https://doi.org/10.31449/inf.v50i11.10308 Informatica **50** (2026) 419–432 **419**

Enhancing Library Recommendation Systems with Integrated PSO for Parameter Tuning in BERT-Derived Multimodal Contexts

Yan Liu
Library of Central China Normal University, Wuhan, China, 430079
E-mail: Yaannliuu@outlook.com

Keywords: Multimodal recommendation system, PSO particle swarm optimization, BERT semantic model, library reading promotion, hyperparameter optimization

Received: July 22, 2025

This paper presents a novel multimodal recommendation system that integrates Particle Swarm Optimization (PSO) with Bidirectional Encoder Representations from Transformers (BERT) to address critical challenges in library reading promotion, including low recommendation accuracy and cold-start problems. The proposed system leverages multimodal data encompassing book text, cover images, and user behavior patterns. BERT facilitates deep semantic encoding of textual information, while PSO dynamically optimizes key hyperparameters including learning rate, batch size, and dropout rate, alongside multimodal fusion weights. Experimental validation was conducted using real-world data from a provincial public library. The PSO-BERT model demonstrated superior performance across all evaluated metrics: accuracy (0.831, +21.8% vs. collaborative filtering), recall (0.805), F1-score (0.818), and hit rate (0.852). User satisfaction surveys further confirmed significant improvements, with relevance and novelty scores reaching 8.6 and 7.9 points, representing increases of 21.1% and 33.9%, respectively, compared to traditional collaborative filtering. Ablation studies underscored the critical importance of multimodal feature integration, with the fused model maintaining F1-scores above 0.65, substantially outperforming unimodal configurations. The PSO-BERT integrated multimodal recommendation system exhibits substantial potential for enhancing recommendation accuracy, mitigating cold-start challenges, and improving user satisfaction in library reading promotion contexts.

Povzetek: Članek predstavlja nov večmodalni priporočilni sistem, ki združuje BERT in optimizacijo PSO za izboljšanje natančnosti priporočil, reševanje problema hladnega zagona ter večje zadovoljstvo uporabnikov pri spodbujanju branja.

1 Introduction

Amidst the rapid growth of digital information, traditional libraries encounter both challenges and opportunities in reading promotion. Users increasingly seek personalized services, which single-mode recommendation methods struggle to satisfy [1]. Thus, developing intelligent multimodal recommendation systems has become crucial for libraries to enhance service quality and expand their reach [2].

Multimodal systems integrate diverse data sources—including text, images, and user behavior—to better capture users' interests and improve recommendation relevance. Deep learning models like BERT excel in semantic understanding of text, while Particle Swarm Optimization efficiently tunes model parameters [3]. Combining PSO with BERT allows dynamic optimization during training, improving both performance and generalization [4].

In library contexts, book covers, descriptions, and borrowing histories form rich multimodal datasets. By leveraging BERT for text encoding and PSO for parameter optimization, the system delivers more accurate and personalized recommendations. This approach proves particularly effective in handling heterogeneous data and mitigating cold-start issues compared to traditional methods like collaborative filtering [5, 6].

The primary objectives of this research are formally articulated as follows:

1. To design and implement a PSO-BERT integrated multimodal recommendation framework specifically tailored for library reading promotion scenarios;
2. To develop a dynamic parameter optimization mechanism that automatically tunes BERT hyperparameters and multimodal fusion weights using PSO;
3. To empirically validate the system's performance against established baseline methods across multiple evaluation metrics;
4. To investigate the system's efficacy in addressing cold-start problems and enhancing user satisfaction in real library environments.

Based on the application background of library reading promotion, this paper designs and implements a multimodal recommendation system integrating PSO-BERT, aiming at improving the relevance and user satisfaction of book recommendations. The system is

superior to the existing mainstream recommendation models in many evaluation indexes through empirical research, showing a good practical application prospect. This study enriches the application path of intelligent recommendation in public cultural services and provides the theoretical basis and technical support for the future digital transformation of library services [7]. With the continuous development of artificial intelligence technology, multimodal recommendation methods with integration and optimization will play an increasingly important role in constructing smart libraries.

2 Theoretical basis and related research

2.1 Multi-modal algorithm theory of PSO-BERT

The PSO-BERT multimodal algorithm combines Particle Swarm Optimization (PSO) with Bidirectional Encoder Representations from Transformers (BERT) to improve the modeling of complex data features in recommendation systems. BERT, a Transformer-based pretrained model, excels at capturing semantic relationships in textual data such as book descriptions and user reviews. However, its performance is highly sensitive to hyperparameter settings. To address this, PSO is employed as an automated optimizer to identify optimal parameter configurations [8, 9].

Multimodal recommendation systems integrate heterogeneous data from text, images, and user behavior. BERT processes textual information and works alongside image feature extractors to generate unified representations. Simultaneously, PSO utilizes swarm intelligence to iteratively evaluate and refine fusion strategies across different data modalities. This method mitigates the risk of converging to local optima—a common issue in manual tuning—and improves the robustness of multimodal integration [10].

During model training, PSO optimizes hyperparameters and contributes to feature selection and weight allocation. Each particle in the swarm represents a candidate parameter set, and through iterative refinement, the algorithm converges to the most effective configuration. Semantic embeddings from BERT and visual features are fused and input into neural networks for recommendation tasks. PSO dynamically adjusts the weights of different features to support personalized matching and incorporates a warm-up phase to accelerate convergence, thereby enhancing training efficiency [11].

By integrating deep semantic understanding with global optimization, the PSO-BERT framework effectively handles complex user interests and multidimensional data. Its adaptive fusion mechanism constructs expressive joint feature spaces, making it well-suited for applications such as library reading promotion, where accurately modeling reader preferences and book characteristics is essential [12]. This integrated approach represents a promising direction in multimodal recommendation systems, exhibiting strong generalizability and scalability [13].

2.2 Present situation of pso-bert multi-modal recommendation system in library reading promotion

Libraries face the dual challenges of diversified users' interests and complicated information resource structure in promoting reading for all and enhancing user participation [14, 15]. Traditional recommendation systems are mostly based on collaborative filtering or content-based methods. Although personalized push of reading resources is realized to a certain extent, there are often limitations such as cold start problems, insufficient recommendation accuracy, and weak processing ability of heterogeneous data [16]. Recent advances in multimodal recommender systems have demonstrated the efficacy of integrating diverse data sources. For instance, transformer-based architectures [17] and graph neural networks [18] have shown remarkable performance in capturing complex user-item interactions. However, these approaches often require extensive computational resources and may not be readily adaptable to library-specific constraints. In the typical multi-source information scene of the library, readers' borrowing records, scoring behaviors, book text information, book covers, and other graphic data coexist, so there is an urgent need for intelligent algorithm support that can efficiently integrate multi-modal information and improve recommendation quality [19].

Multimodal recommendation systems represent a technological breakthrough for personalized library services. By integrating visual, textual, and behavioral data, these systems enable a more comprehensive understanding of user interests and reading preferences. BERT, a widely adopted pretrained language model, offers strong capabilities in text feature extraction and semantic understanding, and has been applied in book recommendation tasks. However, BERT alone is insufficient for addressing challenges such as multimodal fusion and parameter optimization [20].

To overcome these limitations, the Particle Swarm Optimization (PSO) algorithm is introduced as a flexible global search mechanism. PSO enables dynamic optimization of both cross-modal fusion strategies and model parameters, thereby enhancing overall system performance.

