴Shandong Second Medical University, China.

⁵School of Statistics and Mathematics, Shandong University of Finance
and Economics, Jinan, Shandong, China.

⁶School of Software, Shandong University, 1500 Shunhua Road, Jinan,
250000, Shandong Province, China.

⁷RIZHAO STEEL HOLDING GROUP CO., LTD..

*Corresponding author(s). E-mail(s): qiluyiyuanhansai@163.com;

Contributing authors: 779726737@qq.com; fyy@bu.edu;

1527284777@qq.com; 2632148096@qq.com; 242113050@stu.sdufe.edu.cn;

mengyuansun.hope@163.com; dengyongchun@rizhaosteel.com;

lbcraf2018@126.com; Miggecui@163.com; zhangyouzhong@sdu.edu.cn;

18769781275@163.com;

Abstract

Cervical cancer remains a major global health burden, and accurate risk prediction is important for early clinical assessment. Conventional machine-learning models may be limited by sparse records, heterogeneous clinical variables, and

high-dimensional inputs. This study proposes a cervical cancer cross-domain prediction system (CPS) that integrates multi-view feature extraction, auxiliary-data augmentation, and cross-domain correlation learning. The framework combines clinical characteristics with human papillomavirus testing, cytology, colposcopy, and pathological outcomes while learning view-specific contributions to the final prediction. Experiments on an anonymized single-center retrospective cohort compare CPS with logistic regression, support vector machines, k-nearest neighbors, random forests, and gradient-boosted decision trees. CPS achieved the best overall performance among the evaluated models, with a mean squared error of 0.051, an area under the receiver operating characteristic curve of 0.831, an accuracy of 0.802, and a sensitivity of 0.769. Ablation and parameter analyses further supported the contribution of the proposed components. These findings indicate that cross-domain recommendation may provide a useful framework for integrating heterogeneous clinical information. External multicenter validation is required before clinical deployment.

Keywords: Cervical cancer, Cross-domain recommendation, Machine learning, Multi-view learning, Risk prediction

1 Introduction

Cervical cancer (CC) is the fourth most common malignant tumor among women worldwide, with 604,000 new cases and 342,000 deaths reported globally in 2020[19]. With the rise of disease prediction models, machine learning algorithms show great potential in predicting disease occurrence, progression, and complications in recent years[20][21]. Cross-domain recommendation is a recommendation system technology that leverages information and knowledge from different domains or data spaces to provide more accurate and comprehensive prediction results for users in the target domain. This technology already sees mature applications in personalized shopping recommendations and personalized recommendations for healthcare wearable devices[8][15][24].

Although traditional machine learning performs well in many tasks, it also has some notable limitations. Firstly, it heavily relies on feature design, as the quality of features directly impacts model performance. Feature determination and design require high-level domain knowledge and experience, which is time-consuming and prone to errors. Secondly, traditional machine learning models struggle with high-dimensional data, leading to increased computational complexity and reduced model performance. Additionally, traditional machine learning faces data sparsity issues. In cases of sparse data, traditional models find it difficult to learn effective patterns, resulting in overfitting and poor generalization.

In traditional machine learning, to address these issues, we often need to involve senior doctors or professors in the project design, which is constrained by their availability. Moreover, to improve model performance, we spend significant time reducing

high-dimensional data to lower dimensions to aid model learning. To tackle data sparsity and missing data, we resort to imputation methods, which, while partially solving the problem, introduce some degree of data distortion.

Statistical models are commonly used to analyze clinical outcomes and predict cervical cancer outcomes. Traditional machine learning methods, such as Random Forest, SVM, Decision Trees, Logistic Regression, and KNN, build prediction models by identifying relevant features and calculating their correlation with outcomes, followed by scoring and comprehensive feature computation. These methods have been applied in various cancer-related fields, including cancer-causing gene prediction, disease risk prediction, and survival rate prediction at different stages of cancer treatment[20][26][14][9].

However, these methods have their limitations. Decision Trees classify or regress by recursively partitioning data into a tree structure, which is easy to interpret and can handle non-linear data but is prone to overfitting and sensitive to data perturbations. SVMs suffer from high computational complexity and inefficiency with large-scale data, making them less suitable for multi-center, large-volume datasets. Random Forest, while resistant to overfitting and suitable for high-dimensional data, has high model complexity and long training times, which severely limit its use.

In this paper, we propose CPS, a cross-domain recommendation framework for cervical cancer risk prediction. CPS integrates K-means clustering, multi-view learning, supervised learning, and unsupervised learning to exploit complementary information across heterogeneous clinical views.

Our main contributions are summarized as follows:

- We formulate cervical cancer risk prediction as a cross-domain recommendation problem that integrates heterogeneous clinical views.
- We develop view-specific training and auxiliary-data augmentation strategies for patients with sparse pathological records.
- We evaluate CPS on a real-world retrospective dataset and compare it with commonly used machine-learning baselines.

2 Related Work

The following two areas of research are relevant to our work: multi-view clustering and machine learning. Multi-view clustering is a method that uses data from different sources or feature representations for cluster analysis. Its goal is to uncover the intrinsic structure of data by integrating information from multiple views, thereby achieving more accurate clustering results. It can be divided into two main categories: co-training-based methods and graph-based methods. Co-training-based methods improve clustering effectiveness by co-training clustering models across multiple views, leveraging the complementarity and consistency between views.

