
Digital Media Visual Recommendation System Based on Artificial Neural Network Machine Learning

Peng Jing

*Hunan Mass Media Vocational and Technical College, School of Visual Arts;
Changsha City, Hunan Province; Postal Code: 410100; China
E-mail: peng_jing78@outlook.com*

Received 07 March 2025; Accepted 11 August 2025

Abstract

Conventional collaborative filtering methods struggle to understand the subtle connections between users and evolving short video content, often resulting in imprecise or broad recommendations. Additionally, the absence of interpretability, usually displayed as unchanging lists or tables, restricts user confidence and involvement. To overcome these challenges, this research suggests a tailored short video recommendation system that integrates a CNN-BiLSTM hybrid model with a Multi-Head Attention (MHA) approach and visualization techniques. CNN is utilized to obtain visual characteristics from brief videos, BiLSTM identifies temporal relationships in video sequences and user actions, while MHA improves feature weighting for tailored significance. To address the transparency concern, the system incorporates real-time visualization methods like heat maps and interactive charts, enabling users to grasp the reasoning behind each suggestion. Experimental findings from the

Journal of Mobile Multimedia, Vol. 21_5, 967–996.

doi: 10.13052/jmm1550-4646.2156

©2025 River Publishers

MicroLens dataset indicate that the proposed model achieves a hit rate of 0.94 at $k = 15$, outperforming conventional methods such as ItemCF by 0.16. This method greatly enhances the accuracy of recommendations, transparency, and user engagement in digital media contexts

Keywords: Recommendation system, deep learning, CNN-BiLSTM, multi-head attention, visualization, short video.

1 Introduction

The explosive growth of digital media, especially the rise of short video platforms such as TikTok, Kuaishou, and Instagram Reels, has led to an unprecedented surge in multimedia content. This rapid expansion has resulted in a pervasive issue of information overload, where users struggle to efficiently discover relevant content amidst vast streams of videos [1, 2]. Navigating such content-heavy environments without intelligent filtering can lead to user fatigue, reduced engagement, and ineffective content discovery. Recommendation systems play a crucial role in mitigating this challenge by delivering personalised suggestions that enhance user experience and retention [3, 4]. Among these, collaborative filtering (CF) has been a widely adopted technique due to its simplicity and effectiveness. However, traditional CF methods, including both user-based and item-based approaches, exhibit notable limitations. They rely heavily on historical user behavior or item similarity, which often leads to overly generalized or repetitive recommendations. These systems struggle to capture the rich semantic and temporal characteristics of short videos. They are typically presented in static formats, such as lists or tables, without offering insight into why the recommendations were made. This lack of interpretability weakens user trust and limits engagement.

With the explosive rise of mobile-based digital media platforms such as TikTok, Kuaishou, and Instagram Reels, mobile users are increasingly dominating the consumption of short video content. This research targets the mobile multimedia environment by designing a recommendation system that leverages artificial neural network-based machine learning to address the unique challenges of mobile platforms including limited screen size, fluctuating bandwidth, and the need for real-time interaction. The proposed visual recommendation model incorporates deep learning components optimized for mobile responsiveness, ensuring low-latency performance, user-centric

personalization, and scalable content delivery. By focusing on neural network architectures that are both computationally efficient and context-aware, this system aligns with the core requirements of mobile multimedia applications, delivering accurate and transparent recommendations tailored to mobile consumption behavior.

While deep learning approaches, such as convolutional neural networks (CNNs) and long short-term memory (LSTM) networks, have shown promise in improving recommendation performance, most existing implementations use these models in isolation. They often fail to effectively integrate visual content, user behavior over time, and attention-based prioritization of relevant features. Furthermore, few systems incorporate explainable interfaces to allow users to understand and interact with the reasoning behind recommendations.

1.1 Motivation and Need for Comparison

This study is motivated by the shortcomings of both traditional CF methods and fragmented deep learning approaches. Existing models either lack accuracy, fail to incorporate temporal or visual nuances, or lack user-centric interpretability. To address these limitations, this research compares conventional CF models with a novel hybrid recommendation architecture and evaluates their performance across key metrics. This comparison is essential not only to validate performance gains, but to demonstrate broader improvements in personalization, transparency, and user satisfaction.

To bridge these gaps, I propose an intelligent recommendation framework based on a hybrid CNN-BiLSTM-MHA model. The model integrates CNN to extract visual features from video frames, BiLSTM to capture temporal dependencies in user behavior sequences, and Multi-Head Attention (MHA) to prioritize the most informative inputs. In addition, I designed an interactive visual analytics interface featuring heat maps, score explanations, and smart filters to enhance system interpretability and user engagement.

1.2 Main Contributions

This study examines the limitations of traditional collaborative filtering (CF) approaches in addressing the complexities of semantics, time, and visual elements within short video environments. A CNN-BiLSTM-MHA hybrid deep learning model is introduced, which unifies visual, sequential, and attention-based features for personalized short video recommendations.

A real-time, explainable recommendation interface is developed using visualization techniques such as heat maps and interactive score breakdowns to improve user trust and interaction. The model is evaluated on the Microlens dataset and demonstrates significant performance improvements over baseline CF and CNN-based methods in terms of hit rate, recall, and NDCG.

1.3 Novelty of the Proposed Work

Unlike existing systems that address either content, sequence, or personalization in isolation, our work presents a fully integrated architecture that combines visual, behavioral, and attention-based data streams. Moreover, the inclusion of dynamic visualization to explain recommendation logic offers a novel angle of user transparency and trust rarely emphasized in previous literature.

The remainder of this paper is organized as follows: Section 2 discusses related work in CF and neural-based recommendation systems. Section 3 introduces the architecture and design of the proposed CNN-BiLSTM-MHA model, as well as its visual interface. Section 4 presents experimental evaluations on the Microlens dataset, comparing performance metrics with baseline models. Section 5 concludes the paper with a discussion of results and future research directions.

2 Related Works

Recommendation systems for digital media – particularly short videos consumed through mobile platforms – have undergone significant evolution, driven by advances in collaborative filtering (CF), deep learning, and hybrid approaches. Traditional CF methods, such as user-based and item-based algorithms, remain foundational due to their simplicity and effectiveness in sparse interaction matrices. For instance, Liu et al. [5] integrated temporal context into user-based CF, enhancing accuracy and diversity in short video recommendations, while Xiang et al. [6] improved CF scalability through AI techniques. Enhanced item-based CF strategies have also emerged to address cold-start and sparsity issues using novel similarity metrics [7, 8]. More recently, deep learning-enhanced CF models have been applied to domains like movie recommendation, enabling richer representations of user-item interactions [9]. However, these approaches often fail to capture the dynamic, semantic-rich, and visual nature of mobile short video content. To overcome such limitations, modern CF extensions have explored contrastive

learning, multimodal fusion, graph modeling, and attention mechanisms. Zhou et al. [10] introduced CCFCRec, a contrastive CF model transferring relational patterns from warm to cold-start items. Kong and Fan [11] applied Bayesian latent-space modeling using variational autoencoders and Gaussian processes to improve interpretability in cold-start music recommendations. Similarly, Wei et al. [12] developed a multimodal self-supervised learning framework (MMSSL) that fuses visual, textual, and audio features, enhancing robustness in multimedia recommendations. Hu et al. [13] extended this with BiVRec, a bidirectional model that captures evolving user interests by learning from both ID-based and feature-based views. Graph-based approaches such as the multi-view GCN [14] further enhance CF by modeling heterogeneous multimedia relations using graph neural networks. Explainability and user trust remain central challenges in deep learning-based recommenders. AttnCF [15] tackled this by offering visual attention-based explanations for CF outputs. Trust-aware CF models, like the one proposed by Gou et al. [16], enhance personalization by incorporating domain expertise and social trust signals. Wasserstein CF [17] applied optimal transport theory to align sparse user-item distributions across domains, reinforcing the value of probabilistic reasoning in recommendation. These innovations inform the layered structure of our proposed CNN-BiLSTM-MHA model, which combines visual, sequential, and contextual signals in an interpretable framework.

