Journal of Network Intelligence ©2025 ISSN 2414-8105
Taiwan Ubiquitous Information Volume **10**, Number **1**, February **2025**

Intelligent Management of Library and Information Resources Based on Multi-Label Reduction and Multi-Modal Deep Learning

Kai Jiang*
Liaocheng University Dongchang College, Liaocheng 252000, P. R. China
kaijiangabc@163.com

Xiang Yan
Institute of Technology
Pu'er University, Pu'er 665000, P. R. China
15891983287@163.com

Wei-Guo Hwang
College of Engineering
Mariano Marcos State University, Batac City 2906, Philippines
nj9703@163.com

*Corresponding author: Kai Jiang
Received February 19, 2024, revised May 29, 2024, accepted August 2, 2024.

Abstract. *Intending to the existing library intelligence classification and management methods in which some sample data features are high-dimensional and single-modal information is insufficiently expressed, which leads to inefficient classification, this article suggests a library intelligence resource intelligent management method based on multi-label reduction and multi-modal deep learning. Firstly, the optimal singular vector is found as the label projection direction on the ground of Singular Value Decomposition (SVD), and then the sparse labels are encoded into the low-dimensional effective space. After that, relied on the enhanced multi-label dimensionality reduction algorithm, the image and text features of book intelligence are extracted in the low-dimensional space, and then the implicit features of image and text are mined by the deep learning model, and the features of the two modalities are interacted and spliced by the cross-modal attention mechanism to obtain the overall implicit representation. Finally, the spliced vectors are input to softmax classifier to achieve the predicted probability of each type, and a recursive regularization term is added to the error during training to take into account the hierarchical structure information of labels, for the sake of realizing the intelligent management of library intelligence resources. The experimental outcome indicates that the accuracy, precision, and recall of the algorithm suggested in this article are 92.6%, 90.8%, and 93.2%, respectively, which has a good performance of classification management.*
Keywords: Multilabel dimensionality reduction; Multimodal learning; Deep learning; Singular value decomposition; Attention mechanisms

---

1. **Introduction.** As a cultural information service organization, libraries first need to obtain a large number of library intelligence resources, and can better provide readers with information services after their development and utilization [1]. The introduction of artificial intelligence technology makes it easier and faster to obtain library intelligence resources, and plays an important role in digitizing intelligence resources, organizing and

managing resources, etc. It also improves the quality and level of information services, and can provide readers with more personalized and specialized services [2, 3, 4]. At present, the scale of the intelligent library is developing, with a large inventory and a wide range of books, the intelligent management of library and intelligence resources becomes a problem. Intelligent libraries first need to be able to have a high degree of accuracy in the categorization of library information resources in order to ensure its operational development [5, 6]. The inaccuracy of the classification of library information resources management will not only affect the readers' access, but also increase the tediousness of the work of the administrators.

1.1. **Related work.** The research on the classification and management of book resources mainly includes several aspects: digitalization, big data mining, metadata standards, intelligent management and personalized recommendation.

Nohrstedt [7] determined the categories based on the probability of occurrence of the index word in the class and used this method for classification of library and intelligence resources. Zeng et al. [8] suggested the concept of Vector Space Model (VSM) and used it to represent library and intelligence resources. Rodrawangpai and Daungjaiboon [9] utilized a bi-directional transformer as an encoder to fully take into account the information in both the front and back directions of the word, but the feature representation capability is insufficient. Guo et al. [10] proposed self-correcting class weight vectors based on readers' feedback to construct simple linear classifiers to categorize intelligence resources. Meng et al. [11] offered a tree-like classification structure for classifying book intelligence, but the accuracy of the classification is not very high. Anglada [12] proposed a model called Potential Semantic Analysis, a model for intelligent management of library intelligence, which utilizes theoretical methods related to statistical computing in order to obtain potential semantic information between words. Hajbaqeryian and Mahfoozi [13] divided the initial label set into several random small label sets and classified the books based on BP neural network. Zhou [14] used Apriori algorithm and tree structure to compute the large-scale book intelligence resources to generate the text frequent itemset, so as to predict the category of the books, but the prediction accuracy is not high.

Large-scale and high-dimensional intelligence resources bring challenges to book management. Sanhong et al. [15] used to simultaneously optimize the predictability of the low-dimensional label space and the recoverability of the original label space to manage book intelligence resources are crucial for various applications. However, the high dimensionality of the feature space is often neglected. Fu et al. [16] used principal component analysis to spatially downscale books and classify book intelligence resources, but the classification accuracy is not high. Ntakaris et al. [17] combined integrated learning and labeled power set algorithms to integrate multiple classifiers to ensure the completeness of the prediction of library intelligence resources, but the classification efficiency is low. To improve the classification performance, Huang et al. [18] introduced a residual network based on the fusion of a text structure detector model to classify books. Nithya et al. [19] introduced deep learning models such as CNN, LSTM, BERT, etc., to manage library intelligence resources based on the contents of sentences, paragraphs, and chapters.

Some deep neural network-based approaches face the challenge of insufficient multimodal training data, which limits the effectiveness of training and can easily lead to overfitting. He et al. [20] proposed Adversarial Cross-Modal Retrieval (ACMR) to minimize the gap between all the books in a single modality. The representation of all books is in a single modality, leading to insufficient information representation. Dong et al. [21] used a graphical convolutional neural network to establish the dependency relationship between book resources and used a bi-directional attention-based generative network to realize inter-modal transformation, but some of the necessary information was lost.