2.1 Multi-View Clustering

Multi-view clustering offers several advantages, including strong information complementarity, robustness, suitability for complex data structures, support for multi-modal

information fusion, and improved interpretability of results. In our study, we chose to use multi-view clustering to build our prediction model due to its strong multi-modal information fusion capabilities[3].

2.2 Machine Learning

Traditional machine learning is divided into two main categories: supervised learning and unsupervised learning. Supervised learning trains models using labeled data. The goal is to learn the mapping relationship between input features and output labels to predict new input data. It has been widely applied in various medical fields. Common techniques include Logistic Regression, Decision Trees, SVM, and Random Forest[18][25][7][10]. Unsupervised learning identifies patterns or structures from unlabeled data. It also has many applications in medical and biological research, with common techniques including K-Means Clustering, Hierarchical Clustering, and Principal Component Analysis[11][1][5][6].

In cervical cancer, we primarily use supervised learning to build prediction models[6][23][17], particularly in predicting the likelihood of progression from SIL to cervical cancer and in prognosis, where some preliminary explorations have been made[13][22]. Some studies have also explored the prospects of machine learning-based cervical cancer prediction models[12]. However, previous research has not addressed the inherent limitations of traditional machine learning, such as insufficient information sources and the inability to handle multi-domain information. Therefore, developing a more accurate and targeted cervical cancer prediction model is increasingly necessary.

3 Problem Formulation

We use anonymized medical records from 1003 patients, including HPV testing, TCT, colposcopy, and general clinical characteristics. The study protocol was approved by the Ethics Committee of Qilu Hospital of Shandong University, and the requirement for informed consent was waived for this retrospective analysis of anonymized records. All procedures were conducted in accordance with the Declaration of Helsinki.

We denote the source and target domains by D_s and D_t , respectively. Nine clinical factors associated with cervical cancer risk constitute the source domain, whereas pathological diagnosis constitutes the target domain. Let U_o denote overlapping patients with observations in both domains. The cold-start set U_{cs} contains patients without prior pathological records in D_t ; their available information in D_s is used to predict the corresponding pathological outcome. We divide U_{cs} into a validation subset U_{cs1} and a test subset U_{cs2} . Patients in U_o are used to train and validate the model, whereas U_{cs1} and U_{cs2} are used to evaluate prediction under cold-start conditions.

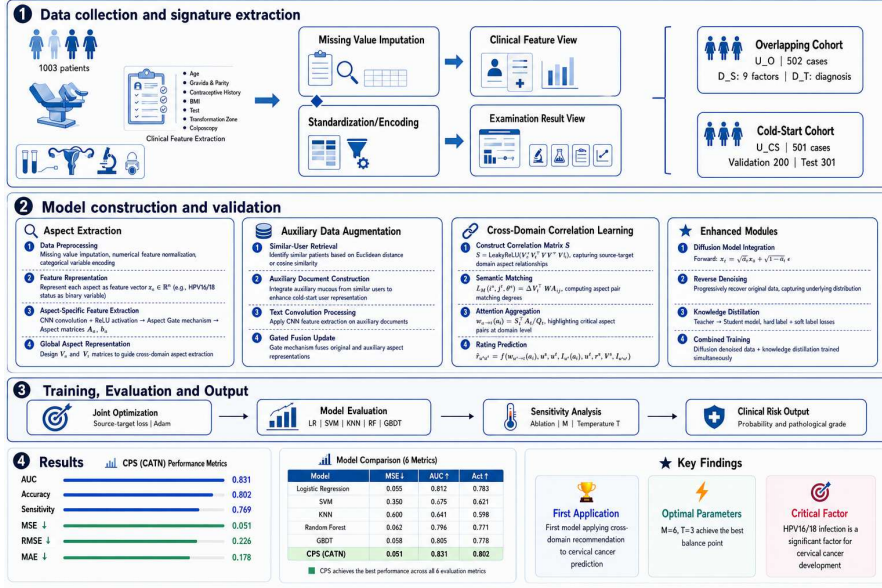


Fig. 1 Overview of the proposed CPS framework.

4 The CPS Framework

4.1 Overview of CPS

The overall CPS architecture is illustrated in Fig. 1. It comprises aspect extraction, auxiliary-data augmentation, and cross-domain aspect-correlation learning. Aspect extraction represents clinical characteristics and examination results from multiple views. Auxiliary-data augmentation improves representations for patients with sparse records by incorporating information from clinically similar patients. Cross-domain aspect-correlation learning identifies the source-domain aspects that are most informative for pathological outcomes in the target domain.

4.2 Aspect Extraction

Aspect extraction is a core component of the CPS framework, with the aim of extracting multiple aspects from source-domain data for prediction in the target domain. This process is achieved through four main steps: data preprocessing, feature representation, aspect-specific feature extraction, and aspect aggregation.