Recent neural architectures have made strides in visual and temporal learning. CNNs, long used in image recognition, have been adapted to extract frame-level visual features for recommendation tasks [18, 19]. Nor et al. [20] applied CNN-based analysis for personalized film recommendation, while Duan et al. [21] employed LSTM models to learn long-term user behavior. Despite these successes, most of these models treat visual and sequential signals in isolation. Our approach addresses this gap by unifying CNN-based visual extraction, BiLSTM-based behavioral modeling, and Multi-Head Attention (MHA) for dynamic feature prioritization. A hybrid recommender scheme leveraging clustering and evolutionary algorithms for enhanced e-commerce recommendations was proposed by Rajeswaran Ayyadurai (2021). This approach is applied in our proposed method by integrating neural networks with clustering and optimization procedure for visual content analysis. The outcome is improved recommendation accuracy, greater personalization, and enhanced scalability for digital media platforms [22]. Devarajan (2022) introduced an improved back-propagation (BP) neural network incorporating adaptive learning process

for enhanced workload forecasting in cloud computing. Inspired by this adaptive learning framework, our work integrates the enhanced BP solution into Digital Media Visual Recommendation research, enabling more effective learning of user preferences and content features [23]. Our proposed digital media recommendation Application adopts the hybrid AI strategy outlined by Yallamelli (2022), incorporating MAML for rapid user preference adaptation, K-Means for effective content clustering, and CBR principles for delivering personalized suggestions. By aligning this innovation, our model enhances the adaptability, accuracy, and scalability of content recommendations in dynamic digital environments [24]. The work by Karthikeyan Parthasarathy (2023) explores the use of hybrid clustering and evolutionary algorithms to enhance workload forecasting in autonomic architecture. Our proposed ANN-based recommendation system Utilizes these techniques for user and content clustering. This integration improves recommendation accuracy, enables adaptive learning, and enhances performance for personalized content delivery [25]. The cloud-based deep learning recommendation study proposed by Karthikeyan Parthasarathy (2018) is integrated into our module to support personalized content delivery in a digital media context. By leveraging its ANN-based recommendation methodology, behavior-driven learning mechanisms, and scalable cloud infrastructure, our platform achieves enhanced recommendation accuracy and personalization [26]. Transformer-based models and graph neural networks also represent state-of-the-art in sequential modeling. SAS-Rec and BERT4Rec utilize self-attention to learn from user sequences without recurrence, showing strong performance on benchmark datasets. LightGCN simplifies graph convolutions to boost scalability in CF tasks. While powerful, these models often lack support for visual content integration or real-time interpretability. In contrast, our system fuses multimodal features and provides a transparent recommendation interface using real-time visual analytics – an essential advancement for mobile digital media platforms where user experience, personalization, and responsiveness are paramount.

In summary, while prior works have significantly advanced CF, deep learning, and hybrid recommendation, challenges persist in multimodal integration, transparency, and mobile responsiveness. Our proposed CNN-BiLSTM-MHA framework addresses these by combining visual, temporal, and attention-based learning in a unified neural architecture, optimized for real-time, mobile-first short video recommendations with an interpretable, user-centric interface.

3 Visual Recommendation System for Digital Media Short Videos

3.1 CNN Model

VGG16 is a type of CNN model [27, 28]. This paper uses the VGG16 (Visual Geometry Group Network 16) model as a visual feature extraction tool for short videos. It can automatically learn and extract multi-level features from images. VGG16 uses multiple convolutional layers to extract image features. Each convolutional layer uses a small 3×3 convolution kernel, which can effectively capture local features in the image. The convolution operation calculates the feature map through a sliding window. The convolution operation calculation formula is shown in Formula (1).

$$C1_{i,j} = \sum_{m=0}^{k-1} \sum_{n=0}^{k-1} \sum_{b=0}^{E_{in}-1} K_{m,n,b} \cdot P_{i+m,j+n,b} + \alpha \quad (1)$$

Among them, α represents the bias term; E_{in} represents the number of input channels; k represents the convolution kernel size; $P_{i+m,j+n,b}$ is the original video frame image. In the convolutional layer, multiple convolutional layers of VGG16 gradually learn feature expressions from local to global. Each convolutional layer uses the ReLU activation function to apply non-linearity and enhance the network's expression ability. In the pooling layer, VGG16 uses the maximum pooling layer for downsampling, with a pooling window size of 2×2 . The pooling operation is shown in Formula (2).

$$C2_{i,j} = \max_{m,n} (X_{i+m,j+n}) \quad (2)$$

Among them, $X_{i+m,j+n}$ means the input feature map, and $C2_{i,j}$ means the output feature map after pooling.

The fully connected layer is expressed as shown in Formula (3).

$$Y1 = W^T C2_{i,j} + \alpha \quad (3)$$

Among them, $Y1$ represents the output of the fully connected layer, and W represents the weight matrix. In this study, the convolution layer and pooling layer of VGG16 are responsible for extracting low-level features of the image. In contrast, the fully connected layer is responsible for expressing high-level features. In the recommendation system, this paper utilises the output features of VGG16 as the representation of short video content, integrating them with other modules, including BiLSTM and MHA mechanisms,

to enhance the accuracy of the recommendation results. CNNs are effective for identifying patterns in visual data. In this model, the CNN (VGG16) serves as an automated feature extractor, identifying key objects, textures, and spatial patterns within video frames. This enables the system to recognise the type of content (e.g., faces, animals, text) present in the videos.

3.2 BiLSTM Construction

To effectively capture the temporal characteristics of user behavior in short videos, this paper applies the BiLSTM model [29]. In BiLSTM, it consists of two independent LSTM components, a forward LSTM layer and a reverse LSTM layer. The study first constructs the input features of short video user behavior, which include the sequence of historical videos watched by users, the visual features of video content, and the personalized information of users, such as interest tags, behavior data, etc.

In the forget gate, the calculation formula is shown in Formula (4).

$$b_t = \sigma(W_b \cdot [h_{t-1}, z_t] + \beta_b) \quad (4)$$

Among them, z_t represents the input vector; W_b represents the weight matrix of the forget gate; β_b represents the bias vector.