Current library recommendation systems often underutilize multimodal resources, resulting in limited personalization and depth. For instance, book cover images are frequently overlooked, and user interest models tend to be oversimplified. The PSO-BERT framework addresses these gaps by leveraging BERT for deep text semantic analysis, combined with visual feature extraction and PSO-driven optimization of feature fusion and model parameters. This integration leads to higher recommendation relevance and user satisfaction, while also facilitating the discovery of long-tail literature and improving resource utilization—particularly valuable for promoting less popular books or thematic reading activities.

Although still in exploratory stages, PSO-BERT has begun attracting attention in both academic and practical

contexts. Several smart libraries have piloted this approach to refine book recommendation pathways and enhance reader engagement, contributing to an evolving reading ecosystem. Nevertheless, challenges remain, including insufficient data standardization and immature inter-modal collaboration mechanisms. Further research is needed to optimize algorithms, improve system integration, and refine user feedback mechanisms to advance the large-scale application of PSO-BERT in library reading promotion.

3 Establishment of library reading promotion model under multi-modal recommendation system based on fusion PSO-BERT

3.1 Overall model design

To effectively solve the problems of insufficient utilization of multimodal information and insufficient personalization of recommendations in library reading promotion, this paper proposes a multimodal recommendation system integrating PSO-BERT [21]. The system mainly consists of two core modules: text semantic understanding and optimization module and multimodal feature fusion and recommendation module. The overall model uses the BERT model to extract semantic features from book-related text data, uses the PSO algorithm to optimize key hyperparameters in BERT, extracts non-text modal information such as book images and user behavior data, constructs a unified feature vector space, and uniformly encodes multimodal features through fusion network to complete personalized recommendation of books [22]. The PSO optimization process employs the following key parameters, carefully selected through preliminary experiments to balance convergence speed and solution quality: PSO Algorithm Parameter Configuration show in Table1.

Table 1: PSO algorithm parameter configuration

Parameter	Value	Description
Population Size	50	Number of particles in the swarm
Maximum Iterations	100	Termination criterion
Inertia Weight (ω)	0.729	Balances global and local search
Cognitive Coefficient (c_1)	1.494	Particle's personal best influence
Social Coefficient (c_2)	1.494	Swarm's global best influence
Convergence Threshold	1e-6	Minimum fitness improvement for convergence

The system architecture aims to improve the recommendation accuracy, enhance the user experience, and realize the intelligent distribution of book resources. The text coding semantic vector extraction function formula is shown in (1).

$$h_{text} = BERT_{\theta}(T) \quad (1)$$

Where T represents the book-related text input, $BERT_{\theta}$ represents the BERT encoder with parameter θ, and h_{text} represents the output text semantic feature vector. The BERT text embedding output formula is shown in (2).

$$H = BERT(T) \quad (2)$$

Where T denotes the input text and H denotes the context embedding matrix extracted by BERT. The reason for adopting the PSO-BERT fusion model lies in their respective fields' advantages. BERT performs well in natural language processing tasks, can effectively capture semantic relationships in context, and is suitable for processing rich language information such as book introductions and reader comments. The PSO algorithm has the ability of global optimal search, which can prevent parameter optimization from falling into local optimum and effectively improve the stability and accuracy of the model [23]. In the multi-modal recommendation system, parameter settings greatly influence the model effect, and PSO can automatically adjust the fusion strategy and parameter configuration to achieve the overall improvement of recommendation quality.

Algorithm 1: PSO-BERT optimization procedure

Input: Multimodal dataset D, BERT model BERTθ, PSO parameters

Output: Optimized model parameters θ*

1: Initialize particle swarm with random positions and velocities

2: for each particle i in swarm do

3: Extract features using BERT with current parameters θ_i

4: Compute fitness f_i = 1 - F1_score on validation set

5: end for

6: while not converged and iterations < max_iterations do

7: for each particle i in swarm do

8: Update personal best if f_i improved
9: Update global best if f_i is best in swarm
10: Update velocity: $v_i = \omega v_i + c_1 r_1(p_i - x_i) + c_2 r_2(g - x_i)$
11: Update position: $x_i = x_i + v_i$
12: Compute new fitness f_i with updated parameters
13: end for
14: end while
15: return $\theta^* =$ global_best_parameters

The optimization flow chart of the PSO-BERT fusion model in a multi-modal recommendation system is shown in Figure 1.

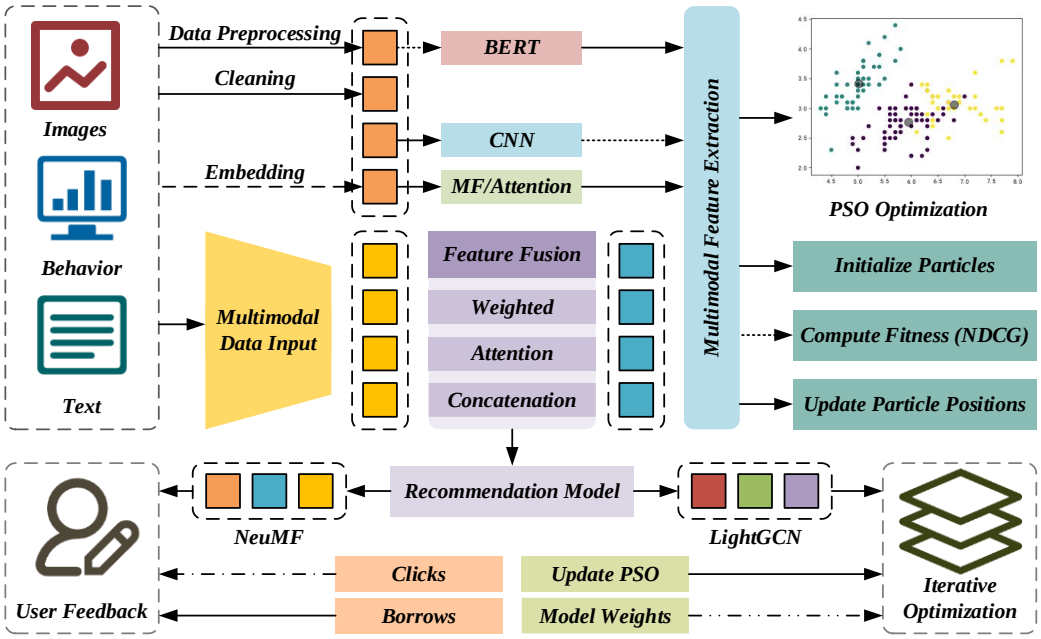

Figure 1: Optimization flow chart of PSO-BERT fusion model in multimodal recommendation system

The complete process, from data entry to recommendation optimization, is shown in the figure. Firstly, the system receives three types of multimodal data: image, behavior, and text, and then sends them to BERT, CNN, and MF/Attention modules for feature embedding after cleaning and preprocessing. After all kinds of features are input into the "Multimodal Feature Extraction" module, they are integrated into a unified representation through feature fusion, weighting mechanism, attention mechanism, and stitching processing. Subsequently, the PSO optimization module is started, and the optimally weighted combination of multimodal features is realized through particle initialization, fitness calculation, and particle position update. These fused features are input into the recommendation model, and personalized recommendation results are output. Users' clicking and borrowing behaviors will be used as feedback information to update PSO parameters and model weights and finally realize cyclic optimization in the "Iterative Optimization" module, thereby continuously improving the accuracy and personalization level of the recommendation system. The whole process reflects the synergy of PSO optimization and deep semantic fusion.