4.2.1 Data Preprocessing

In this step, the raw data in the source domain is preprocessed to ensure data consistency and quality. This includes handling missing data, normalizing numerical features, and encoding categorical variables. Suppose we have a source-domain dataset D_s , and after preprocessing, we get a refined dataset D'_s .

4.2.2 Feature Representation

Each aspect (such as HPV status, TCT results) is represented as a feature vector. For example, if we consider an aspect a with n possible states, we can represent it as a vector $e_a \in R^n$. For instance, whether there is an HPV16/18 infection can be represented as a binary variable. Given a user - related document D_u in the source domain, where $D_u = [w_1, w_2, \dots, w_l]$, we first project each word w_j to its embedding representation e_j . So, $E_u = [e_1, e_2, \dots, e_l]$, where $e_j \in R^d$, l is the document length, and d is the word - embedding dimension.

4.2.3 Aspect - Specific Feature Extraction and Aggregation

In this step, we first focus on extracting relevant features for each aspect. Contextual features $c_{j,u}$ for the j - th word are obtained through a convolution operation on E_u with specific convolution filters and a ReLU activation function, forming a feature matrix C_u . Then, an aspect-specific gate mechanism is used to identify which features are relevant to each aspect. For the m - th aspect, aspect-specific features $g_{m,j,u}$ of word w_j are calculated using certain matrices and bias vectors, and element-wise product operation with the sigmoid activation function. As a result, we get M aspect-specific words' contextual features G_u for further aspect extraction.

Next, we consider that reviews from different domains emphasize different aspects. We design two matrices of global - sharing aspect representations for the source domain D_s and target domain D_t to guide the aspect extraction. The representation $a_{m,u}$ of the m - th aspect is derived by aggregating the relevant features in $G_{m,u}$ with weights $\beta_{m,j,u}$ which indicate the importance of each word towards the m - th aspect. In this way, we obtain the aspect matrix A_u for D_u and A_i for the item - related document D_i . It should be noted that the parameters for aspect extraction for D_u and D_i are shared in each learning flow, while a distinct set of parameters is used in each learning flow. Also, the matrices V_s and V_t are shared in their corresponding domains respectively. By extracting multiple aspects from the source-domain data, CPS can comprehensively capture the multi - dimensional characteristics of cervical cancer risk factors and make more accurate and interpretable predictions, with the combined aspects forming the input for the cross-domain aspect correlation learning component.

4.2.4 Addressing Data Sparsity with Auxiliary Reviews in CPS

In the CPS framework, data sparsity poses a major challenge, particularly for cold-start users with limited data in the source domain. To tackle this, CPS utilizes auxiliary reviews. First, for a given user - item pair, an auxiliary review is defined as one from another user who gave the same rating for that item. The auxiliary document for a user is created by combining such reviews from the historical items the user purchased within the same domain, focusing only on non - overlapping users to boost data diversity.

The process of enhancing with auxiliary reviews involves several steps. Initially, CPS identifies similar users for each cold-start user based on review - related profiles, using metrics like Euclidean distance or cosine similarity. Then, the auxiliary reviews

of these similar users are integrated into the cold-start user’s context. This integration includes text convolution on the auxiliary document to extract a feature matrix, followed by applying the same Aspect Gate Control and Aspect Attention processes to obtain an aspect matrix from the auxiliary document. Finally, a gate mechanism is employed to effectively update the original aspect representation with the aspect representation from the auxiliary document.

Overall, incorporating auxiliary reviews helps CPS alleviate data sparsity and improve the representation quality of cold-start users. The enhanced user profile, combining original and auxiliary review information, serves as input for the cross-domain aspect correlation learning component, thereby significantly enhancing the model’s predictive performance even with incomplete review data.

4.3 Cross - Domain Aspect Correlation Learning

The cross-domain aspect correlation learning component plays a crucial role in the CPS framework. Its main task is to learn the correlations between aspects in the source and target domains, which is essential for achieving accurate predictions, especially for cold-start users.

4.3.1 Constructing the Correlation Matrix

CPS constructs a correlation matrix to capture the relationships between aspects in the source and target domains. Recall that we have a set of global aspect representations V_s and V_t which were previously used to guide the aspect extraction process. Here, we utilize these static aspect representations to calculate the global cross-domain aspect correlation matrix S as follows:

$$S = \text{LeakyReLU}(V_s^\top W V_t) \quad (1)$$

where $S(p, q)$ reflects the importance of preference transfer based on aspect p from the source domain and aspect q from the target domain. $S \in \mathbb{R}^{M \times M}$, and $W \in \mathbb{R}^{k \times k}$ is a learnable matrix for affinity projection. The LeakyReLU activation function is adopted, and by setting the corresponding α to be a very small value (e.g., 0.01), it supports the sparse aspect correlation across domains.

4.3.2 Applying the Attention Mechanism

To focus on the most relevant aspects, CPS employs an attention mechanism. First, we calculate the semantic matching between each aspect pair between the abstract aspect features A_u (for user u) and A_i (for item i) as follows:

$$S_{u,i} = A_u^\top W A_i \quad (2)$$

Similar to the calculation of S , $S_{u,i}(p, q)$ reflects the matching degree between the corresponding aspects, and the matrix W is shared for affinity projection.