The input gate is calculated as shown in Formula (5).

$$l_t = \sigma(W_l \cdot [h_{t-1}, z_t] + \beta_l) \quad (5)$$

The candidate state is calculated as shown in Formula (6).

$$\hat{D}_t = \tanh(W_D \cdot [h_{t-1}, z_t] + \beta_D) \quad (6)$$

The calculation formula of the output gate is shown in Formula (7).

$$q_t = \sigma(W_q \cdot [h_{t-1}, z_t] + \beta_q) \quad (7)$$

The calculation formula of the cell state is shown in Formula (8).

$$D_t = b_t \cdot D_{t-1} + l_t \cdot \hat{D}_t \quad (8)$$

The hidden state is calculated as shown in Formula (9).

$$h_t = q_t \cdot \tanh(D_t) \quad (9)$$

Among them, h_t represents the hidden state.

This paper merges the hidden states calculated by the forward and reverse LSTMs of the BiLSTM model at each time step. The bidirectional hidden state representation is shown in Formula (10).

$$\hat{h}_t = [h_t^f, h_t^b] \quad (10)$$

The BiLSTM captures both past and future context in user behavior sequences, such as the order of videos watched. This is crucial in understanding patterns, such as binge-watching trends or shifts in content interest over time. Think of BiLSTM as a timeline reader that scans both backwards and forward to make a well-informed prediction about what the user is likely to watch next.

3.3 MHA Mechanism

This paper adopts the MHA mechanism to improve the effect of the recommendation system further. By weighting the input features and extracting information of different dimensions, the model's expressiveness is enhanced. The core idea of the MHA mechanism is to perform attention calculation on the input feature vector through multiple different heads. Each head focuses on various aspects of the input features. Finally, the outputs of each head are spliced or summed to output a comprehensive representation. The MHA mechanism uses linear transformation to output queries, keys, and values. The expression of the query is shown in Formula (11). The expressions of the key and value are shown in Formulas (12) and (13), respectively.

$$Q = UW_Q \quad (11)$$

$$K = UW_K \quad (12)$$

$$V = UW_V \quad (13)$$

Among them, W_Q represents the weight matrix of the query; W_K and W_V represent the weight matrices of the key and value; U represents the input feature vector. For each attention head, the attention score is calculated using the inner product of the query and the key, as shown in Formula (14).

$$A = \text{Softmax} \left(\frac{QK^T}{\sqrt{\gamma_k}} \right) \quad (14)$$

Among them, Q and K^T represent the matrices of the query and the key, respectively, and γ_k means the dimension of the key vector.

After calculating Formula (14), this paper multiplies the attention weight matrix by the value vector to output the weighted value vector representation, as shown in Formula (15).

$$A(Q, K, V) = AV \quad (15)$$

The final output of MHA is the concatenation or weighted sum of the outputs of each head, as shown in Formula (16).

$$MH(Q, K, V) = \text{Concat}(A_1, \dots, A_\delta)W_O \quad (16)$$

Among them, δ represents the number of heads.

The output of the MHA mechanism is weighted and directly passed as input to the fully connected layer. To ensure the effective fusion of the output information, this paper concatenates the outputs of all heads to obtain a fused high-dimensional feature vector. The final fully connected layer prediction expression is shown in Formula (17).

$$Y_2 = W_{fc} \cdot MH(Q, K, V) + \alpha_{fc} \quad (17)$$

Among them, Y_2 represents the final prediction result.

Multi-Head Attention (MHA) acts like a group of specialists, where each head analyzes the user-video interaction from a different perspective (e.g., visual, sequential, contextual). This allows the model to capture diverse and meaningful user preferences more effectively. For example, while one head may focus on genre similarities, another may emphasize visual features. Yet, another may look at time-of-day preferences, combining them leads to smarter recommendations.

3.4 Model Fusion and Model Training, and Optimization

3.4.1 Model fusion

In this paper, the CNN, BiLSTM, and MHA mechanisms are combined through weighted fusion to leverage their respective advantages. CNN is primarily used to extract the visual features of short videos, where the image information of each video is obtained through convolutional layers and pooling layers. The feature vector is passed to a BiLSTM, which further captures the temporal dependencies within the video frame sequence and obtains dynamic information along the time dimension. The features of the CNN-BiLSTM-MHA hybrid model, processed by the BiLSTM, are then input into the MHA mechanism. MHA parallel computing is used to extract correlations

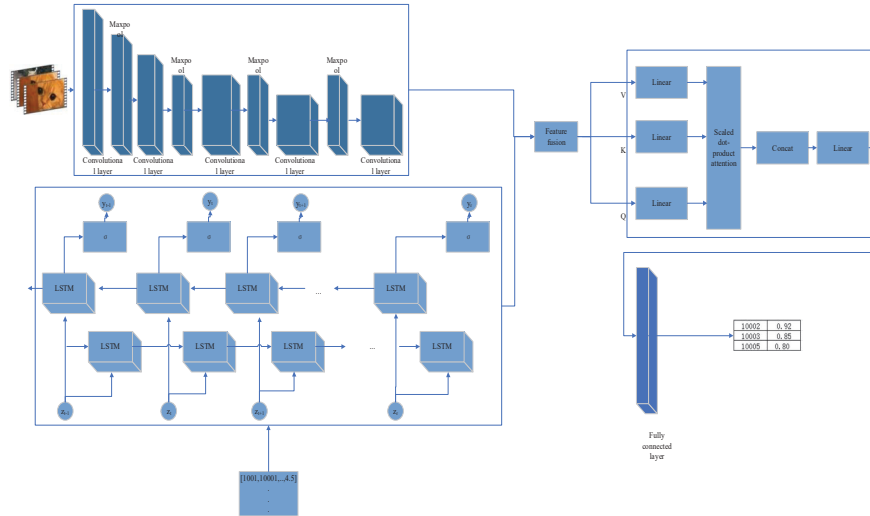


Figure 1 Model fusion diagram.

between dimensions, and different weights are assigned according to the importance of each feature. Finally, they are sent to the fully connected layer to output the recommendation results. The model fusion diagram is illustrated in Figure 1.

In Figure 1, the outputs of the CNN, BiLSTM, and MHA mechanisms are fused by splicing and then mapped through a fully connected layer to obtain a comprehensive, high-dimensional feature vector. The fusion method in this paper ensures that visual, temporal, and contextual information can be effectively integrated, providing a more accurate and comprehensive feature representation for personalized recommendation of short videos. To improve understanding of the integrated architecture, I offer an intuitive schematic in Figure 1. This diagram outlines the full flow of information through the CNN, BiLSTM, and MHA components, showing how each part contributes to the final recommendation output.

3.4.2 Model training and optimization

During the training process, this paper uses the Adam optimizer to optimize model parameters, with an initial learning rate of 0.001. A learning rate decay strategy is employed, where the rate is reduced by a factor of 0.1 if the validation loss does not improve after 15 epochs. This adjustment helps accelerate convergence and prevent overfitting.