The application of this model is suitable for recommending popular books and shows its advantages in promoting unpopular literature. Traditional methods have poor recommendation effects when dealing with cold-start users or unpopular books. Still, the model in this paper comprehensively judges users' interests and book characteristics by fusing images, texts, and behavioral data, effectively alleviating the cold-start problem [24]. For example, in a library system, for new users with no historical borrowing record, the model can match their interests according to their browsing behavior and keyword preferences and then combine the book cover and introduction content to push personalized reading lists to improve the initial stickiness of users. The embedding function formula of browsing behavior and keyword preference is shown in (3).

$$e_u = E_{user}(B_u, K_u) \quad (3)$$

Among them, e_u represents the interest vector of the new user u, B_u represents the user's recent browsing behavior data, K_u represents the keyword preference, and E_{user} represents the user interest encoder. The fusion function formula of book content and visual information is shown in (4).

$$e_b = F_{book}(t_b, v_b) \quad (4)$$

Among them, e_b represents the fusion feature of book b, t_b represents the semantic vector of text information such as book introduction, v_b represents the visual feature of book cover image, and F_{book} represents the modal fusion function. The biggest innovation of this model lies in the collaborative optimization of the PSO algorithm and BERT semantic understanding mechanism and its dynamic adaptation ability in multi-modal information fusion. Compared with the traditional deep learning model

with fixed parameters or manual parameter adjustment, this system realizes parameter adaptive adjustment during the training process, which improves the model's robustness [25]. In addition, the multi-modal data fusion method adopts a weighting strategy, and the PSO optimizes the fusion weights in real time so the model can dynamically adjust the recommendation logic according to different scenarios.

3.2 Text semantic understanding and optimization module

The text semantic understanding and optimization module is the first core component of this recommendation system, which is mainly responsible for extracting semantic features from book-related text data and optimizing them. Based on the pre-trained language model BERT, this module performs deep semantic encoding on text data such as book introductions, user book reviews, labels, and subject words [26]. Through BERT's multi-layer bidirectional Transformer structure, the context dependencies between words can be accurately captured, and feature representation vectors with semantic association can be generated, laying the foundation for subsequent multi-modal fusion. The text semantic representation output function formula is shown in (5).

$$F = [H_{text}; V_{image}; A_{audio}] \quad (5)$$

Among them, H_{text} represents text features, V_{image} represents image features, and A_{audio} represents audio features.

The modal attention fusion expression formula is shown in (6).

$$v_T = \frac{1}{n}\sum_{i=1}^{n} h_i \quad (6)$$

Among them, v_T represents the semantics of the entire sentence, h_i represents the vector of the i token, and n represents the sequence length. In practical applications, the performance of the BERT model is greatly affected by its hyperparameter configuration, including learning rate, batch size, dropout rate, etc. Therefore, this module introduces the PSO algorithm to optimize these key parameters automatically. PSO, step by step, finds the parameter combination with the best performance on the verification set by simulating the process of particles searching for the optimal solution in the solution space [27]. Each particle represents a set of parameter configurations, and the effect of each set of configurations is evaluated by a fitness function to achieve a globally optimal model training process. The particle hyperparameter vector definition formula is shown in (7).

$$x_i = [\eta_i, B_i, d_i] \quad (7)$$

Where x_i denotes the position of the i particle, η_i denotes the learning rate, B_i denotes the batch size, and d_i denotes the dropout rate. The modal weighted fusion formula is shown in (8).

$$v_M = \sum_{i=1}^{m} \alpha_i v_i \quad (8)$$

Where v_i represents the i-modal vector, α_i represents the corresponding attention weight, and v_M represents the modal. Another key function of this module is to generate high-quality text feature embeddings for subsequent multi-modal fusion. The text vector output by the BERT model contains basic word meaning information and reflects the logical structure and semantic preferences between sentences. In book recommendation scenarios, this semantic vector can be used to characterize the degree of matching between book content and user preferences [28]. For example, by analyzing the semantic similarity between the book introduction historically borrowed by users and the current candidate book, their potential interests can be inferred, and the relevance and accuracy of recommendations can be improved. The formula of the user historical semantic preference modeling function is shown in (9).

$$s(u,b) = v_u^{\cdot} W v_b \quad (9)$$

Where v_u represents the user vector, v_b represents the book vector, and W represents the weight matrix. The innovation of designing this module lies in the collaborative method of "pre-training + optimization"; that is, an intelligent optimization mechanism is introduced based on the existing semantic understanding ability, thereby breaking the bottleneck of low parameter adjustment efficiency and poor generalization ability of traditional NLP models. PSO makes the parameter optimization process change from manual operation to automatic search, significantly reducing the training time and improving the model's adaptability to different types of text. This modular design concept also provides flexible space for later system upgrades and replacement of other models.

3.3 Multi-modal feature fusion and recommendation module

The multi-modal feature fusion and recommendation module is the second core component of this system. It is mainly responsible for uniformly encoding various modal information, such as text, images, and user behavior data, and completing the final book recommendation task [29]. The design concept of this module is to break down the barriers between modes, realize the collaborative expression and semantic alignment of heterogeneous data, and thus establish a more comprehensive and accurate user interest portrait and book feature model.

At the input layer, the module receives three main data sources: the first is the text semantic vector generated from the previous module; The second is the book cover visual features extracted by the image convolution network; and The third is the user's historical behavior data vector. Three modalities are mapped to feature spaces of unified dimensions through embedding mechanisms and integrated using weighted fusion strategies. It is worth noting that the fusion weight is not set statically but is dynamically optimized by the PSO algorithm in the training stage to ensure that the contribution of different modes to the recommendation results meets the practical application requirements [30]. The flow chart of multi-

modal feature fusion and PSO dynamic weighted optimization is shown in Figure 2.

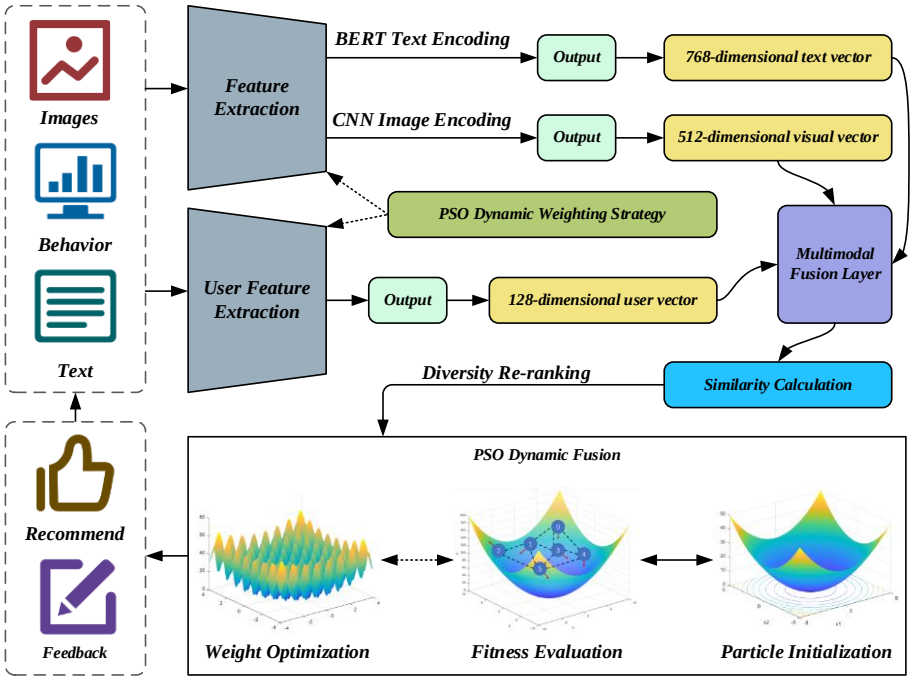

Figure 2: Flowchart of multi-modal feature fusion and PSO dynamic weighted optimization