We then use the correlation matrix S as the attention weights to aggregate the pairwise aspect matchings. This is done to highlight the important aspect pairs at the domain level for better prediction. The aggregated result $S_{u,i}^r$ is calculated as:

$$S_{u,i}^r = S \odot S_{u,i} \quad (3)$$

4.3.3 Making Predictions and Evaluation

Based on the weighted aspects, CPS makes predictions for cold-start users in the target domain. The final rating prediction $\hat{r}_{u,i}$ is calculated as:

$$\hat{r}_{u,i} = \frac{1}{M^2} \sum_{p=1}^M \sum_{q=1}^M S_{u,i}^r(p, q) + b_u + b_i \quad (4)$$

Here, b_u and b_i are the user bias and item bias respectively.

After making the predictions, we compare the predicted results with the actual pathological diagnostic outcomes. This comparison helps us evaluate the performance of the model. For example, in the context of cervical cancer prediction, since HPV16/18 is a significant factor in the development of cervical cancer, the attention mechanism assigns it a higher weight. By learning the correlations between aspects in the source and target domains, CPS can provide more accurate and interpretable predictions for cold-start users.

4.4 Optimization Strategy Experimental Process

In the model training phase, we adopt an optimization strategy that leverages the interactions of overlapping users in the source domain D_s and target domain D_t to optimize model parameters.

Let O_s and O_t be batches of observed user - item rating pairs in D_s and D_t respectively, limited to the overlapping user set U_o . The loss functions L_s and L_t are defined as:

$$L^* = \frac{1}{|O^*|} \sum_{(u,i) \in O^*} (r_{u,i} - \hat{r}_{u,i})^2 + \lambda \|\Theta^*\| \quad (* = s \text{ or } t) \quad (5)$$

where λ is the regularization coefficient to prevent overfitting, Θ^* are the trainable parameters, $r_{u,i}$ is the actual rating of user u for item i , $\hat{r}_{u,i}$ is the predicted rating, and $|O^*|$ is the size of the rating pair set O^* .

We execute two alternating learning flows: predicting $r_{u,i}$ in the target domain using D_u from the source domain and D_i from the target domain, and predicting $r_{u,i}$ in the source domain using D_u from the target domain and D_i from the source domain. Each training batch consists of shuffled O_s and O_t in a fixed proportion $\frac{|O_s|}{|O_t|} = \frac{|R_s|}{|R_t|}$, where $|R_s|$ and $|R_t|$ are the number of ratings given by U_o in D_s and D_t respectively.

We use the Adam optimizer to update the model parameters.

Regarding the context view, we first obtain the context representations. After that, we calculate the aspect attention matrix S_{attn} based on the global aspect representations v_u and v_i :

$$S_{attn} = \text{LeakyReLU}(\text{Concat}(\mathbf{a}_u)^\top \mathbf{W}_c \text{Concat}(\mathbf{a}_i)) \quad (6)$$

where $W_{convert}$ is a conversion matrix.

Next, we compute the context attention weights and generate context representations:

$$(\boldsymbol{\alpha}_x^{\text{ctx}}, \mathbf{c}_x) = \text{ContextAttn}(\mathbf{R}_x^{\text{ctx}}, \mathbf{v}_x), \quad x \in \{u, i\} \quad (7)$$

To avoid overfitting, we add a regularization term for context attention to the loss function:

$$\boldsymbol{\theta} \leftarrow \text{Adam}\left(\boldsymbol{\theta}, \nabla_{\boldsymbol{\theta}} \left[L_{\text{ratings}} + \lambda \left(\|\boldsymbol{\alpha}_u^{\text{ctx}}\|_2^2 + \|\boldsymbol{\alpha}_i^{\text{ctx}}\|_2^2 \right) \right] \right) \quad (8)$$

where L_{ratings} is the rating prediction loss, α is the regularization rate, and $\boldsymbol{\theta}$ represents the model parameters.

Through this comprehensive optimization strategy and experimental process, we iteratively adjust the model parameters during training. In subsequent evaluations, we will compare the predicted results with actual ratings to verify the model's performance and the effectiveness of the optimization strategy.

4.5 Improved Framework in Experimental Section

In this subsection, we introduce the enhanced experimental framework that integrates a diffusion model and a knowledge distillation model into the training process. The proposed improvements aim to boost the performance and generalization ability of the model.

4.5.1 Diffusion Model Integration

The diffusion model is incorporated to gradually denoise the input data, which helps in capturing the underlying data distribution more effectively. The diffusion process consists of a forward diffusion step and a reverse denoising step. In the forward diffusion step, Gaussian noise is iteratively added to the original data x_0 according to the predefined schedule of β values. The formula for generating the noisy data x_t at time step t is given by:

$$x_t = \sqrt{\alpha_t}x_0 + \sqrt{1 - \alpha_t}\epsilon, \quad \epsilon \sim \mathcal{N}(0, I) \quad (9)$$

where α_t is calculated based on the β values, and ϵ represents the random noise.