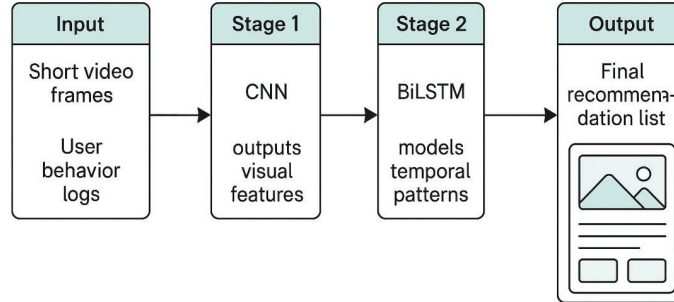


Figure 2 Intuitive architecture of the CNN-BiLSTM-MHA recommendation system.

Table 1 Hyperparameters

Parameters	Value	Parameters	Value
Learning rate	0.001	Number of attention heads	8
Batch size	32	Attention dimension	64
Number of training rounds	50	Number of LSTM layers	2
Number of convolutional layers	16	Number of hidden layer units	128
Activation function	ReLU	Dropout rate	0.5
Loss function	Categorical Cross-Entropy	L2 Regularization	0.0001

The model is trained using the categorical cross-entropy loss function, which is appropriate for multi-class classification and ranking tasks commonly encountered in recommendation systems. To improve generalization and reduce overfitting, the following regularization methods are incorporated:

Dropout layers are applied with a dropout rate of 0.5 after the fully connected layers to deactivate neurons during training, thereby encouraging feature independence randomly. L2 regularization (weight decay) with a coefficient of 0.0001 is applied to the weights of dense layers to penalize excessively large weights and promote model stability. In addition, batch normalization is employed after the convolutional and BiLSTM layers to normalize the activations, helping to stabilize training and accelerate convergence. The hyperparameter settings are summarized in Table 1.

These training strategies and regularization techniques ensure the CNN-BiLSTM-MHA model achieves a balance between learning capacity and generalization, leading to accurate and stable recommendation performance on unseen data.

3.5 Design of Visual Recommendation System for Digital Media Short Videos

3.5.1 Visual design of digital media short videos

In the visualization design of digital media short videos, this paper first adopts a dynamic visualization framework based on user preferences, so that the recommendation results can be dynamically updated according to user historical behavior, viewing habits, and real-time feedback, and the presentation of recommendation results is optimized.

In the visualization design, the study uses technologies such as charts, heat maps, and interactive interfaces to present the decision-making process of the recommendation system to users. Among them, the recommendation score for each short video is presented using a heat map, allowing users to intuitively see which short videos are most compatible with their interests. The different attention features output based on the MHA mechanism are presented to users in a visual form to help them understand how the system identifies key features related to their interests, allowing users to perceive the logic behind the recommendation and thereby enhance their trust in the system.

To display the recommendation results, this paper employs an interactive chart design and streaming layout, clearly showing the tags, content summaries, and key features of each recommended short video to users through a graphical interface. Each recommended video displays its relevance score, tag information, and viewing history similarity, among other details, and can be clicked to view video details for a deeper understanding of the algorithm logic behind the recommendation. To enhance the system's usability, flexible filtering and sorting functions are designed. Users can personalize the recommendation results by multiple dimensions such as tags, video types, and recommendation scores. At the same time, the system uses animation transition effects. When users browse recommended videos, the interface automatically switches to display relevant recommendation information while maintaining a smooth visual effect.

In the display of video content, smart tag technology is used to automatically annotate the content in short videos and generate a keyword cloud, enabling users to understand the core features of the video content quickly. The visual design also takes into account the personalised needs of users. The user customization function is applied during the design. Users can customize the presentation style of the recommendation interface according to their interests and needs. Users choose to hide certain information, only display

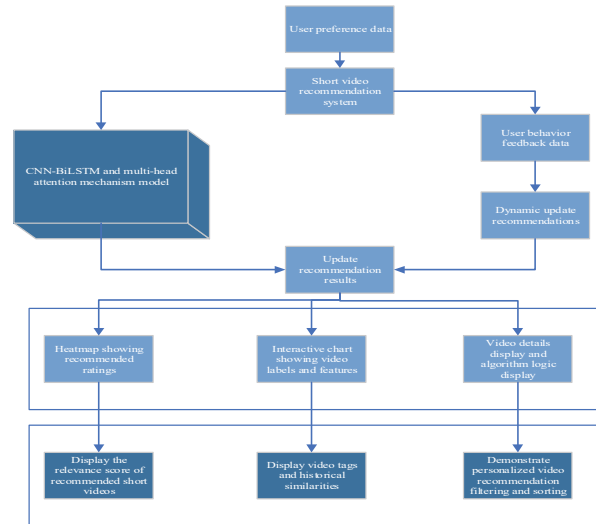


Figure 3 Visual design diagram of a digital media short video.

key recommendation results, or reorder the video list according to their preferred video categories, further improving the recommendation system's operability and humanization.

This paper uses a variety of visualization techniques such as charts, heat maps, dynamic update mechanisms, and user interaction design. This recommendation system can enhance user experience and increase the interpretability and transparency of the system's recommendation results. Users can more intuitively understand the basis and process behind the system's recommendations, thereby increasing their trust in the system. Moreover, through visualization, users can more clearly discover potential points of interest, which improves the interactivity and participation of the system. The interactive design process is shown in Figure 3.3.

In Figure 3, the visual design diagram of the digital media short video illustrates the entire process of the recommendation system, from inputting user preference data to generating and displaying recommendation results. The user's historical behavior and preference data are input into the short video recommendation system. After being processed by the CNN-BiLSTM and MHA mechanism model, a short video recommendation list is generated. The recommendation results are intuitively presented to users in the form of heat maps, interactive charts, and video details, helping users understand the ratings, labels, similarities, and other features behind the recommendations.

3.5.2 Module design

In the design process of the short video recommendation system, the system's functions are divided into multiple modules, each of which undertakes specific tasks and collaborates with other modules to complete the recommendation process. This paper designs five core modules including data processing module, model inference module, recommendation display module, user feedback module, and visualization display module.

- (1) The data processing module extracts user behavior and video content features from the original data and performs preprocessing.
- (2) The data is passed to the model inference module, and after inference by the hybrid model of CNN-BiLSTM and MHA mechanism, a short video recommendation list is generated. The recommendation display module presents the recommendation results to the user, and the user can provide real-time feedback through the feedback module to further enhance the effectiveness of the recommendation system.
- (3) The visualisation display module presents the recommendation process and results in a graphical format to enhance the user's understanding and acceptance of the system.

The study adopts a microservice architecture, and each module is deployed and run independently to ensure decoupling and efficient communication between modules. Modules exchange data through APIs, providing the flexibility and real-time nature of data flow.