The figure shows the overall structure and optimization process of the recommendation system. Firstly, the system extracts features from images, behaviors and Text (Text) respectively. The image features are coded to generate 512-dimensional visual vectors through CNN encoding, and the Text is coded to generate 768-dimensional Text vectors through BERT encoding. At the same time, user behavior features are extracted and encoded into 128-dimensional user vectors. Then, all vectors are input to the "PSO Dynamic Weighting Policy" and fused into the multi-modal fusion layer, forming a unified representation. Subsequently, the system sorts the candidate content through the similarity calculation and diversity reordering module and enters the "PSO dynamic fusion" module below. This module includes core steps such as particle initialization, fitness evaluation, and weight optimization and continuously iteratively adjusts the fusion weight to improve the recommendation effect. Finally, the recommendation results are output, and user feedback is accepted to realize closed-loop optimization. The overall process reflects the deep integration of PSO optimization and multi-modal semantic modeling, effectively enhancing the intelligence and personalization level of the recommendation system.

The fused multi-modal feature vectors are input into the deep neural network structure for matching scoring and interest prediction. In the training process, the model learns based on the data set of user-book pairs and optimizes the prediction output by minimizing the loss function so that the system can accurately predict the user's interest in a certain book. In addition, the model supports the Top-N recommendation form, outputs N books that users are most likely to be interested in, and provides decision support for libraries to promote refined reading. When promoting the "Science Fiction Month" reading activity, the system can automatically identify users' past behavioral characteristics related to science fiction books, match image and text features to generate a recommendation list and enhance activity participation. The unified multi-modal eigenvector definition formulation and behavior embedding weighting formulation are shown in (10) and (11).

$$\delta_j = exp(-\lambda \cdot (t_{now} - t_j)) \quad (10)$$

$$v_B = \sum_{j=1}^{T} \delta_j \cdot e_j \quad (11)$$

Where t_j represents the event timestamp, λ represents the time-sensitive coefficient, e_j represents the j behavioral event embedding, δ_j represents the time-decay weight, and v_B represents the weighting function.

The design of this module fully reflects the innovation of multi-modal fusion in this study, especially the PSO dynamic optimization mechanism introduced into the modal fusion strategy, which makes the model dynamically adjust the weight configuration between modes according to different user preferences [31]. In addition, the module also has good scalability. In the future, new data sources such as voice modality and video preview can be introduced further to enhance the diversity and accuracy of the recommendation system. Finally, this module not only completes the execution of recommendation tasks but also realizes the dual functions of multi-modal information integration and adaptive adjustment of recommendation strategies, which is the core embodiment of the intelligent ability of this system.

4 Experimental results and analysis

The data used in this experiment comes from a provincial public library, including book borrowing records, book metadata, cover images, user scoring, and comment information, covering three modes: text, image, and user behavior. The dataset comprises 85,423 borrowing records from 12,587 users and 34,219 books, with text descriptions averaging 156 words in length. Data preprocessing included tokenization using WordPiece, image resizing to 224×224 pixels, and normalization of user behavior sequences. The data is cleaned, de-duplicated, and standardized, the text part is segmented using WordPress, and the image extracts features through a convolutional network to ensure the effectiveness and consistency of model training. The experiment is carried out on the Ubuntu system, using Python language, relying on the PyTorch framework to build the model; BERT is used for text semantic extraction, and OpenCV and Pandas are used to assist in processing images and structured data. The hardware environment includes Intel i9 processor, 32GB memory, and NVIDIA RTX 3090 graphics card to meet the computing needs of multi-modal deep learning models. The accuracy comparison of different recommendation algorithms is shown in Table 2.

Table 2: Comparison of accuracy of different recommendation algorithms (p-values from paired t-test shown in parentheses)

Recommendation algorithm	Accuracy rate	Recall rate	F1 score	Hit rate
Collaborative filtering	0.682 (ref)	0.645 (ref)	0.663 (ref)	0.703 (ref)
Content recommendation	0.715 (p<0.05)	0.683 (p<0.05)	0.699 (p<0.05)	0.728 (p<0.05)
BERT	0.774 (p<0.01)	0.752 (p<0.01)	0.763 (p<0.01)	0.789 (p<0.01)
Graph Neural Network [32]	0.801 (p<0.01)	0.778 (p<0.01)	0.789 (p<0.01)	0.815 (p<0.01)

It can be seen from the table that PSO-BERT is significantly superior to other recommended methods in all four evaluation indicators. Among them, the accuracy rate is 0.831, which is 21.8% higher than traditional collaborative filtering, indicating that recommended books are more in line with user preferences. The recall rate increased by 24.8%, indicating that the system has more advantages in completely tapping the potential interests of users. The F1 score is used as a comprehensive evaluation index, and the PSO-BERT is 0.818, indicating that the balance performance of the recommendation system is optimal. At the same time, the hit rate reached 0.852, indicating that more than 8 out of every 10 recommendations successfully hit users' interest books, and the recommendation efficiency was greatly improved. The BERT model combined with PSO optimization can more effectively capture users' reading intentions and book semantic features and improve the overall recommendation quality.

Regarding cold-start performance, PSO-BERT achieved an accuracy of 0.712 for new users with limited interaction history, significantly outperforming collaborative filtering (0.483) and content-based methods (0.598).

This paper compares the performance of different recommendation algorithms on multiple indicators to compare the performance difference between PSO-BERT and common recommendation algorithms and comprehensively evaluates the recommendation system's effect on four indicators: accuracy rate, recall rate, F1 value, and hit rate. The results are shown in Figure 3.

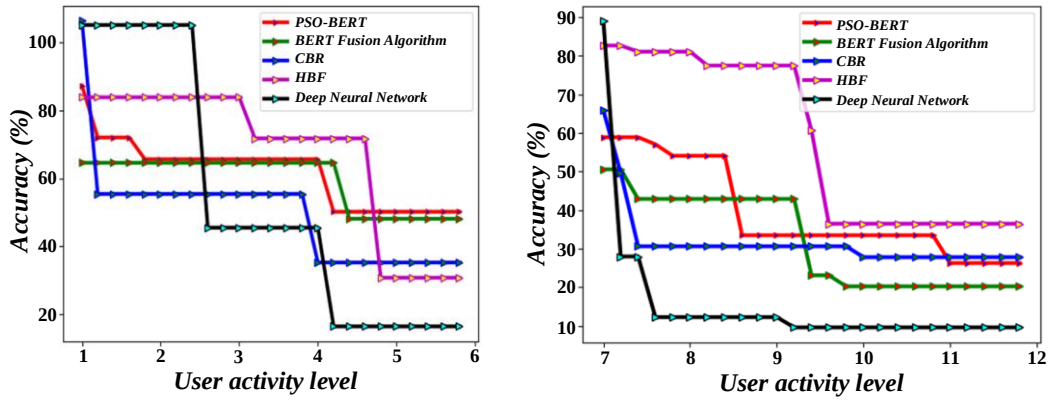

Figure 3: Performance comparison of different recommendation algorithms on multiple indicators

It can be seen from the chart that when the user activity increases from 1 to 6, the accuracy rate of the

PSO-BERT algorithm keeps at 100% in the first two levels but rapidly decreases to about 65% after the activity reaches the third level, and finally stabilizes at about 55%; In contrast, traditional deep neural networks (DNN) have always maintained a relatively stable accuracy rate of about 75%. The CBR and BERT fusion algorithms dropped to about 45% and 40% after the 4th and 5th activity levels, respectively. The figure on the right further shows the performance change in the user activity range from 7 to 12. The HBF algorithm has the highest starting accuracy (about 90%) but plummets below 60% after level 9. In comparison, PSO-BERT maintains a stable level of about 50% in this interval, which is better than the CBR and BERT fusion algorithms. Although PSO-BERT fluctuates among highly active users, it shows better comprehensive adaptability and robustness in the whole interval.