In the reverse denoising process, the model aims to recover the original data from the noisy data x_t step by step. The sampling formula for obtaining x_{t-1} from x_t is:

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(x_t - \frac{\beta_t}{\sqrt{1 - \alpha_t}} \epsilon_{\theta}(x_t, t) \right) + \sigma_t z \quad (10)$$

where $\epsilon_{\theta}(x_t, t)$ is the noise predicted by the model, σ_t is the variance at time step t , and $z \sim \mathcal{N}(0, I)$. The diffusion loss in (11) measures the difference between the predicted and actual noise.

$$\mathcal{L}_{\text{diffusion}} = \mathbb{E}_{t, x_0, \epsilon} \left[\|\epsilon - \epsilon_{\theta}(x_t, t)\|_2^2 \right] \quad (11)$$

4.5.2 Knowledge Distillation Model

Knowledge distillation is employed to transfer the knowledge from a pre-trained teacher model to a student model. The teacher model, which is usually more complex and has better performance, serves as a knowledge source. The student model, with a simpler architecture, aims to mimic the behavior of the teacher model.

The distillation process uses hard-label and soft-label losses. The hard-label loss $\mathcal{L}_{\text{hard}}$ is the mean squared error between the student prediction $\hat{y}_s$ and the ground-truth label y , as shown in (12).

$$\mathcal{L}_{\text{hard}} = \frac{1}{N} \sum_{i=1}^N \|\hat{y}_{s,i} - y_i\|_2^2 \quad (12)$$

The soft-label loss $\mathcal{L}_{\text{soft}}$ is computed using the Kullback–Leibler divergence between the temperature-scaled outputs of the student and teacher models, as shown in (13).

$$\mathcal{L}_{\text{soft}} = \frac{1}{N} \sum_{i=1}^N \text{KLD} \left(\text{softmax} \left(\frac{\hat{y}_{s,i}}{T} \right), \text{softmax} \left(\frac{\hat{y}_{t,i}}{T} \right) \right) \quad (13)$$

The total distillation loss is a weighted combination of the hard-label and soft-label losses, as shown in (14).

$$\mathcal{L}_{\text{distillation}} = (1 - \alpha) \mathcal{L}_{\text{hard}} + \alpha T^2 \mathcal{L}_{\text{soft}} \quad (14)$$

where α is the weight parameter that balances the contributions of the two losses.

4.5.3 Combined Training Process

During combined training, the input data are first processed by the diffusion model. The resulting denoised features are then provided to the teacher and student models for knowledge distillation. The diffusion and distillation losses jointly update the student model, allowing it to learn from both the data distribution and the teacher predictions.

5 Experimental Study

5.1 Dataset Description and Preprocessing

We utilize anonymized clinical data from patients who underwent Cold Knife Conization (CKC) procedures at Qilu Hospital of Shandong University between 2016 and 2023. All personally identifiable information is removed to ensure no privacy breach risks. We systematically identify and independently categorize all clinically recognized factors influencing cervical cancer development into 10 distinct feature sets (triplets). These encompass both diagnostic test results (including HPV status, TCT findings, and colposcopy reports) and general patient characteristics (such as age, gravidity, and parity). Each feature is associated with pathological outcomes through auxiliary annotations. The dataset is randomly partitioned into common users and cold-start

users. The group of Common users has 502 cases with complete scoring data across training, validation, and test sets. The group of Cold-start users has 501 cases further divided into validation cold-start users (Ucs1): 200 cases (40% of cold-start cohort) And the group of Test cold-start users (Ucs2) has 301 cases (60% of cold-start cohort) This stratification enable comprehensive evaluation of model performance across different user scenarios while maintaining rigorous data privacy standards. All data processing complies with ethical guidelines for retrospective studies using anonymized medical records.

5.2 Selection of Predictive Factors

Based on prior studies and established clinical guidance, we select age, gravidity, parity, contraceptive method, HPV test results, TCT results, transformation-zone type, and colposcopy findings as predictive factors. HPV16 and HPV18 are represented separately because of their established association with cervical carcinogenesis [2]. High-grade squamous intraepithelial lesion is treated as a clinically important pre-cancerous endpoint in accordance with risk-based management recommendations [16]. Colposcopy-related factors are defined according to relevant Chinese and international consensus recommendations [4, 16].

5.3 Comparison With Commonly Used Clinical Prediction Models

We further test our dataset using the mainstream machine learning predictive models in the current medical field to determine whether the accuracy of our model’s prediction has an advantage. We select Logistic Regression, Support Vector Machine (SVM), and K-Nearest Neighbors (KNN) for prediction. Logistic regression has the advantages of a simple model and easy interpretability. It directly provides the influence degree (i.e., coefficients) of each feature on the prediction result, which helps to understand which factors have a greater impact on the occurrence of cervical cancer. In addition, logistic regression performs stably in handling binary classification problems and has high computational efficiency. SVM has strong generalization ability and achieve good results when dealing with small sample data. At the same time, by choosing different kernel functions, SVM can handle various complex classification problems. In the prediction of cervical cancer, the data has certain complexity and nonlinear relationships, and SVM is expected to capture these relationships through appropriate kernel functions to improve the prediction accuracy. The KNN algorithm is simple and intuitive and easy to implement. It does not need to make assumptions about the distribution of data and is applicable to various types of data. In the prediction of cervical cancer, KNN classifies by considering the similarity between samples, and it performs well for some data with local characteristics.