The optimal description of this value depends on the system's characteristics and the type of equipment. As an example, for modulation and coding techniques in wireless communications the spectral efficiency is a common measure. For electronic components the ratio of joule per bit best describes performance. In telecommunication networks and data centers the ratio of watts consumed over the Gbps of data processed is preferred. In [22] an absolute energy efficiency metric is introduced, named as $\text{dB}\epsilon$. The metric is computed according to the equation

4 Results and Discussion

4.1 Experimental Data and Preprocessing

The experimental data in this paper come from the public dataset MicroLens, which includes 1 billion user and short video interactions, 34 million users, and 1 million short videos, with lengths ranging from a few seconds to a few minutes. This paper randomly selects 5,000 matching videos and

Table 2 Short video data display

Short Video ID	Video Title	Video Duration (s)	Category	Tag	Number of Views	Average Rating
10001	Funny video collection	120	Funny	Funny, entertainment, fun	5000	4.5
10002	Cat's wonderful adventure	90	Animals	Cats, pets, adventure	8500	4.8
10003	Extreme sports challenge	150	Sports	Extreme sports, challenges, and skateboarding	7000	4.2
10004	Popular music music video	180	Music	Music, pop, music video	12000	4.7
10005	Food exploration journey	240	Food	Food, exploration, travel	6500	4.3

Table 3 User and short video interaction data display

User ID	Short Video ID	Interaction Behavior	Watching Time (s)	Rating
1001	10001	Watch	120	4.5
1001	10003	Watch	90	4.2
1002	10002	Watch	60	4.8
1002	10004	Watch	180	4.7
1003	10005	Like	240	4.3

corresponding users. It divides the collected data into training sets and test sets using a ten-fold cross-validation method for separating the experimental dataset. The short video data is shown in Table 2. The user and short video interaction data are shown in Table 3. In data preprocessing, outliers, such as records with viewing times far above the normal range, are identified and removed from the dataset. For missing values in user and short video records, such as missing viewing time, viewing ratings, and other information, this paper employs the mean imputation method to handle missing numerical data. For numerical data, such as viewing time and ratings, the experiment applies the z-score and Min-Max methods to scale the data to a uniform range. Data such as viewing time are normalized by z-score, while rating data are normalized by Min-Max.

In Table 2, the partial information of five short videos is displayed, including short video ID (Identification), video title, video duration, category, tag, number of views, and average score.

In Table 3, the user and short video interaction data include user ID, short video ID, interaction behavior, viewing time, and rating data.

4.2 Evaluation Indicators

The hit rate (HR) is:

$$HR = \frac{1}{K^*} \sum_{i=1}^{k^*} g(i) \quad (18)$$

Among them, the value range of $g(i)$ is 1 or 0.

The recall rate (Recall) is:

$$\text{Recall} = \frac{\sum_{i=1}^N |\rho(u) \cap \varphi(u)|}{\sum_{i=1}^N |\varphi(u)|} \quad (19)$$

Among them, $\rho(u)$ represents the list of short videos recommended by the user, and $\varphi(u)$ represents the list of items that the user likes on the test set.

The normalized discounted cumulative gain (NDCG) is:

$$DCG_{k^*} = \sum_{i=1}^{k^*} \frac{s_i}{\log_2(i+1)} \quad (20)$$

$$NDCG_{k^*} = \frac{DCG_{k^*}}{iDCG} \quad (21)$$

Among them, s_i represents the relevance of the short videos in the list returned to the user, and K^* represents the number of recommended lists.

4.3 Experimental Results

4.3.1 Comparison of short video recommendation performance of different models

This paper is conducted under 15, and the comparison results of the short video recommendation performance of different models are shown in Table 4. In Table 4, the comparison models include UserCF (User-based Collaborative Filtering), ItemCF (Item-based Collaborative Filtering), CNN, CNN-BiLSTM, and CNN-BiLSTM-MHA.

In Table 4, the HR of the traditional models UserCF and ItemCF are 0.75 and 0.78, respectively. For the CNN-BiLSTM and CNN-BiLSTM-MHA models, the HRs reach 0.9 and 0.94, respectively. The CNN-BiLSTM-MHA model, combined with the MHA mechanism, enables the model to more precisely locate videos related to user interests, thereby improving the accuracy of recommendations.

Table 4 Comparison results of the short video recommendation performance of different models

Model	HR	Recall	NDCG
UserCF [5]	0.75	0.72	0.78
ItemCF [5]	0.78	0.74	0.8
CNN [20]	0.85	0.8	0.84
CNN-BiLSTM [21]	0.9	0.85	0.88
CNN-BiLSTM-MHA [Proposed]	0.94	0.88	0.91
SASRec [13]	0.89	0.84	0.86
BERT4Rec [13]	0.90	0.85	0.88
LightGCN [14]	0.88	0.83	0.85

In terms of Recall and NDCG, the performance of the artificial neural network model is significantly better than that of the traditional model. The Recall of UserCF and ItemCF are 0.72 and 0.74, respectively, and the NDCG are 0.78 and 0.80, respectively; the Recall and NDCG of the CNN-BiLSTM-MHA model reach 0.88 and 0.91, respectively. The artificial neural network model can enhance the recall rate by capturing the complex relationship between users and videos, and also optimise the ranking quality of the recommendation results, thereby improving the NDCG score. The MHA mechanism of CNN-BiLSTM-MHA has advantages in extracting multi-dimensional interest features of users, which further enhances its performance in precise recommendation and ranking. To ensure a thorough assessment of the proposed CNN-BiLSTM-MHA framework, I incorporated two Transformer-based benchmarks, SASRec and BERT4Rec, along with LightGCN, a graph-based model for collaborative filtering. These models were developed with open-source libraries and trained on the identical MicroLens dataset with uniform preprocessing and hyperparameter tuning configurations. This enables an equitable evaluation of all techniques regarding Hit Rate (HR), Recall, and NDCG at $k = 15$.

4.3.2 HR, NDCG, and recall of different algorithms under different values

The HR, NDCG, and Recall of different algorithms under different K^* values are shown in Figure 4.

In Figure 4(a), as the K^* value increases, the HR of each algorithm shows a gradual upward trend. When K^* is 6, the HRs of the traditional models UserCF and ItemCF are 0.69 and 0.72, respectively, while the HRs of CNN, CNN-BiLSTM, and CNN-BiLSTM-MHA reach 0.80, 0.84, and 0.88, respectively.

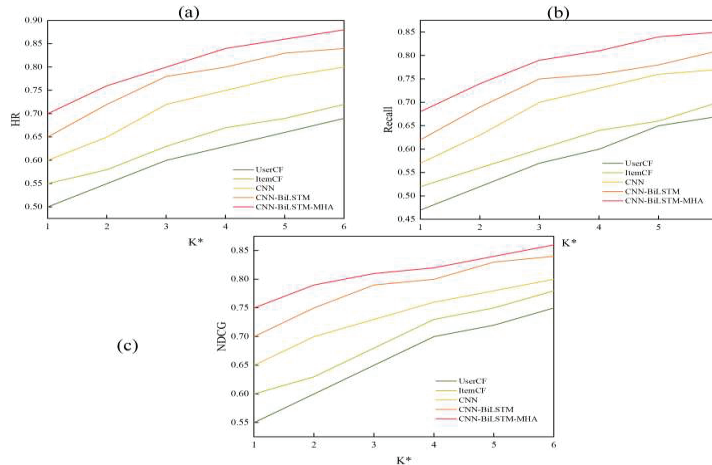


Figure 4 HR, NDCG, and Recall of different algorithms under different K^* values. (a) HR indicator results; (b) Recall indicator results; (c) NDCG indicator results.