1.2. **Motivation and contribution.** Aiming at the issues of insufficient expression of single-modal information and missing data of some samples in the existing research, this article designs a method of intelligent management of library intelligence resources based on multi-label reduction and multi-modal deep learning.

Firstly, the optimal projection direction of a set of label space is found based on Singular Value Decomposition (SVD), and the sparse labels are embedded into the low-dimensional space. Besides relied on the optimized multi-label dimensionality reduction algorithm, a set of low-dimensional label representations is achieved by maximizing the dependency between low-dimensional features and potential labels, image and text features of book intelligence are extracted on the low-dimensional space.

Then the features of the two modalities interact with each other through the cross-modal attention mechanism, and spliced together, to get the implicit representations of book intelligence. Finally, the spliced vectors are input to the classifier to get the predicted probability of each category. The hierarchical structure information of the labels is considered during the training, so as to realize the intelligent management of book intelligence resources. The experimental outcome indicates that the suggested method has high classification accuracy and efficiency, and can be better applied to the field of library intelligence resource management.

2. Theoretical analysis.

2.1. **Multi-label dimensionality reduction algorithm.** Linear Discriminant Analysis (LDA) method [22] is a typical supervised linear dimensionality reduction method, which is only applicable to single-label multi-class classification. In fact, the class labels among multi-label data are interdependent, and the dependence of sample feature attributes and labels can be considered to solve the problem of poor class separation. The Multi-label LDA (MLDA) method [23] takes into account the dependency between features and labels by assigning different weights to different labels.

Define a non-negative weight matrix $V_{m\times p} = [v_1, ..., v_j, ..., v_m]^T = [v^1, ..., v^j, ..., v^p]$, where $v_j = [v_{j1}, v_{j2}, ..., v_{jp}]^T$ and $v^i = [v_{i1}, v_{i2}, ..., v_{im}]^T$ denote the weight vector of the i-th sample and the weight vector of the j-th class label, respectively, and $v_{ij} \geq 0$ denotes the weight of the j-th class label belonging to the i-th sample. Given an input space $X_{m\times c} = [x_1, x_2, ..., x_m]^T$, denote m instances in a d-dimensional space. In MLDA, the inter-class matrix T_c and intra-class matrix T_v are represented as follows.

$$T_c = \sum_{i=1}^{p} T_c^{(i)}, \quad T_c^{(i)} = \left(\sum_{j=1}^{m} v_{ij}\right)(n_i - n)(n_i - n)^T \tag{1}$$

$$T_v = \sum_{i=1}^{p} T_v^{(i)}, \quad T_v^{(i)} = \sum_{j=1}^{m} V_{ij}(x_j - n_i)(x_j - n_i)^T \tag{2}$$

where n_i denotes the mean of the j-th class and n is the labeled global mean.

The idea of MLDA is to minimize the intra-class scattering matrix and maximize the inter-class scattering matrix, and the final standard optimization objective is as follows.

$$G^* = \arg\max_G \left(\frac{G^T T_c G}{G^T T_v G}\right) \tag{3}$$

2.2. **Multimodal Learning.** Multimodal Machine Learning (MMML) [24] mainly includes representation, transformation, alignment and fusion methods, of which the most central one is the multimodal fusion method. The model architecture is indicated in Figure 1.

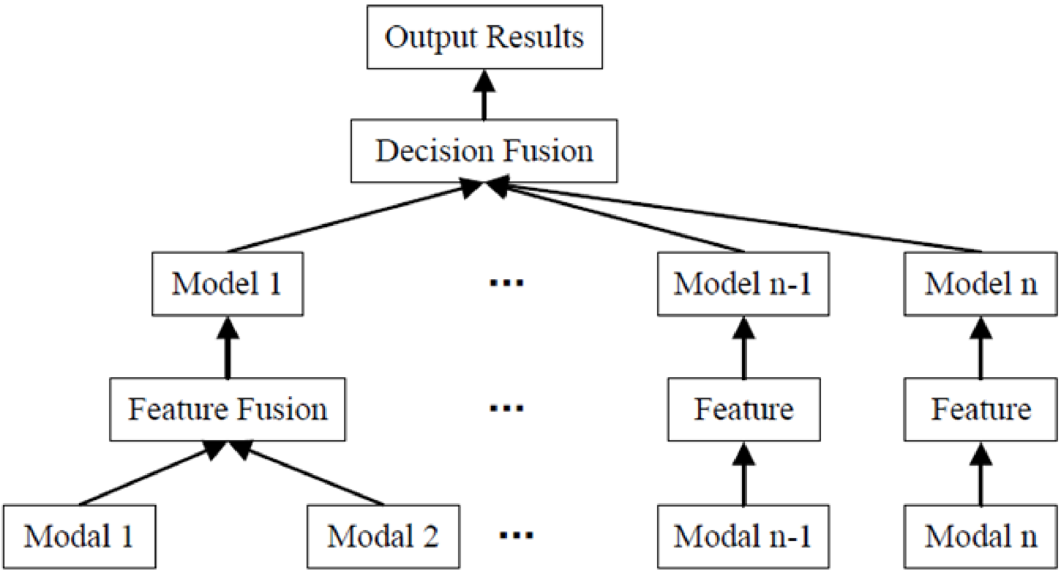

Figure 1. Architecture of the multimodal fusion model

Model-based multimodal methods utilize deep learning models to address the multimodal fusion issue, and the image model approach fuses shallow and deep graphics through image segmentation, splicing, and prediction to generate modal fusion results. The neural network approach is one of the most widely used methods, which provides flexible single-modal implicit representation extraction and multimodal representation splicing for different modal inputs, and utilizes the strong learning ability and classification performance of neural networks to implicitly adjust the fusion parameters, which is more scalable.