As shown in Table 1, CPS achieved the best performance across all six evaluation metrics.

Table 1 Comparison of the predictive performance of different models in cervical cancer prediction.

Model	MSE↓	RMSE↓	MAE↓	AUC↑	Accuracy↑	Sensitivity↑
Logistic Regression	<u>0.055</u>	<u>0.234</u>	<u>0.189</u>	<u>0.812</u>	<u>0.783</u>	<u>0.745</u>
SVM	0.550	0.742	0.601	0.675	0.621	0.583
KNN	0.600	0.775	0.634	0.643	0.598	0.552
Random Forest	0.062	0.249	0.201	0.798	0.771	0.732
GBDT	0.058	0.241	0.195	0.805	0.779	0.738
CPS	0.051	0.226	0.178	0.831	0.802	0.769

Note: ↓ indicates that a lower value is better, whereas ↑ indicates that a higher value is better. Bold values denote the best results, and underlined values denote the second-best results.

5.4 Split Testing

Split Testing is an experimental method that determines the optimal solution by comparing the effects of different versions. The target population is randomly divided into several groups, and different strategies are applied to each group, and the best solution is selected through statistical comparison of the results. In order to test the effectiveness of each part of our framework, we conduct split testing on the above three models. The data of the experimental group and the validation group follow our previous grouping situation to ensure that there are no significant differences in key features between the two groups ($p > 0.05$). All features are standardized (Z-score), and categorical variables are encoded using one-hot encoding. Taking the presence or absence of HPV16/18 infection as an example, after we remove this factor, all three models show a large bias, indicating the necessity of this factor for our model. We conduct split testing on all factors for verification.

Table 2 Ablation experiment

experimental setup	MSE
Remove all contextual views	0.053
Remove the contextual view of the item	0.052
Remove the user’s contextual view	0.054
CPS	0.051

5.5 Parameter Variation Experiment

The parameter variation experiment is an experimental method that observes the impact on the output results by systematically adjusting specific parameters in the model or system. It is important to determine key parameters: identify the parameters that have the greatest impact on the system performance. It plays an important role in determining key parameters; optimizing parameter combinations to find the parameter values that optimize the objective function; and understanding parameter

sensitivity to analyze the patterns of how the model behaves as the parameters change. Colposcopy is an examination method that allows us to visually observe the lesions of the cervix and vagina. The lesion areas can be indicated through the acetic acid test and iodine test under colposcopy. We conduct parameter variation experiments on all data, including whether the colposcopy result is HSIL or above. We assign weights according to different variation ranges.

5.5.1 Effect of Parameter M

To evaluate the influence of the parameter M on the predictive performance, we varied M from 2 to 10. The corresponding MSE values were recorded to determine the optimal parameter setting.

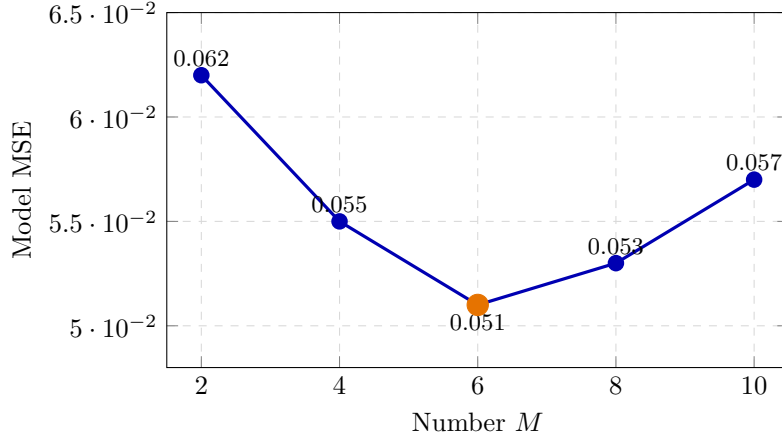


Fig. 2 Effect of parameter M on model performance.

Figure 2 illustrates the effect of M on predictive performance. As M increases from 2 to 10, the MSE first decreases and then increases. The minimum MSE of 0.051 occurs at $M = 6$, which is therefore used in subsequent experiments.

This trend indicates that an excessively small (M) results in an overly coarse feature representation, which may fail to capture important transition patterns. In contrast, an excessively large (M) produces an overly fine-grained representation and may introduce redundant or noisy information, thereby weakening the generalization ability of the model. At ($M=6$), the model achieves an effective balance between feature granularity and generalization performance. Therefore, ($M=6$) is selected as the optimal setting for the subsequent experiments.

5.5.2 Effect of Distillation Temperature T

The temperature parameter T controls the smoothness of the output probability distribution produced by the teacher model during knowledge distillation. To investigate its influence on model performance, we varied T from 1 to 6 and recorded the corresponding MSE values.