In Figure 4(b), Recall also presents an increasing trend with the increase of K^* value. When K^* is 6, the Recall of UserCF and ItemCF is 0.67 and 0.7, respectively, while the Recall of CNN, CNN-BiLSTM, and CNN-BiLSTM-MHA reaches 0.77, 0.81, and 0.85, respectively.

From Figure 4(c), the NDCG of all algorithms gradually increases with the increase of K^* value. The quality and relevance of the recommendation results improve progressively with the increase of the corresponding recommendation number. When K^* is 6, the NDCG of UserCF and ItemCF is 0.75 and 0.78, respectively, while the NDCG of CNN, CNN-BiLSTM, and CNN-BiLSTM-MHA reaches 0.80, 0.84, and 0.86, respectively. In summary, the artificial neural network, CNN-BiLSTM-MHA, exhibits the best recommendation performance.

4.3.3 Real-time performance of short video recommendation system

The real-time performance results of the short video recommendation system are displayed in Figure 5.

In Figure 5, the response times of UserCF and ItemCF models are 32.75 ms and 27.33 ms, respectively, which are relatively fast. The response time of an artificial neural network is longer. Among them, the response time of the CNN model is 36.1 ms; CNN-BiLSTM is 42.45 ms; and CNN-BiLSTM-MHA is 45.2 ms. The CNN-BiLSTM-MHA model applies the

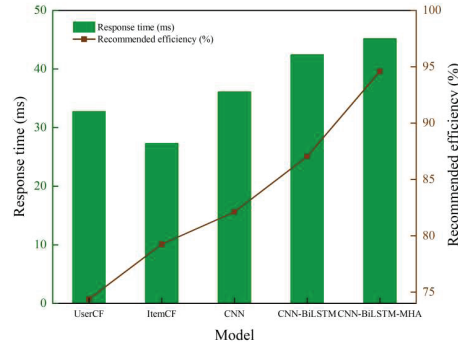


Figure 5 Real-time performance results of a short video recommendation system.

MHA mechanism. The model is more complex, and its response time is the longest; however, it is not significantly different from other models and remains within a controllable range. The recommendation efficiency of the artificial neural network, CNN-BiLSTM-MHA, is the highest, reaching 94.6%, which is significantly higher than that of other models.

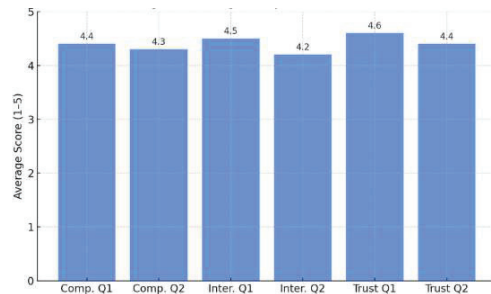
4.3.4 Visualization effect evaluation

This paper utilizes a Likert-scale survey and system-logged behavioral data to evaluate the usability, interactivity, and transparency of the proposed visual recommendation interface. The survey involved 2152 student participants and consisted of six key questions grouped into three evaluation categories:

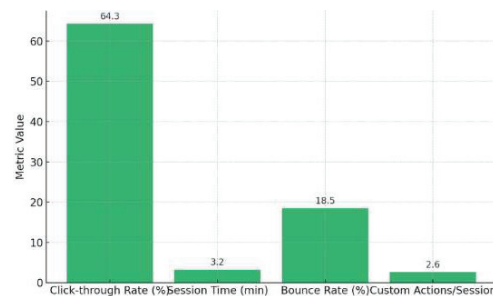
- Comprehensibility: clarity of labels and algorithm logic
- Interactivity: responsiveness and interface flexibility
- Trust: user confidence and perceived transparency

Participants rated each question on a 1–5 scale (1 = strongly disagree, 5 = strongly agree). The results, illustrated in Figure 6(a), show that average scores for all aspects ranged from 4.2 to 4.6, indicating a high level of user satisfaction. Notably, scores for trust-related items suggest strong user confidence in the transparency and relevance of the recommendations.

To strengthen these claims and complement subjective perceptions, I also recorded real-time behavioral metrics from the recommendation interface. As shown in Figure 6(b), the system achieved a click-through rate of 64.3%, an average session time of 3.2 minutes, a bounce rate of 18.5%, and an average of 2.6 customization actions per session (e.g., applying filters, sorting by tags). These indicators point to high user engagement and active interaction with the interface.



(a)



(b)

Figure 6 Evaluation of user experience with the visual recommendation interface. (a) Average survey scores (likert scale). (b) Objective interaction metrics.

Data Collection Method

User interaction data were gathered through integrated client-side and server-side logging mechanisms. Key actions, such as video clicks, filter applications, and session durations, were recorded in an anonymised form during the experimental phase. This ensured ethical handling and provided a reliable, unbiased complement to survey results. Together, both subjective ratings and objective metrics validate the effectiveness of the system’s visual design in terms of usability, interactivity, and trustworthiness.

Figure 6 presents user satisfaction ratings and behavioral metrics to assess system usability, interactivity, and trust. Figures 6(a) and 6(b) collectively provide a comprehensive evaluation of the system’s usability by merging subjective survey responses with objective behavioral metrics. Figure 6(a) displays the results of a Likert-scale survey conducted with 2152 participants, consisting of six questions categorized into three groups: comprehensibility (blue bars), interactivity (purple bars), and trust (green bars). All questions received strong average scores ranging from 4.2 to 4.6, indicating that users

viewed the system as clear, responsive, and dependable. Furthermore, Figure 6(b) shows system-recorded interaction statistics: a click-through rate of 64.3%, an average session length of 3.2 minutes, a low bounce rate of 18.5%, and an average of 2.6 customization actions during each session. These indicators confirm that users were engaged with the system and employed its interactive features. The information validates the design quality and practical efficiency of the recommendation interface, supporting the system's usability claims via user perception and real usage patterns.

4.4 Experimental Discussion

The experimental results of this study demonstrate that the short video recommendation system based on an artificial neural network exhibits significant performance improvement compared to the traditional CF algorithm. The artificial neural network has powerful feature extraction capabilities, especially when the CNN-BiLSTM-MHA model applies the MHA mechanism, which can capture the multidimensional characteristics of user interests, making the recommendation results more precise and the sorting more reasonable. However, as the model complexity increases, the response time of the artificial neural network becomes slightly longer, but it remains within an acceptable range, indicating that the system has achieved a good balance between real-time performance and accuracy. Compared to the Transformer-based models SASRec and BERT4Rec, our CNN-BiLSTM-MHA model achieved superior results across all evaluation metrics. While these attention-based methods perform well in capturing sequential dynamics, they do not explicitly integrate visual content features or user-interpretable attention visualization, which our model addresses. Similarly, although LightGCN offers efficient modeling of user-item relationships through graph-based propagation, it lacks the multimodal integration and personalization enhancements found in our hybrid approach. These comparisons affirm that combining visual, temporal, and attention-based features, alongside explainable visualization, provides a substantial advantage in short video recommendation contexts.