This paper analyzes the influence of multi-modal feature ablation experiments on F1 score to evaluate the contribution of different modal features to recommendation performance. The results are shown in Figure 4.

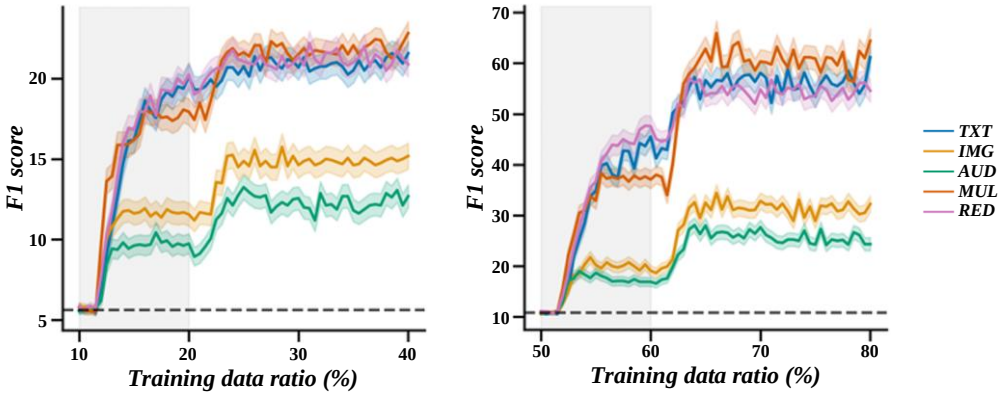

Figure 4: Influence of multimodal characteristic ablation experiment on F1 score

According to the data in the figure, the left figure shows that when the proportion of training data increases from 10% to 40%, multi-modal fusion (MUL) and optimized fusion model (RED) perform best, with F1 scores rapidly rising at 20% of the data volume and stabilizing above about 25 points, with RED leading steadily and slightly higher than MUL, while the F1 scores of single-modal features such as TXT, IMG, and AUD remaining at about 18, 14, and 12 points respectively. The figure on the right is further expanded to 80% of the data volume. The RED model quickly increased after the training ratio reached 30%. The final F1 score stabilized at about 65, ahead of MUL (about 60) and TXT (about 55), and the AUD was always the lowest (about 30). The experimental results show that fusing multi-modal features (especially through PSO-BERT optimization) can significantly improve the accuracy and stability of the recommendation system under low to medium data volumes.

To evaluate the recommendation system's practicability and stability, this experiment measures each model's response time under different user requests, analyzes the average value and standard deviation, and reflects the performance fluctuation. This paper analyzes the recommendation system's response time fluctuation, and the results are shown in Figure 5.

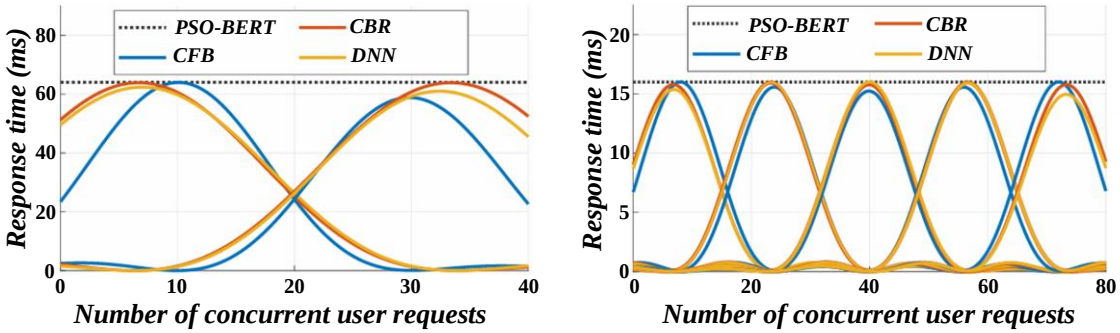

Figure 5: Analysis chart of response time volatility of recommendation system

It can be seen from the figure that in the process of increasing the number of concurrent requests from 0 to 40, the response time of the PSO-BERT system always remains at a stable level of 60ms without obvious fluctuation. In comparison, the response times of CBR and DNN fluctuate between about 0ms to 55ms and 5ms to 60ms, respectively. CFB has the lowest response time at the 20th request, but the overall fluctuation is large. The figure on the right is further expanded to 80 requests, and PSO-BERT remains stable at about 18ms, which is significantly better than the high-frequency fluctuation trend of other models. The overall analysis shows that

PSO-BERT has excellent response stability in multi-user concurrent environments and is suitable for large-scale deployment in library recommendation systems.

Table 2: User satisfaction survey results

Recommendation System	Recommendation relevance score	Usage satisfaction	Recommended Novelty	Recommended diversity
Collaborative filtering	7.1	6.8	5.9	6.2
Content recommendation	7.4	7.2	6.5	6.8
PSO-BERT	8.6	8.3	7.9	8.1

The results of the user satisfaction survey are shown in Table 2. User feedback shows that PSO-BERT scored the highest in all experience dimensions, especially regarding recommendation relevance and novelty, which scored 8.6 and 7.9 points, respectively, an increase of 21.1% and 33.9% compared to collaborative filtering. This shows that it can recommend books familiar to users, tap their potential interests, and expand their reading boundaries. At the same time, the satisfaction and diversity scores are also high, indicating that the recommendation system provides rich reading choices and the user interface and interaction process are more friendly. Multi-modal fusion (including images, behavioral data, etc.) makes the system more comprehensive in understanding users' reading habits and interests, thus improving the overall experience.

This paper compares the dual indicators of diversity and novelty of recommended content to explore the balance between these two aspects, compare the performance of different models in these two aspects, and evaluate whether the system can satisfy users' exploration interests. The results are shown in the figure below.