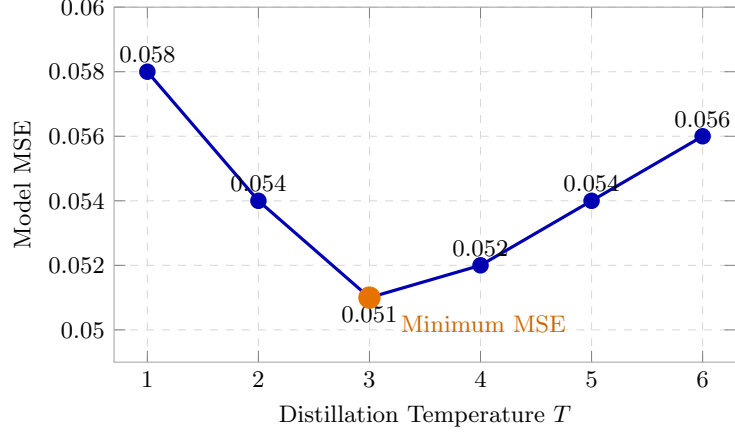


Fig. 3 Effect of the distillation temperature T on model performance.

As shown in Figure 3, the MSE initially decreases as the distillation temperature increases. Specifically, the MSE decreases from 0.058 at $T = 1$ to its minimum value of 0.051 at $T = 3$. However, when T is further increased, the MSE gradually rises to 0.052, 0.054, and 0.056 at $T = 4$, $T = 5$, and $T = 6$, respectively.

A relatively low temperature produces an overly sharp probability distribution, which limits the amount of soft knowledge transferred from the teacher model. In contrast, an excessively high temperature may over-smooth the output distribution and weaken the discriminative information contained in the teacher predictions. Therefore, $T = 3$ provides an appropriate balance between knowledge transfer and prediction accuracy and is selected for the subsequent experiments.

6 Conclusion

This study presented CPS, a cross-domain recommendation framework for integrating heterogeneous clinical information in cervical cancer risk prediction. On the evaluated single-center retrospective cohort, CPS outperformed the selected machine-learning baselines across the reported error and classification metrics. The ablation analysis supported the contribution of the contextual views, while the parameter analyses identified suitable settings for the number of aspects and the distillation temperature. These results support the feasibility of cross-domain learning for clinical risk prediction; however, they do not establish clinical effectiveness. Prospective multicenter validation, calibration analysis, assessment across clinically relevant subgroups, and evaluation against current screening pathways are required before deployment.

Declarations

Funding

This work was supported by the National Key Research and Development Program of China under Grant 2023YFC2705700, the National Natural Science Foundation

of China under Grants 82401876 and 82673960, and the Shandong Natural Science Foundation under Grants ZR2022QH359.

Author Contributions

Conceptualization, J.X.Z.; methodology, J.X.Z.; software, Y.F.; validation, A.D.; formal analysis, Y.W.; investigation, J.X.Z.; resources, J.R.Z.; data curation, B.L.; writing—original draft preparation, Y.F.; writing—review and editing, Y.C. and Y.D.; visualization, M.S.; supervision, Y.Z. and S.H.; project administration, S.H.; funding acquisition, S.H. and Q.C. All authors have read and approved the manuscript.

Data Availability

The anonymized data supporting the findings of this study are available from the corresponding author upon reasonable request, subject to ethical and institutional restrictions.

Ethics Approval

Approval of all ethical and experimental procedures and protocols was granted by the Ethics Committee of Qilu Hospital of Shandong University under Approval Nos. KYLL2021(ZM)-159 and KLYLL-2026(ZM)-1256. The study was performed in accordance with the Declaration of Helsinki.

Consent to Participate

The requirement for informed consent was waived for this retrospective study using anonymized clinical records.

Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

References

- [1] Bi Q, Miao K, Xu N, et al (2024) Habitat radiomics based on MRI for predicting platinum resistance in patients with high-grade serous ovarian carcinoma: A multicenter study. *Acad Radiol* 31(6):2367–2380. <https://doi.org/10.1016/j.acra.2023.11.038>
- [2] Burd EM (2003) Human papillomavirus and cervical cancer. *Clin Microbiol Rev* 16(1):1–17. <https://doi.org/10.1128/CMR.16.1.1-17.2003>
- [3] Chao G, Sun S, Bi J (2021) A survey on multi-view clustering. *IEEE Trans Artif Intell* 2(2):146–168. <https://doi.org/10.1109/TAI.2021.3065894>