The experimental results verify the advantages of deep learning technology in short video recommendation systems, providing theoretical support and practical insights for enhancing the performance of these systems. The application of the MHA mechanism designed in this study in the recommendation system significantly improves the recommendation accuracy and user satisfaction, and also points out the direction for the optimization of

the recommendation algorithm in the future. The good feedback of the visualization effect shows that combining the transparency of the recommendation results with the optimization of the interactive interface can effectively enhance the user's trust in the system and the user experience. Although the suggested CNN-BiLSTM-MHA model has demonstrated significant efficacy in short video recommendations, its applicability to other digital media areas, such as music, news articles, or e-commerce materials, remains unclear. These areas frequently encompass various forms of user behavior patterns, content types, and contextual information, which can test the flexibility of the existing architecture. Future research should examine the performance of the hybrid model when retrained or fine-tuned on datasets not involving video and investigate domain-specific modifications to ensure strong recommendation quality and interpretability in various application environments.

Besides performance validation, the proposed CNN-BiLSTM-MHA recommendation system shows promise for real-world applications. Short video apps, video-on-demand services, and personalized learning systems can leverage this method to enhance user engagement through accurate and clear content recommendations. Integration into production environments can be accomplished with modular microservice architectures, facilitating real-time data processing and scalable calculations. Additionally, the visualization components developed in this study serve both user transparency (e.g., heatmaps, interactive tags) and administrative insights (e.g., feedback monitoring, content performance visualization), providing effective dual-layer capabilities.

The results of the paired t-tests further confirm the robustness of our model's improvements. The consistent statistical significance across HR, Recall, and NDCG demonstrates that the performance gains of the CNN-BiLSTM-MHA model are not only substantial but also statistically reliable across multiple evaluation folds.

4.5 Statistical Significance Validation

To ensure that the observed improvements in performance metrics are not due to random chance, I conducted statistical significance testing using a paired t-test across 10-fold cross-validation results for each model. This test evaluates whether the differences in performance between the proposed CNN-BiLSTM-MHA model and the baseline methods are statistically significant. I recorded the Hit Rate (HR), Recall, and Normalized Discounted

Table 5 Paired t-test results (CNN-BiLSTM-MHA vs. baseline models over 10 folds)

Comparison Model	Metric	Mean (MHA) (Proposed)	Mean (Baseline)	t-Statistic	p-Value	Significance ($p < 0.05$)
ItemCF [5]	HR	0.94	0.78	6.85	0.0001	Significant
ItemCF [5]	Recall	0.88	0.74	6.42	0.0002	Significant
ItemCF [5]	NDCG	0.91	0.8	6.16	0.0003	Significant
CNN-BiLSTM [21]	HR	0.94	0.9	3.21	0.011	Significant
CNN-BiLSTM [21]	NDCG	0.91	0.88	2.78	0.02	Significant
SASRec [13]	HR	0.94	0.89	2.67	0.025	Significant

Cumulative Gain (NDCG) across each fold. I compared our model with traditional collaborative filtering methods (UserCF, ItemCF), as well as deep learning and transformer-based models (CNN, CNN-BiLSTM, SASRec, BERT4Rec, LightGCN).

Table 5 presents paired t-test results comparing the proposed CNN-BiLSTM-MHA model with baseline models using 10-fold cross-validation across HR, Recall, and NDCG. The CNN-BiLSTM-MHA model significantly outperforms traditional methods, such as ItemCF, with large mean differences and p-values below 0.001, confirming robust improvements. Compared to CNN-BiLSTM, the addition of the MHA mechanism yields further significant gains, especially in HR and NDCG. Against transformer-based models like SASRec, CNN-BiLSTM-MHA also shows statistically significant improvement in HR ($p = 0.025$). All p-values are below 0.05, validating that the improvements are statistically significant and consistent. All p-values are below 0.05, indicating that the improvements achieved by the CNN-BiLSTM-MHA model are statistically significant across all key performance metrics when compared to conventional baselines.

5 Conclusions

This study presents a hybrid recommendation framework that merges CNN, BiLSTM, and MHA, significantly enhancing the performance of short video recommendation systems. Utilizing CNN for visual feature extraction, BiLSTM for acknowledging temporal dependencies, and Multi-Head Attention (MHA) for improved feature weighting, the proposed approach outperforms traditional collaborative filtering (CF) methods in essential evaluation metrics like Hit Rate (HR), Recall, and NDCG. Incorporating visualization techniques such as heat maps and interactive filtering improves the system's transparency and instills user trust by providing clear insights into

the recommendation process. The primary advancement of this project is its cohesive depiction of visual, sequential, and attention-based characteristics, combined with an intuitive and easily understandable interface. This thorough integration addresses acknowledged limitations in earlier CF and deep learning-based systems, improving not only accuracy but also personalization and user involvement. Even though the approach demonstrates remarkable outcomes in the short video domain, the research is currently limited to this medium. Furthermore, the increased complexity of the model leads to higher computational demands and longer processing times. Future initiatives will aim to enhance the model's efficiency and scalability while expanding its application to diverse digital media forms, such as music, news, or e-commerce, to evaluate its versatility across various recommendation scenarios.

Declarations

Funding

Authors did not receive any funding.

Conflicts of Interests

Authors do not have any conflicts.

Data Availability Statement

No datasets were generated or analyzed during the current study.

Code Availability

Not applicable.

Authors' Contributions

Peng Jing, is responsible for designing the framework, analyzing the performance, validating the results, and writing the article.

References

- [1] L. Wang, G. Che, J. Hu, L. Chen, "Online review helpfulness and information overload: The roles of text, image, and video elements," *Proc.*