3. **Optimization of Multi-label Dimensionality Reduction Algorithms.** Focusing on the issues of feature height and label sparsity in the MLDA dimensionality reduction method, this article suggests a multi-label dimensionality reduction algorithm based on singular value decomposition and stream regularization, which firstly finds the optimal projection direction of a set of label space on the ground of the SVD method, and then embeds the sparse labels into the low-dimensional space. On this basis, a linear discriminant analysis relied on stream regularization is suggested to ensure that the dependence among the low-dimensional characteristics and the embedded tags in the geometric space is maximized. Finally, the classification model is studied on the low-dimensional effective space, which lays the foundation for the subsequent classification of library intelligence resource management.

In the preprocessing stage, the high-dimensional sparse labeling space $Y_{m\times p}$ is transformed into the potential labeling space $Y'_{m\times l}$ using the SVD method [25]. It is known that $Y_{m\times p}$ is able to be disintegrated to three matrices.

$$Y_{m\times p} = W_{m\times m} \cdot \Sigma_{m\times p} \cdot V_{p\times p} \tag{4}$$

where matrix Σ is a diagonal matrix when $i \neq j$, $\Sigma_{i,j} = 0$, and the diagonal elements of Σ are the peculiar values of the labeled matrix Y in descending order. In order to reduce the noise, the singular values smaller than $\sum_{l,j}$ and their corresponding basis vectors are ignored.

Equation (4) is simplified as follows.

$$Y'_{m\times p} \approx W'_{m\times l} \cdot \Sigma'_{l\times l} \cdot V'^{T}_{l\times p} \tag{5}$$

where the matrix $Y'_{m\times p}$ is close to $Y_{m\times p}$, i.e., $||Y - Y'||_{\kappa} < \mathcal{Q}$. $W'_{m\times l}$, $\Sigma'_{l\times l}$, and $V'_{l\times p}$ are simplified matrices after removing the $c-l$ singular values. The orthonormal basis $V'_{l\times p}$ is able to be considered as a tiny projection matrix of Y that maps the p-dimensional labels into the l-dimensional space ($l < p$), i.e., $Y'_{m\times l} = Y_{m\times p}V'^{T}_{l\times p}$.

To effectively exploit the dependency of low-dimensional labels and feature attributes, the MLDA method is optimized using stream regularization [26]. If samples x_i and x_j are similar in the geometric structure of feature distribution, their corresponding labels are also similar, otherwise the opposite is true. The flow regularization is defined as below.

$$\sum_{i,j} \frac{v_{i,j}}{2} ||x'_i - x'_j||^2 = \mathrm{tr}(X^T K X) \tag{6}$$

where x'_i denotes the sample characteristic of the latent feature space, $K = C - V'$ is the Laplace matrix of the degree matrix V', and C is the diagonal matrix $C_{ii} = \sum_{j=1}^{m} v_{i,j}$. The degree matrix V is achieved through computing the Gaussian distance among the tag vectors in the latent label space.

$$v_{i,j} = \begin{cases} \exp\left(-\frac{||y'_i - y'_j||^2}{g^2}\right), y'_i \in N_l(y'_j) \text{ or } y'_j \in N_l(y'_i), \\ 0, \quad \text{otherwise} \end{cases} \tag{7}$$

where $N_l(x_i)$ denotes the l nearest neighbors of x_i.

The linear combination function of stream shape regularization and MLDA is then adopted as the whole objective operation as indicated below.

$$\begin{aligned} G^* &= \arg \max_{P^T P = I} \mu \mathrm{tr}(P^T X^T X P) - (1-\mu)\mathrm{tr}(P^T X^T K X P) \\ &= \arg \max_{P^T P = I} \mathrm{tr}(P^T X^T (\mu + (1-\mu)K) X P) \end{aligned} \tag{8}$$

where μ is the equilibrium parameter, the problem can be converted into a feature solving problem by finding the top p largest eigenvalues and their corresponding eigenvectors, which form an optimal feature expulsion matrix P^*, and then finding a set of valid low-dimensional feature representation $X'_{m\times p} = X_{m\times c}P^*_{c\times p}$.

4. **Intelligent management of library and intelligence resources based on Multi-Label reduction and multi-modal deep learning.** On the ground of the above enhanced multi-label dimensionality reduction algorithm, firstly, a set of low-dimensional label representations are achieved by maximizing the dependency between low-dimensional features and potential labels, and the image and text features of book intelligence are extracted on the low-dimensional space.

Then the implicit features of images and texts are mined by the deep learning model, and the features of the two modalities interact with each other through the cross-modal attention mechanism, and then spliced together, to obtain the book implicit representation of intelligence. Finally, the spliced vectors are input to softmax classifier to get the predicted probability of each category, and the recursive regularization term is added to the error during training to take into account the hierarchical structure information of the labels, so as to realize the intelligent management of book intelligence resources. The overall model structure is indicated in Figure 2.