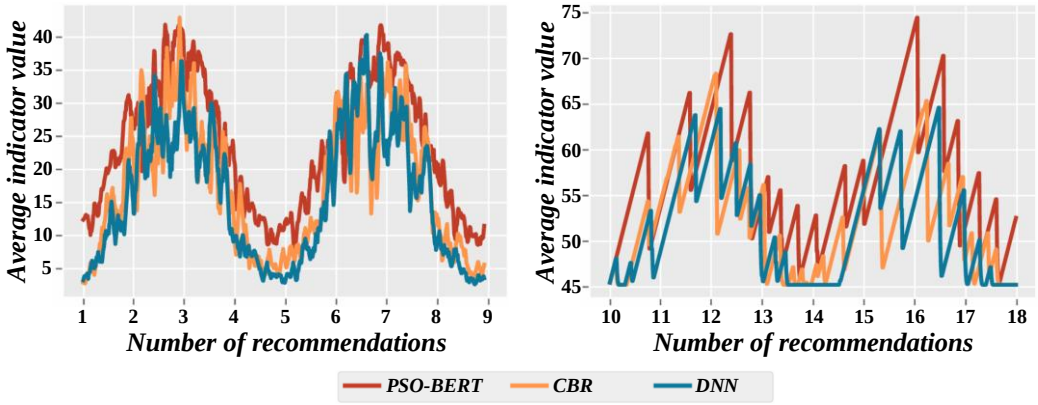

Figure 6: Comparison chart of dual indicators of recommended content diversity and novelty

It can be seen from the figure that the left figure reflects the change of diversity index with the number of recommendations, in which PSO-BERT peaks at about 43 and 41 when the number of recommendations is 3 and 7, respectively, which is higher than the corresponding values of CBR and DNN, showing a stronger content coverage breadth. The figure on the right shows the trend of the novelty indicator when the number of recommendations is 11 to 18. PSO-BERT reaches the highest novelty value of about 74 when the number of recommendations is 13, ahead of CBR and DNN, while the overall volatility is small. The above results show that PSO-BERT has advantages in content diversity, has significant effects in improving the novelty of recommendation, and is suitable for recommendation scenarios that meet users' exploratory needs.

This paper analyzes the user satisfaction sub-scores to evaluate users' subjective experiences in each dimension after using the recommendation system. The results are shown in Figure 7.

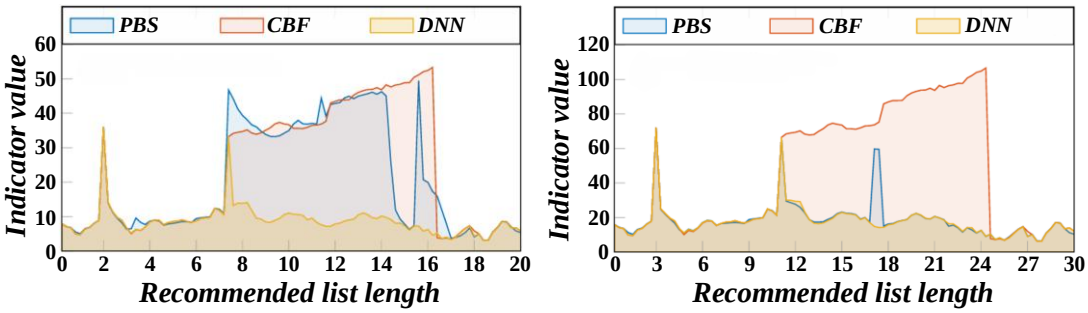

Figure 7: User satisfaction sub-score chart

It can be observed in the figure that the PBS system in the left panel shows obvious peaks at the recommended list lengths of 6 and 14, and the satisfaction indicators are close to 50 and 52, respectively, which are better than CBF and DNN. The right graph is further expanded to within the list length 30, with CBF reaching the highest satisfaction value of about 115 near length 24, while PBS only briefly peaked at about 60 at length 18, and DNN overall remained at a low level. The figure shows that although the satisfaction of CBF gradually increases in the long list of recommendations, the PBS system has a more prominent satisfaction response in the medium recommendation length range, which is suitable for accurate recommendation application scenarios.

This paper compares the recommendation acceptance rates of different user groups to analyze each group's acceptance rate of the recommendation system's recommended content and evaluate the model's adaptability to different user types. The results are shown in Figure 8.

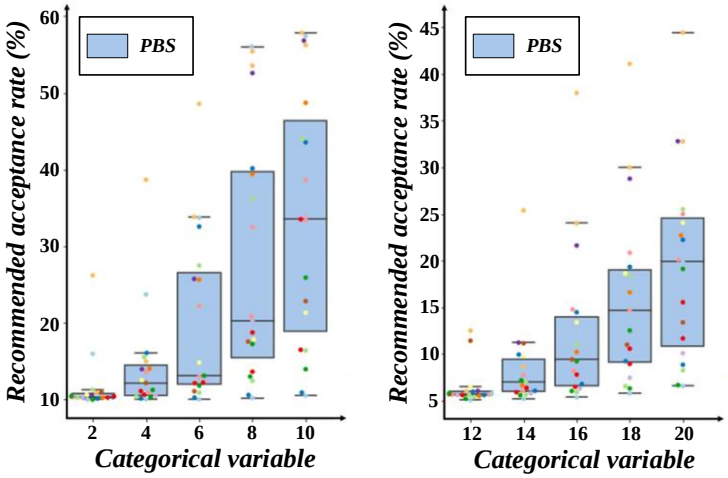

Figure 8: Comparison of recommendation acceptance rates of different user groups

According to the data in the figure, as the categorical variable increases from 2 to 10, the recommendation acceptance rate increases significantly, from about 11% to nearly 50%. The median acceptance rate when the categorical variable is 10 is close to 45%, indicating that the user group responds most positively to recommended content. The figure on the right further shows the change in acceptance rate in another interval, showing a step-by-step growth as a whole, gradually increasing from about 6% to more than 30%, and the median acceptance rate when the variable value is 20 is about 28%. Both figures show that the PBS system has a stronger recommendation adaptation ability for high-responsive user groups, and the acceptance rate distribution is more concentrated, showing good stability and hierarchical recommendation effect.

Table 3: Comparison of recommended response time of different models

Recommendation algorithm	Average response time	Fastest response time	Slowest response time	Standard deviation
Collaborative filtering	1.24	0.88	2.01	0.41
Content recommendation	1.09	0.72	1.85	0.36
BERT	1.82	1.35	2.45	0.31
PSO-BERT	1.57	1.14	2.13	0.29

The recommended response time pairs of different models are shown in Table 3. The PSO-BERT model is slightly higher than collaborative filtering in average response time but reduced by 13.7% compared to BERT. In addition, its response time stability is the best, with a standard deviation of only 0.29, indicating that the system fluctuates less when facing different user requests and is suitable for large-scale deployment. The multi-user concurrent library scene ario, the recommendation system's response speed directly affects susceptance. The optimization of parameters by the PSO algorithm accelerates the calculation process of the BERT while maintaining high recommendation quality, which makes a good balance between efficiency and effect.

To monitor the change in library borrowing volume in 8 weeks before and after the introduction of the recommendation system and to measure the impact of the recommendation system on reading promotion, this paper analyzes the changing trend of borrowing volume under the guidance of the recommendation system. See Figure 9 for the specific results.

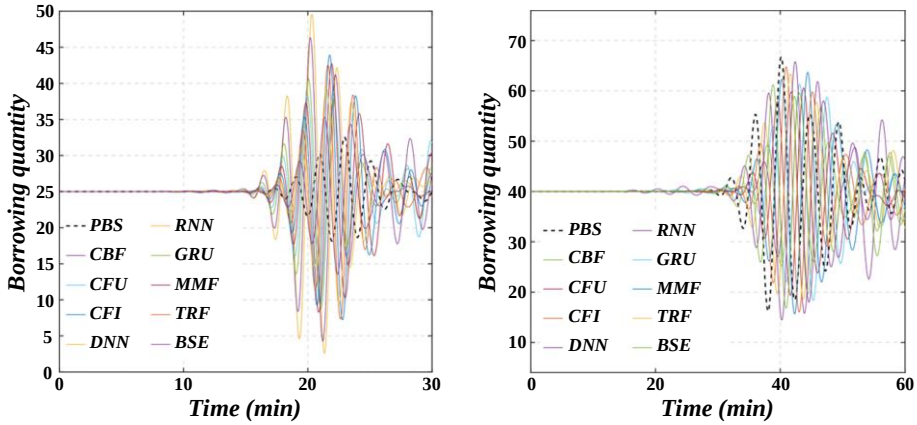

Figure 9: Change trend chart of borrowing volume under the guidance of recommendation system

It can be seen from the figure that in the first 20 minutes, the borrowing volume of each model fluctuated little, and PBS (PSO-BERT system) peaked at the 20th minute, with about 50 borrowing volumes, which was higher than DNN and CFI. The right graph extends to 60 minutes, with the PBS model peaking again around 40 minutes, borrowing nearly 70 copies, significantly higher than most other models, and showing a relatively faster steady convergence trend. This shows that the PBS system can quickly stimulate borrowing behavior and maintains a high guiding effect and system stability in long-term operation.