- [4] Egemen D, Cheung LC, Chen X, et al (2020) Risk estimates supporting the 2019 ASCCP risk-based management consensus guidelines. *J Low Genit Tract Dis* 24(2):132–143. <https://doi.org/10.1097/LGT.0000000000000529>
- [5] Fong S, Pabis K, Latumalea D, et al (2024) Principal component-based clinical aging clocks identify signatures of healthy aging and targets for clinical intervention. *Nat Aging* 4(8):1137–1152. <https://doi.org/10.1038/s43587-024-00646-8>
- [6] Grabski IN, Street K, Irizarry RA (2023) Significance analysis for clustering with single-cell RNA-sequencing data. *Nat Methods* 20(8):1196–1202. <https://doi.org/10.1038/s41592-023-01933-9>
- [7] Guan S, Xu Z, Yang T, et al (2024) Identifying potential targets for preventing cancer progression through the PLA2G1B recombinant protein using bioinformatics and machine learning methods. *Int J Biol Macromol* 276(Pt 1):133918. <https://doi.org/10.1016/j.ijbiomac.2024.133918>, retracted
- [8] Huang L, Huang XD, Zou H, et al (2025) Knowledge-reinforced cross-domain recommendation. *IEEE Trans Neural Netw Learn Syst* 36(7):12880–12894. <https://doi.org/10.1109/TNNLS.2024.3500096>
- [9] Huang S, Yang J, Shen N, et al (2023) Artificial intelligence in lung cancer diagnosis and prognosis: Current application and future perspective. *Semin Cancer Biol* 89:30–37. <https://doi.org/10.1016/j.semcancer.2023.01.006>
- [10] Huang X, Hu M, Sun T, et al (2023) Multi-kingdom gut microbiota analyses define bacterial-fungal interplay and microbial markers of pan-cancer immunotherapy across cohorts. *Cell Host Microbe* 31(11):1930–1943.e4. <https://doi.org/10.1016/j.chom.2023.10.005>
- [11] Liu Z, Zhang Z, Zhang Y, et al (2024) Spatial transcriptomics reveals that metabolic characteristics define the tumor immunosuppression microenvironment via iCAF transformation in oral squamous cell carcinoma. *Int J Oral Sci* 16(1):9. <https://doi.org/10.1038/s41368-023-00267-8>
- [12] Mathivanan SK, Francis D, Srinivasan S, et al (2024) Enhancing cervical cancer detection and robust classification through a fusion of deep learning models. *Sci Rep* 14(1):10812. <https://doi.org/10.1038/s41598-024-61063-w>
- [13] Mehmood M, Rizwan M, Gregus ml M, et al (2021) Machine learning assisted cervical cancer detection. *Front Public Health* 9:788376. <https://doi.org/10.3389/fpubh.2021.788376>
- [14] Moon I, LoPiccolo J, Baca SC, et al (2023) Machine learning for genetics-based classification and treatment response prediction in cancer of unknown primary. *Nat Med* 29(8):2057–2067. <https://doi.org/10.1038/s41591-023-02482-6>

- [15] Ni R, Cai W, Jiang Y (2024) Contrastive cross-domain sequential recommendation via emphasized intention features. *Neural Netw* 179:106488. <https://doi.org/10.1016/j.neunet.2024.106488>
- [16] Perkins RB, Guido RS, Castle PE, et al (2024) 2019 ASCCP risk-based management consensus guidelines: Updates through 2023. *J Low Genit Tract Dis* 28(1):3–6. <https://doi.org/10.1097/LGT.0000000000000788>
- [17] Rahimi M, Akbari A, Asadi F, et al (2023) Cervical cancer survival prediction by machine learning algorithms: A systematic review. *BMC Cancer* 23(1):341. <https://doi.org/10.1186/s12885-023-10808-3>
- [18] Sanchez-Aguilera A, Masmudi-Martín M, Navas-Olive A, et al (2023) Machine learning identifies experimental brain metastasis subtypes based on their influence on neural circuits. *Cancer Cell* 41(9):1637–1649.e11. <https://doi.org/10.1016/j.ccell.2023.07.010>
- [19] Sung H, Ferlay J, Siegel RL, et al (2021) Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 71(3):209–249. <https://doi.org/10.3322/caac.21660>
- [20] Swanson K, Wu E, Zhang A, et al (2023) From patterns to patients: Advances in clinical machine learning for cancer diagnosis, prognosis, and treatment. *Cell* 186(8):1772–1791. <https://doi.org/10.1016/j.cell.2023.01.035>
- [21] Tran KA, Kondrashova O, Bradley A, et al (2021) Deep learning in cancer diagnosis, prognosis and treatment selection. *Genome Med* 13(1):152. <https://doi.org/10.1186/s13073-021-00968-x>
- [22] Wang X, Su R, Li L, et al (2024) Machine learning-based radiomics for predicting outcomes in cervical cancer patients undergoing concurrent chemoradiotherapy. *Comput Biol Med* 177:108593. <https://doi.org/10.1016/j.compbiomed.2024.108593>
- [23] Xu T, Jiang J, Xiang X, et al (2024) Conduction and validation of a novel prognostic signature in cervical cancer based on the necroptosis characteristic genes via integrating of multiomics data. *Comput Biol Med* 168:107656. <https://doi.org/10.1016/j.compbiomed.2023.107656>
- [24] Yu X, Zhan D, Liu L, et al (2022) A privacy-preserving cross-domain healthcare wearables recommendation algorithm based on domain-dependent and domain-independent feature fusion. *IEEE J Biomed Health Inform* 26(5):1928–1936. <https://doi.org/10.1109/JBHI.2021.3069629>
- [25] Zhang X, Zhao L, Ngo LH, et al (2024) Prediagnostic plasma proteomics profile for hepatocellular carcinoma. *J Natl Cancer Inst* 116(8):1343–1355. <https://doi.org/10.1093/jnci/djae079>

- [26] Zhong X, Lin Y, Zhang W, et al (2023) Predicting diagnosis and survival of bone metastasis in breast cancer using machine learning. Sci Rep 13(1):18301. <https://doi.org/10.1038/s41598-023-45438-z>