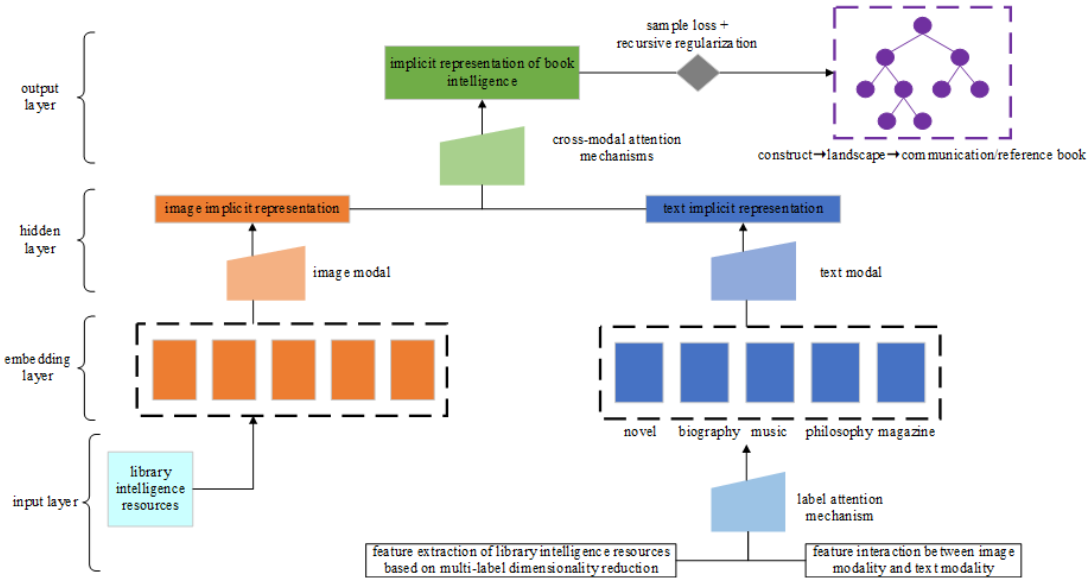

Figure 2. The overall model structure of the designed method

4.1. **Feature extraction of library intelligence resources based on multi-label dimensionality reduction.** Assume that the training set of library and information resources is $X \in \mathbb{R}^{m\times c}$, the label matrix is $Y \in \mathbb{R}^{m\times p}$, η is the equilibrium parameter, and q is the low-dimensional feature space dimension. The steps in detail are as follows.

1) The feature dependence matrix $\Sigma_{c\times c} = X^T X$ is constructed from the eigenmatrix $X_{m\times c}$, the feature decomposition dependency matrix $\Sigma_{c\times c}$ is constructed, and then the eigenvectors corresponding to the first q largest eigenvalues are selected to form the optimal projection matrix $Q^* \in \mathbb{R}^{c\times q}$.

2) Project the eigenmatrix $X_{m\times c}$ to the low-dimensional eigenmatrix: $X'_{m\times q} = X_{m\times c} \cdot Q^*_{c\times q}$. Solve the eigenvector: $A = Y^T(I + \beta WKW)Y$ and then select the eigenvector corresponding to the first L maximum eigenvalues to form the optimal label projection matrix $V^* \in \mathbb{R}^{p\times k}$.

$$\min_{V^TV=I} E(V) = \max_{V^TV=I} \operatorname{tr}(V^TY^T(I + \eta WKW)YV) \tag{9}$$

where $V \in \mathbb{R}^{p\times l}$ is a linear projection matrix, W denotes the feature vector, $\operatorname{tr}(\cdot)$ denotes the feature space encoding based on dependency maximization, and K is a symmetric matrix that searches for l nearest neighbors for each book.

3) Project the labeling matrix $Y_{m\times p}$ to the lower-dimensional labeling matrix: $Z_{m\times l} = Y_{m\times p} \cdot V^*_{p\times k}$.

4) The SVD decomposition of the library intelligence resource feature matrix X, i.e., $X = U \times \Sigma \times S^T$, where U and S represent the two orthogonal matrices of X. The n-rank optimal approximation matrix of X is: $X_n = U_n \times \Sigma'_n \times S_n^T$, where U_n denotes the first n columns of matrix V, U_n^T denotes the first n rows corresponding to matrix U^T, and $\Sigma'_n = \operatorname{diag}(\delta_1, \delta_2, \ldots, \delta_n)$ contains the diagonal matrices of the first n largest singular values. The image and text matrices after feature extraction. The i-th book x_i is decomposed by SVD as follows.

$$X_k = X_i^T U_l \Sigma_n^{-1} \tag{10}$$

The features of book X_i are reduced from k to n. Next, the downscaled processed $X = (x_1, x_2, \ldots, x_m)^T$ is used as an input to the deep learning model for training the classifier.

4.2. **Multimodal Integration of Library and Intelligence Resources.** Since library intelligence resources are divided into picture and text information, which are in different modalities, the implied feature vectors of different modalities need to be spliced together to form longer feature vectors, which are then input into the classifier for classification prediction. In addition, the feature lengths and dimensions of text and images are not the same to compute the similarity, so it is necessary to use the attention mechanism [27] to initialize the weight parameters and update the parameters automatically through training.

Assuming that there are text modality a and image modality b to be interacted with attentively, taking modality a as the question and modality b as the knowledge base, the similarity between the question and each key in the knowledge base is computed to obtain the corresponding weight coefficients, which are calculated as below.