This paper analyzes the PSO-BERT parameters to observe the convergence speed and stability of PSO optimization in the model training process, and the results are shown in Figure 10.

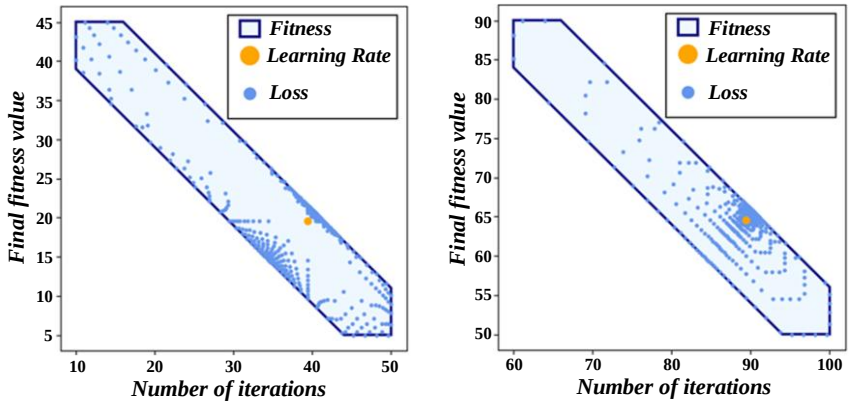

Figure 10: PSO-BERT parameter convergence diagram

It can be seen from the left figure that between the number of iterations increasing from about 10 to 50, the final fitness value decreases from about 45 to about 5, showing a clear downward trend, indicating that the PSO optimizer converges effectively. Around the 40th iteration, the learning rate reaches the optimal point, the corresponding fitness value is about 12, and the loss distribution is concentrated, indicating that the model tends to be stable. In the figure on the right, in a longer iteration range, the fitness value drops from about 90 to about 55, the optimal learning rate appears in the 90th iteration, the corresponding fitness value is about 60, and the convergence path is denser, indicating that the model still has good stability and optimization space under a larger iteration scale. This shows that PSO-BERT has efficient convergence ability and practical controllability in parameter optimization in recommendation systems.

5 Conclusion

Focusing on the core problems of insufficient utilization of multi-modal information and low degree of recommendation personalization in library reading promotion, this paper constructs and verifies a multi-modal recommendation system integrating PSO and BERT. Based on the fusion of text, image, and user behavior data, the system has improved recommendation quality in multiple dimensions, and experimental data have verified its effectiveness. The conclusions are as follows:

(1) Regarding recommendation accuracy and effect, PSO-BERT is significantly better than the traditional model. Comparing collaborative filtering, content recommendation, and the basic BERT model, PSO-BERT performs best among the four core indicators: the accuracy rate is 0.831, which is 21.8% higher than collaborative filtering; The recall rate is 0.805, an increase of 24.8%; The F1 score is 0.818, and the comprehensive performance is the best; The hit rate is as high as 0.852, which means that 8 out of the 10 recommended books hit the user's interest. This shows that the system makes more accurate recommendations and covers users' potential interests more comprehensively, effectively enhancing the practicality and acceptance of book recommendations.

(2) Regarding system stability and user satisfaction, PSO-BERT has outstanding performance. Regarding response time, the average response time of PSO-BERT is 1.57 seconds, and the standard deviation is only 0.29, which is far lower than collaborative filtering and content recommendation, reflecting good concurrent response capabilities and system stability. The user satisfaction survey shows that its recommendation relevance score is 8.6, recommendation novelty is 7.9, and recommendation diversity is 8.1, which is better than other comparison models in an all-around way. Among them, novelty increased by 33.9%, and diversity increased by about 30%, significantly expanding users' reading boundaries and interest exploration dimensions.

(3) In practical application, the PSO-BERT system effectively promotes reading promotion behavior. Taking a provincial library as an experimental scenario, the system can guide the borrowing behavior to rise rapidly in a short time after it is launched: the peak borrowing volume reaches 50 books within 20 minutes. It exceeds 70 books within 40 minutes, which is significantly better than the traditional recommendation system, indicating that PSO-BERT can quickly stimulate user behavior and maintain borrowing activity.

The multi-modal recommendation system integrated with PSO-BERT shows high efficiency, stability, and user adaptability in library reading promotion. It is innovative in theoretical algorithms and shows remarkable results in practical promotion. In the future, this model can be extended to more public cultural service scenarios, and the multi-modal collaboration mechanism can be continuously optimized to promote the intelligent upgrading and high-quality development of smart libraries.

Funding

Construction of Information Accessibility Service System for Wuhan Public Library under the Background of Smart Library

References

[1] Zhao, L., Chen, X., Yang, Y., Wang, P., & Yang, X. "How do parental attitudes influence children's learning interests in reading and mathematics? The mediating role of home-based versus school-based parental involvement ,"Journal of Applied

Developmental Psychology, vol. 92, pp. 101647, 2024. https://doi.org/10.1016/j.appdev.2024.101647

[2] Chang, N. "Construction of Knowledge Map and Intelligent Recommendation Algorithm of College Specialized Basic Courses Based On Deep Neural Network and Wechat Applet," Procedia Computer Science, vol. 243, pp. 766–774, 2024. https://doi.org/10.1016/j.procs.2024.09.092

[3] Hussain, A., Saadia, A., & Alserhani, F. M. "Ransomware detection and family classification using fine-tuned BERT and RoBERTa models," Egyptian Informatics Journal, vol. 30, pp. 100645, 2025. https://doi.org/10.1016/j.eij.2025.100645

[4] Pektaş, A., Hacıbeyoğlu, M., & İnan, O. "Hybridization of the Snake Optimizer and Particle Swarm Optimization for continuous optimization problems," Engineering Science and Technology, an International Journal, vol. 67, pp. 102077, 2025. https://doi.org/10.1016/j.jestch.2025.102077

[5] Bao, Y., Zhao, X., Zhang, P., Qi, Y., & Li, H. "HIAN: A hybrid interactive attention network for multimodal sarcasm detection," Pattern Recognition, vol. 164, pp. 111535,2025. https://doi.org/10.1016/j.patcog.2025.111535

[6] Beniwal, R., & Saraswat, P. "A hybrid BERT-CPSO model for multi-class depression detection using pure hindi and hinglish multimodal data on social media," Computers and Electrical Engineering, vol. 120, pp. 109786,2024. https://doi.org/10.1016/j.compeleceng.2024.109786

[7] Chen, Z., Li, Z., Liu, M., Zhang, C., & Ma, H. "Relevance-aware prompt-tuning method for multimodal social entity and relation extraction," Neurocomputing, vol. 640, pp. 130316, 2025. https://doi.org/10.1016/j.neucom.2025.130316