The input matrix for modality a is $X_a \in \mathbb{R}^{T_a \times c_a}$, where T_a refers to the sequence length of text modality a and c_a refers to the feature dimension of each representation of modality a. The input matrix for modality b is $X_b \in \mathbb{R}^{T_b \times c_b}$, where T_b and c_b refer to the sequence length and feature dimension of modality b.

A learnable parameter matrix $H_{Q_a} \in \mathbb{R}^{c_* \times c_*}$ is constructed from the dimensions of the input matrix X_a. Q_a is computed as indicated in Equation (11).

$$Q_a = X_a H_{Q_a} \in \mathbb{R}^{T_a \times c_*} \tag{11}$$

Two learnable parameter matrices $H_{K_b} \in \mathbb{R}^{c_* \times d}$ and $H_{V_b} \in \mathbb{R}^{c_* \times d}$ are constructed based on the dimensions of the input matrix X_b. K_b and V_b are computed as indicated in Equation (12) and Equation (13), respectively.

$$K_b = X_b H_{K_b} \in \mathbb{R}^{T_b \times c_*} \tag{12}$$

$$V_b = X_b H_{V_b} \in \mathbb{R}^{T_b \times c_*} \tag{13}$$

Thereby, question Q_a and key K_b are mapped to the same feature space, and the similarity between each element in Q_a and each element in K_b is computed and the weights obtained are indicated below.

$$Weight = Softmax\left(\frac{Q_a K_b^T}{\sqrt{c_k}}\right) \in \mathbb{R}^{T_a \times T_b} \tag{14}$$

where $Softmax(x) = e^x / \sum_{j=1}^{c} e^x$.

To prevent the results of the dot product of Q_a and K_b^T from being too large, the results of each component are divided by $\sqrt{c_k}$ thereby converting each component of the dot product into a number with a mean of 0 and variance of 1.

Finally, F_a is obtained by weighting and summing the values according to the weights of the keys.

$$F_a = Weight * V_b \in \mathbb{R}^{T_a \times c_*} \tag{15}$$

In summary, the whole equation for the cross-modal attention mechanism is as follows.

$$F_a = Softmax\left(\frac{X_a H_{Q_a} (X_b H_{K_b})^T}{\sqrt{c_k}}\right) X_b H_{V_b} \tag{16}$$

A new splice vector $D = [Q_a, K_b, V_b, F_a]$ is generated by computing the attention mechanism from text modality to picture modality and from picture modality to text modality, respectively, and then stitching the two together.

4.3. **Intelligent management of library and intelligence resources based on multi-label reduction and multi-modal deep learning.** The above input text modal and picture modal splicing vector D, through the classification model to obtain the overall implicit representation of book intelligence $h \in \mathbb{R}^{f_*}$. Through the $f_* \times c$-dimensional fully connected layer, the weight matrix is multiplied with the input vector plus the bias, and the c-dimensional $(-\infty, +\infty)$ real numbers are mapped to the c-dimensional $[0, 1]$ real numbers through the softmax function to ensure that the sum is 1, so as to obtain the probability distribution of the current book intelligence on each category as follows.

$$\hat{y} = P(y|h) = Softmax(H^T h + g) \tag{17}$$

where $H \in \mathbb{R}^{f_* \times c}$ is the weight matrix and $g \in \mathbb{R}^c$ is the bias term.

For each book intelligence sample i, the cross-entropy loss is as below.

$$I_i^{CE} = -y_i \log \hat{y}_i - (1 - y_i) \log(1 - \hat{y}_i) \tag{18}$$

Then this paper considers adding a regularization term of the label structure to the loss function to introduce the category hierarchy information, so as to improve the accuracy of the classification management. Suppose that for any node label $l \in L$ in the hierarchical classification tree, the corresponding fully connected layer parameter of the softmax classifier is w_l. All of its children nodes are labeled as $l' \in \pi(l)$, and the corresponding parameter is $w_{l'}$. The following regularization term is added to the loss function.

$$R(w) = \sum_{l \in L} \sum_{l' \in \pi(l)} \frac{1}{2} \|w_l - w_{l'}\|^2 \tag{19}$$

For the current classification scenario of library intelligence resources, all the books go to a unique leaf node, and the fully connected layer does not have the parameters of each parent node, for this situation, this paper utilizes Equation (20) to compute w_l, so that it can be applied to the classification management of library intelligence.

$$w_l = \frac{1}{|\pi(l)|} \sum_{l' \in \pi(l)} w_{l'} \tag{20}$$

Finally, recursively upward in order to get the parameter matrix of all nodes of the classification tree, and then calculate the recursive regularization loss based on these parameters, the smaller the loss, the more similar the parameters of each sibling node, the overall loss function is calculated as bellow.

$$L_{Total} = L_{FL} + \lambda \cdot R(W) \tag{21}$$

where λ is a hyperparameter that balances the sample loss and recursive regularization loss.

5. **Performance testing and analysis.**

5.1. **Experimental results and analysis.** To estimate the performance of the designed algorithm, this article randomly selects book category data from a university library for simulation experiments, including titles, descriptions, and cover images of five categories of books, such as novels, biographies, etc. The dataset contains a total of 19,832 bibliographic items, and is divided into training, testing and validation sets in accordance with 5:3:2. ECOI [13], AGCN [21] and OURs (the classification management algorithm for library intelligence designed in this paper), are trained and tested respectively. All models adopt softmax classifiers, and the optimizer uses the Adam model with a learning rate of 0.0005. The decay coefficient of 0.95. Small batch training is set, and the size of each batch is set to 64. The experimental platform is configured with Windows 11, 64-bit operating system, Intel(R) Core (TM) i5-13500H CPU @ 120Hz, 16 GB RAM, and the simulation experiment environment is Matlab R2017a.

This article adopts a combination of multi-branch classification loss function (LOSS) [28], Accuracy (Acc), Precision (Prec), Recall (Rec) and F1 value to evaluate the effectiveness of classification management of library intelligence resources. Table 1 demonstrates the comparative results of each metric for different models. Through the analysis, it is concluded that the loss of OURs method is less than 0.8 compared with ECOI and AGCN library and intelligence classification management methods, which is reduced by 11.71% and 5.82% compared with ECOI and AGCN, respectively, and improved in four evaluation indexes, namely, Acc, Prec, Rec and F1 value. It was improved by 21.2% and 11.43% on Acc, 16.26% and 7.33% on Prec, 18.73% and 12.42% on Rec, and 17.5% and 9.92% on F1, respectively.

This model outperforms the ECOI method based on traditional multi-label dimensionality reduction and the AGCN method based on multimodal and graph convolutional networks on all five metrics, which is attributed to the following reasons: the OURs method improves the traditional multi-label dimensionality reduction method, maximizes the dependency between low-dimensional features and potential labels, and deeply extracts the implicit representations of images and texts in book intelligence, fully considering the label's hierarchical structure information, and therefore can get more accurate classification outcome.

Table 1. Performance comparison results for different models

Model	LOSS	Acc	Prec	Rec	F1
ECOI	0.897	0.764	0.781	0.785	0.783
AGCN	0.841	0.831	0.846	0.829	0.837
OURs	0.792	0.926	0.908	0.932	0.920

In addition, the main statistical evaluation indexes used in this paper include ROC curve and AUC value, as indicated in Figure 4, the ROC curve is drawn with the false-positive rate as the horizontal coordinate and the T-true-positive rate as the vertical coordinate. When the ROC curve is applied to the model evaluation, the smaller the horizontal coordinate and the bigger the vertical coordinate are, the higher the accuracy of the model is. Moreover, the AUC value of each model can be calculated to compare the classification performance of different methods horizontally.

The AUC values of OURs and AGCN methods are greater than 0.5, indicating that the classification stability and stability of these two methods are stronger. The AUC value of ECOI method is the smallest, because it only uses the traditional multi-label dimensionality reduction method to process the book intelligence, and does not optimize the algorithm itself and does not take into account the internal features of the book, which leads to a poorer classification effect. The AGCN method only uses the deep learning models to mining features of different modalities, but does not consider the dependency among low-dimensional characteristics and potential tags, resulting in a lower classification effect than the OURs method.

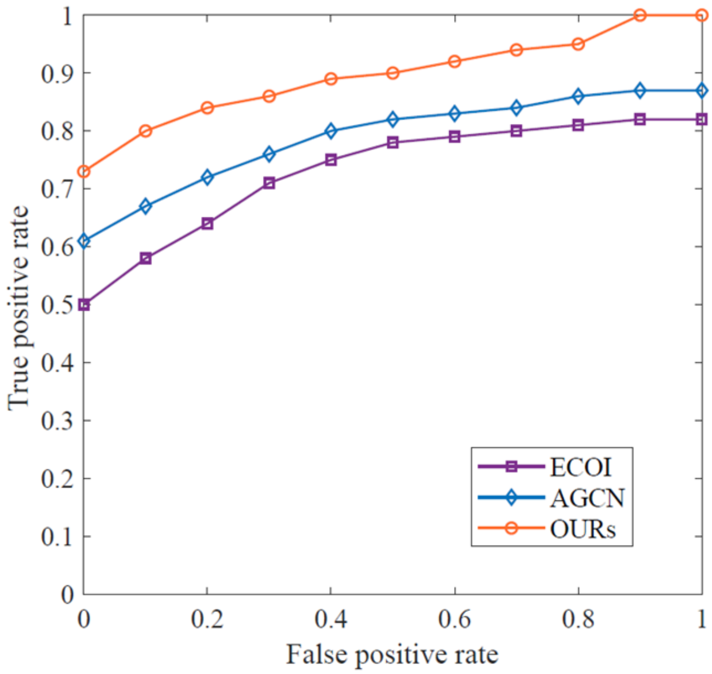

Figure 3. Comparison of ROC curves for different methods

5.2. **Ablation experiment.** To better validate the impact of the multi-label dimensionality reduction algorithm and the multimodal fusion module in this paper's model OURs, ablation experiments are conducted on the dataset respectively, and two comparative models are designed to be analyzed, and the evaluation metrics are adopted as MAE, MAPE, RMSE, and RMSPE.

(1) Remove the multilabeled dimensionality reduction module. Directly use the multimodal deep learning model to categorize and manage book intelligence resources, denoted as OURs-MDR.

(2) Removal of Multimodal Fusion Module. The sparse labels are embedded into the low-dimensional space using an optimized multi-label dimensionality reduction algorithm, and the deep learning model is used to classify and manage the book intelligence resources, denoted as OURs-MMF.