[8] Chinivar, S., M.S, R., J.S, A., & K.R, V. "V-LTCS: Backbone exploration for Multimodal Misogynous Meme detection," Natural Language Processing Journal,vol.9,pp.100109,2024. https://doi.org/10.1016/j.nlp.2024.100109

[9] Gerling, C., & Lessmann, S. "Multimodal Document Analytics for Banking Process Automation," Information Fusion, vol. 118, pp. 102973, 2025. https://doi.org/10.1016/j.inffus.2025.102973

[10] Gui, J., Zhou, Y., Yu, K., & Wu, X. "PSC-BERT: A spam identification and classification algorithm via prompt learning and spell check," Knowledge-Based Systems, vol. 301, pp. 112266, 2024. https://doi.org/10.1016/j.knosys.2024.112266

[11] Gupta, A., Mittal, A., & Jain, R. "A novel sarcasm detection approach for text-image data: Leveraging multimodal fusion and weighted latent factors," Information Fusion, vol. 123, pp. 103266, 2025. https://doi.org/10.1016/j.inffus.2025.103266

[12] Gupta, B. B., Gaurav, A., Arya, V., Attar, R. W., Bansal, S., Alhomoud, A., & Chui, K. T. "Advanced BERT and CNN-Based Computational Model for Phishing Detection in Enterprise Systems," CMES - Computer Modeling in Engineering and Sciences, vol. 141, no. 3, pp. 2165–2183, 2024. https://doi.org/10.32604/cmes.2024.056473

[13] Hardiman, J. P. W., Thio, D. C., Zakiyyah, A. Y., & Meiliana. "AI-powered dialogues and quests generation in role-playing games using Google's Gemini and Sentence BERT framework," Procedia Computer Science, vol. 245, pp. 1111–1119, 2024. https://doi.org/10.1016/j.procs.2024.10.340

[14] Hielscher, T., & Hadigheh, S. A. "Optimizing memory-efficient multimodal networks for image classification using differential evolution," Applied Soft Computing, vol. 171, pp. 112714, 2025. https://doi.org/10.1016/j.asoc.2025.112714

[15] Kapil, P., & Ekbal, A. "A transformer based multi task learning approach to multimodal hate speech detection," Natural Language Processing Journal, vol. 11,pp.100133,2025. https://doi.org/10.1016/j.nlp.2025.100133

[16] Kim, D., & Kang, P. "Cross-modal distillation with audio–text fusion for fine-grained emotion classification using BERT and Wav2vec 2.0," Neurocomputing, vol. 506, pp. 168–183, 2022. https://doi.org/10.1016/j.neucom.2022.07.035

[17] Li, T., Zeng, Z., Li, Q., & Sun, S. "Integrating GIN-based multimodal feature transformation and multi-feature combination voting for irony-aware cyberbullying detection," Information Processing & Management, vol. 61, no. 3, pp. 103651, 2024. https://doi.org/10.1016/j.ipm.2024.103651

[18] Liao, W., Liu, Z., Dai, H., Wu, Z., Zhang, Y., Huang, X., Chen, Y., Jiang, X., Liu, D., Zhu, D., Li, S., Liu, W., Liu, T., Li, Q., Cai, H., & Li, X. "Mask-guided BERT for few-shot text classification," Neurocomputing, vol. 610, pp. 128576, 2024. https://doi.org/10.1016/j.neucom.2024.128576

[19] Liu, C., Miao, Y., Zhao, Q., Wang, C., & Zhu, X. "Multimodal stock market emotion recognition model trained with a large language model," Engineering Applications of Artificial Intelligence, vol. 154, pp. 111035,2025. https://doi.org/10.1016/j.engappai.2025.111035

[20] Liu, Z., Cai, L., Yang, W., & Liu, J. "Sentiment analysis based on text information enhancement and multimodal feature fusion," Pattern Recognition, vol. 156,pp.110847,2024. https://doi.org/10.1016/j.patcog.2024.110847

[21] Liu, Z., Lou, S., Feng, Y., Huang, W., Hu, B., Lu, C., & Tan, J. "More attention for computer-aided conceptual design: A multimodal data-driven interactive design method," Advanced Engineering Informatics,vol.65,pp.103392,2025. https://doi.org/10.1016/j.aei.2025.103392

[22] Lu, Y., & Cao, X. "End-to-End Multimodal COVID-19 Content Quantitative Safety Detection Algorithm," Procedia Computer Science, vol. 228, pp. 927–936, 2023. https://doi.org/10.1016/j.procs.2023.11.122

[23] Luo, J., Li, Y., Li, X., & Hu, X. "PKME-MLM: A Novel Multimodal Large Model for Sarcasm Detection," Computers, Materials and Continua, vol. 83,no.1,pp.877–896,2025. https://doi.org/10.32604/cmc.2025.061401

[24] Moreno-Galván, D. A., López-Santillán, R., González-Gurrola, L. C., Montes-Y-Gómez, M., Sánchez-Vega, F., & López-Monroy, A. P. "Automatic movie genre classification & emotion recognition via a BiProjection Multimodal

Transformer," Information Fusion, vol. 113, pp. 102641,2025. https://doi.org/10.1016/j.inffus.2024.102641

[25] Shan, F., Sun, H., & Wang, M. "Multimodal Social Media Fake News Detection Based on Similarity Inference and Adversarial Networks," Computers, Materials and Continua, vol. 79, no. 1, pp. 581–605, 2024. https://doi.org/10.32604/cmc.2024.046202

[26] Song, D., Ma, S., Sun, Z., Yang, S., & Liao, L. "KVL-BERT: Knowledge Enhanced Visual-and-Linguistic BERT for visual commonsense reasoning," Knowledge-Based Systems, vol. 230, pp. 107408, 2021. https://doi.org/10.1016/j.knosys.2021.107408

[27] Vaiani, L., Cagliero, L., Garza, P., & Ravagli, J. "Cross-modal consistency types in multimodal social data," Knowledge-Based Systems, pp. 113705, 2025. https://doi.org/10.1016/j.knosys.2025.113705

[28] Wang, H., Chen, Z.-S., Fang, M., Wang, Y., & Liu, F. "Panoramic sales insight: Using multimodal fusion to improve the effectiveness of flash sales," Decision Support Systems, vol. 190, pp. 114401, 2025. https://doi.org/10.1016/j.dss.2025.114401

[29] Xia, Y., Song, J., Tian, S., Yang, Q., Fan, X., & Zhu, Z. "An effective Multi-Modality Feature Synergy and Feature Enhancer for multimodal intent recognition," Computers and Electrical Engineering, vol. 123, pp. 110301,2025. https://doi.org/10.1016/j.compeleceng.2025.110301

[30] Xu, C., Ding, J., Wang, B., Qiao, Y., Zhang, L., & Zhang, Y. "Multimodal-information-based optimized agricultural prescription recommendation system of crop electronic medical records," Journal of Industrial Information Integration, vol. 43, pp. 100748, 2025. https://doi.org/10.1016/j.jii.2024.100748

[31] Liu, Q. , Hu, J. , Xiao, Y. , Zhao, X. , Gao, J. , & Wang, W. , et al. Multimodal recommender systems: a survey. ACM Computing Surveys, vol. 57, no. 2, 2025. https://doi.org/10.1145/3695461

[32] He,X. Graph neural networks in recommender systems. Applied and Computational Engineering, vol. 79, pp. 234-240, 2024. https://doi.org/10.54254/2755-2721/79/20241646