The outcome of the ablation experiment are indicated in Figure 3, the MAE values of the three models decrease gradually with the increase of the number of library information and then stabilize, and the OURs method always has smaller MAE values. Meanwhile, comparing the multimodal deep learning model OURs-MDR and the model OURs-MMF, which only optimizes the multilabel dimensionality reduction algorithm without considering the internal representation of modality, it can be seen that there is not much difference between the two models in terms of MAE value when the number of books is small. However, with the increase of book bibliographies, the OURs in this paper decrease at a higher rate than the OURs-MDR model and the OURs-MMF model, which suggests that the optimized multi-label dimensionality reduction algorithm maximizes the dependency between the low-dimensional features and the potential labels, and the use of multimodal deep learning to further mine the internal features of images and texts in book intelligence, which results in a smaller classification management error and proves that the OURs have a better performance.

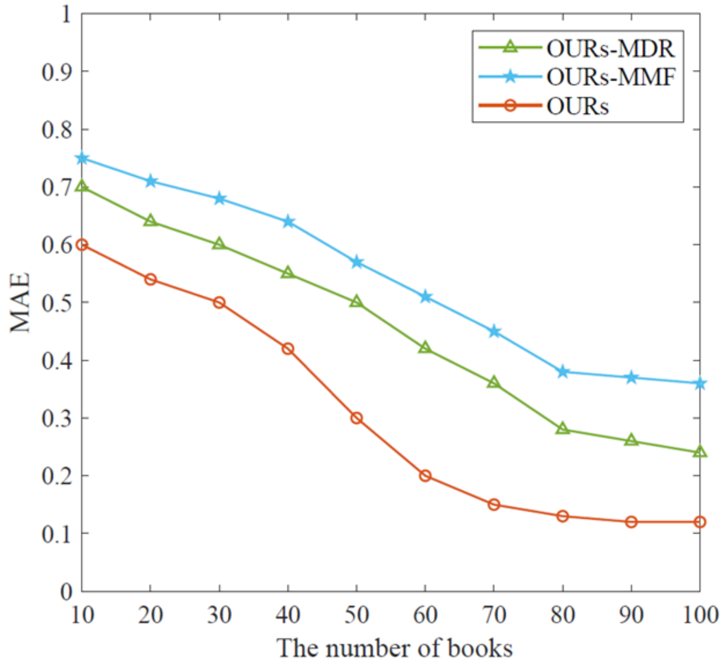

Figure 4. Comparison of MAE for different models

The outcome of the three model evaluation indexes are given in Table 2, and the accuracy evaluation indexes of OURs are significantly better than those of OURs-MDR and OURs-MMF. Among them, the MAE of OURs is 0.1257, which is reduced by 0.1158 and 0.2371 compared with that of OURs-MDR and OURs-MMF, respectively, and the RMSPE of OURs method is 0.1589, which is reduced by 0.1262 and 0.2259 compared with that of OURs-MDR and OURs-MMF, respectively. OURs-MDR and OURs-MMF are reduced by 0.1262 and 0.2259 respectively. Moreover, in the classification results, the MAPE value and RMSE value of the OURs model are lower than 0.15, which reinforces the fact that the OURs method is more suitable for solving the problem of library intelligence management.

Table 2. Comparison of classification accuracy of different models

Method	MAE	MAPE	RMSE	RMSPE
OURs-MDR	0.2415	0.2632	0.3074	0.2651
OURs-MMF	0.3628	0.3928	0.4326	0.3648
OURs	0.1257	0.1432	0.1492	0.1389

6. **Conclusion.** Focusing on the issue of inefficient classification in the existing library intelligence resource management algorithms, this article suggests an intelligent management method for library intelligence resources based on multi-label dimensionality reduction and multimodal deep learning. Firstly, the optimal projection direction of a set of label space is found based on SVD, and the sparse labels are embedded into the low-dimensional space. Then relied on the optimized multi-label dimensionality reduction algorithm, a set of low-dimensional label representations is achieved by maximizing the dependency between low-dimensional features and potential labels, image and text

features of book intelligence are extracted on the low-dimensional space, then the implicit features of images and texts are mined, and the features of the two modalities are interacted and spliced by the cross-modal attention mechanism. Finally, the spliced vectors are input to the classifier to get the predicted probability of each category, and the hierarchical structure information of labels is taken into account during training, so as to realize the intelligent management of book intelligence resources. The experimental outcome indicates that the suggested algorithm has higher classification performance and efficiency compared with the comparison algorithms.

Acknowledgment. This work is supported by the Pu'er School level project (No. PEXYXJYB202310).

REFERENCES

[1] M. Y. Ali, S. B. Naeem, and R. Bhatti, "Artificial intelligence tools and perspectives of university librarians: An overview," *Business Information Review*, vol. 37, no. 3, pp. 116–124, 2020.

[2] P. S. Nitse, and K. R. Parker, "Library Science, Knowledge Management, Competitive Intelligence: Archive Theory—The Common Link," *The Reference Librarian*, vol. 38, no. 79–80, pp. 395–407, 2003.

[3] A. Asemi, A. Ko, and M. Nowkarizi, "Intelligent libraries: a review on expert systems, artificial intelligence, and robot," *Library Hi Tech*, vol. 39, no. 2, pp. 412–434, 2020.

[4] T.-Y. Wu, H. Li, S. Kumari, and C.-M. Chen, "A Spectral Convolutional Neural Network Model Based on Adaptive Fick's Law for Hyperspectral Image Classification," *Computers, Materials & Continua*, vol. 79, no. 1, pp. 19–46, 2024.

[5] Y. Ma, Y. Peng, and T.-Y. Wu, "Transfer learning model for false positive reduction in lymph node detection via sparse coding and deep learning," *Journal of Intelligent & Fuzzy Systems*, vol. 43, no. 2, pp. 2121–2133, 2022.

[6] F. Zhang, T.-Y. Wu, and G. Zheng, "Video salient region detection model based on wavelet transform and feature comparison," *EURASIP Journal on Image and Video Processing*, vol. 2019, 58, 2019.

[7] D. Nohrstedt, "Shifting resources and venues producing policy change in contested subsystems: a case study of Swedish signals intelligence policy," *Policy Studies Journal*, vol. 39, no. 3, pp. 461–484, 2011.

[8] W. Zeng, H. Xu, and J. Zhang, "Term extraction and correlation analysis based on massive scientific and technical literature," *International Journal of Computational Science and Engineering*, vol. 15, no. 3–4, pp. 248–255, 2017.

[9] B. Rodrawangpai, and W. Daungjaiboon, "Improving text classification with transformers and layer normalization," *Machine Learning with Applications*, vol. 10, 100403, 2022.

[10] P.-F. Guo, L.-X. Du, and J.-X. Qi, "A Library Book Intelligence Classification System based on Multi-agent," *Physics Procedia*, vol. 24, pp. 2187–2193, 2012.

[11] J. Meng, H. Lin, and Y. Yu, "A two-stage feature selection method for text categorization," *Computers & Mathematics with Applications*, vol. 62, no. 7, pp. 2793–2800, 2011.

[12] L. M. Anglada, "Collaborations and alliances: social intelligence applied to academic libraries," *Library Management*, vol. 28, no. 6, pp. 406–415, 2007.

[13] T. Hajbaqeriyan, and M. Mahfoozi, "Criticism of a Book Titled as Class BP: Extended Classification of Islam in Library of Congress Classification," *Library and Information Science Research*, vol. 10, no. 1, pp. 152–184, 2020.

[14] Y. Zhou, "Design and implementation of book recommendation management system based on improved Apriori algorithm," *Intelligent Information Management*, vol. 12, no. 3, pp. 75–87, 2020.

[15] D. Sanhong, F. Yuyangzi, and W. Hao, "Multi-Label Classification of Chinese Books with LSTM Model," *Data Analysis and Knowledge Discovery*, vol. 1, no. 7, pp. 52–60, 2017.

[16] J. Fu, C. Huang, J. Xing, and J. Zheng, "Pattern classification using an olfactory model with PCA feature selection in electronic noses: Study and application," *Sensors*, vol. 12, no. 3, pp. 2818–2830, 2012.

[17] A. Ntakaris, M. Magris, J. Kanniainen, M. Gabbouj, and A. Iosifidis, "Benchmark dataset for mid-price forecasting of limit order book data with machine learning methods," *Journal of Forecasting*, vol. 37, no. 8, pp. 852–866, 2018.

[18] S. Huang, R. Dai, J. Huang, Y. Yao, Y. Gao, F. Ning, and Z. Feng, "Automatic modulation classification using gated recurrent residual network," *IEEE Internet of Things Journal*, vol. 7, no. 8, pp. 7795–7807, 2020.

[19] K. Nithya, M. Krishnamoorthi, S. V. Easwaramoorthy, C. Dhivyaa, S. Yoo, and J. Cho, "Hybrid approach of deep feature extraction using BERT–OPCNN & FIAC with customized Bi-LSTM for rumor text classification," *Alexandria Engineering Journal*, vol. 90, pp. 65–75, 2024.

[20] S. He, W. Wang, Z. Wang, X. Xu, Y. Yang, X. Wang, and H. T. Shen, "Category alignment adversarial learning for cross-modal retrieval," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 5, pp. 4527–4538, 2022.

[21] X. Dong, L. Liu, L. Zhu, L. Nie, and H. Zhang, "Adversarial graph convolutional network for cross-modal retrieval," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 3, pp. 1634–1645, 2021.

[22] A. Tharwat, T. Gaber, A. Ibrahim, and A. E. Hassanien, "Linear discriminant analysis: A detailed tutorial," *AI Communications*, vol. 30, no. 2, pp. 169–190, 2017.

[23] X. Shu, H. Xu, and L. Tao, "A least squares formulation of multi-label linear discriminant analysis," *Neurocomputing*, vol. 156, pp. 221–230, 2015.

[24] T. Baltrušaitis, C. Ahuja, and L.-P. Morency, "Multimodal machine learning: A survey and taxonomy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 2, pp. 423–443, 2018.

[25] V. Klema, and A. Laub, "The singular value decomposition: Its computation and some applications," *IEEE Transactions on Automatic Control*, vol. 25, no. 2, pp. 164–176, 1980.

[26] B. Geng, D. Tao, C. Xu, L. Yang, and X.-S. Hua, "Ensemble manifold regularization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 6, pp. 1227–1233, 2012.

[27] Z. Niu, G. Zhong, and H. Yu, "A review on the attention mechanism of deep learning," *Neurocomputing*, vol. 452, pp. 48–62, 2021.

[28] X. Luan, Y. Zhao, W. Ou, L. Liu, W. Li, Y. Shu, and H. Geng, "Collaborative learning with a multi-branch framework for feature enhancement," *IEEE Transactions on Multimedia*, vol. 24, pp. 929–941, 2021.