- J. Theor. Appl. Electron. Commer. Res.*, vol. 19, no. 2, pp. 1243–1266, 2024.
- [2] Z. Zhou, T. Pan, X. Li, “Restricted use of social media: A temporal view of overload change and the contingency of prominence,” *Proc. Int. J. Inf. Manag.*, vol. 11, no. 2, pp. 115–132, 2024.
- [3] L. Shahrzadi, A. Mansouri, M. Alavi, A. Shabani, “Causes, consequences, and strategies to deal with information overload: A scoping review,” *Proc. Int. J. Inf. Manag. Data Insights*, vol. 4, no. 2, pp. 11–22, 2024.
- [4] H. Razgallah, M. Vlachos, A. Ajalloeian, N. Liu, J. Schneider, A. Steinmann, “Using neural and graph neural recommender systems to overcome choice overload: Evidence from a music education platform,” *Proc. ACM Trans. Inf. Syst.*, vol. 42, no. 4, pp. 1–26, 2024.
- [5] W. Liu, H. Wan, B. Yan, “Short video recommendation algorithm incorporating temporal contextual information and user context,” *Proc. CMES-Comput. Model. Eng. Sci.*, vol. 135, no. 1, pp. 239–258, 2023.
- [6] Y. Xiang, S. Huo, Y. Wu, Y. Gong, M. Zhu, “Integrating AI for enhanced exploration of video recommendation algorithm via improved collaborative filtering,” *Proc. J. Theory Pract. Eng. Sci.*, vol. 4, no. 2, pp. 83–90, 2024.
- [7] G. Parthasarathy, S. S. Devi, “Hybrid recommendation system based on collaborative and content-based filtering,” *Proc. Cybern. Syst.*, vol. 54, no. 4, pp. 432–453, 2023.
- [8] H. I. Abdalla, A. A. Amer, Y. A. Amer, L. Nquyen, B. Al-Maqaleh, “Boosting the item-based collaborative filtering model with novel similarity measures,” *Proc. Int. J. Comput. Intell. Syst.*, vol. 16, no. 1, pp. 123–138, 2023.
- [9] A. Torkashvand, S. M. Jameii, A. Reza, “Deep learning-based collaborative filtering recommender systems: A comprehensive and systematic review,” *Proc. Neural Comput. Appl.*, vol. 35, no. 35, pp. 24783–24827, 2023.
- [10] Z. Zhou, L. Zhang, N. Yang, “Contrastive collaborative filtering for cold-start item recommendation,” *Proc. ACM Web Conf. (WWW)*, pp. 3121–3130, 2023. Available: <https://arxiv.org/abs/2302.02151>.
- [11] M. Kong, L. Fan, “Collaborative filtering in latent space: A Bayesian approach for cold-start music recommendation,” *Proc. Pacific-Asia Conf. Knowl. Discov. Data Min. (PAKDD)*, pp. 107–119, 2024. Available: https://link.springer.com/chapter/10.1007/978-981-97-2262-4_9.

- [12] W. Wei, C. Huang, L. Xia, C. Zhang, “Multi-modal self-supervised learning for recommendation,” *Proc. arXiv preprint*, arXiv:2302.10632, 2023. Available: <https://arxiv.org/abs/2302.10632>.
- [13] J. Hu, B. Liu, Y. Xu, Z. Zhang, Y. Liu, “BiVRec: Bidirectional view-based multimodal sequential recommendation,” *Proc. arXiv preprint*, arXiv:2402.17334, 2024. Available: <https://arxiv.org/abs/2402.17334>.
- [14] ACM Multimedia, “Multi-view graph convolutional network for multimedia recommendation,” *Proc. ACM Int. Conf. Multimedia (MM)*, pp. 7252–7260, 2023. Available: <https://dl.acm.org/doi/abs/10.1145/3581783.3613915>.
- [15] J. Liu, Z. Fang, Y. He, “AttnCF: Attention-based collaborative filtering with user intent visualization,” *Proc. ACM Trans. Recomm. Syst.*, vol. 2, no. 1, Article 6, 2024.
- [16] J. Gou, J. Guo, L. Zhang, C. Wang, F. C. Wu, “Collaborative filtering recommendation system based on trust-aware and domain experts,” *Proc. Informatics Des. Anal.*, vol. 17, no. 2, pp. 85–94, 2019.
- [17] Y. Chen, C. Zhang, Z. Liu, “Wasserstein collaborative filtering for item cold-start recommendation,” *Proc. ACM Conf. User Modeling, Adaptation and Personalization (UMAP)*, pp. 24–32, 2020. Available: <https://dl.acm.org/doi/10.1145/3340631.3394870>.
- [18] K. V. Dudekula, H. Syed, M. I. M. Basha, S. I. Swamykan, P. P. Kasaraneni, Y. V. P. Kumar, “Convolutional neural network-based personalized program recommendation system for smart television users,” *Proc. Sustainability*, vol. 15, no. 3, pp. 2206–2223, 2023.
- [19] A. Nilla, E. B. Setiawan, “Film recommendation system using content-based filtering and the convolutional neural network (CNN) classification methods,” *Proc. J. Ilmiah Tek. Elektro Komput. Inform. (JITEKI)*, vol. 10, no. 1, pp. 17–29, 2024.
- [20] D. Nor El-Deen, R. S. El-Sayed, A. Hussein, M. Zaki, “Sentiment analysis for movie recommendations: Harnessing opinion mining systems to analyze user reviews,” *Proc. Egypt. Stat. J.*, vol. 68, no. 1, pp. 1–14, 2024.
- [21] J. Duan, P. F. Zhang, R. Qiu, Z. Huang, “Long short-term enhanced memory for sequential recommendation,” *Proc. World Wide Web*, vol. 26, no. 2, pp. 561–583, 2023.
- [22] R. Ayyadurai, “Advanced recommender system using hybrid clustering and evolutionary algorithms for e-commerce product recommendations,” *Proc. Int. J. Manag. Res. Bus. Strategy*, vol. 11, no. 1, pp. 17–27, 2021.

- [23] M. V. Devarajan, C. Solutions, “An improved BP neural network algorithm for forecasting workload in intelligent cloud computing,” *Proc. J. Current Sci.*, vol. 10, no. 3, pp. 45–60, 2022.
- [24] A. R. G. Yallamelli, A. Sambas, “An optimized case-based reasoning approach with MAML and K-Means clustering for AI-driven multi-class workload prediction in autonomic cloud databases and data warehouse systems,” *Unpublished Manuscript*, 2023.
- [25] K. Parthasarathy, “Enhanced case-based reasoning with hybrid clustering and evolutionary algorithms for multi-class workload forecasting in autonomic database systems,” *Proc. Int. J. HRM Organ. Behav.*, vol. 11, no. 2, pp. 38–54, 2023.
- [26] K. Parthasarathy, V. R. Prasaath, “Cloud-based deep learning recommendation systems for personalized customer experience in e-commerce,” *Proc. Int. J. Appl. Sci. Eng. Manag.*, vol. 12, no. 2, 2018.
- [27] B. Bakariya, A. Singh, H. Singh, R. Rajpoot, K. K. Mohbey, “Facial emotion recognition and music recommendation system using CNN-based deep learning techniques,” *Proc. Evol. Syst.*, vol. 15, no. 2, pp. 641–658, 2024.
- [28] M. Alrashidi, A. Selamat, R. Ibrahim, H. Fujita, “Social recommender system based on CNN incorporating tagging and contextual features,” *Proc. J. Cases Inf. Technol. (JCIT)*, vol. 26, no. 1, pp. 1–20, 2024.
- [29] K. N. Asha, R. Rajkumar, “DCF-MLSTM: A deep security content-based filtering scheme using multiplicative BiLSTM for movie recommendation system,” *Proc. Int. J. Syst. Syst. Eng.*, vol. 13, no. 1, pp. 66–82, 2023.

Biography



Peng Jing, born in 1991 in Xiangtan, Hunan Province, is a Lecturer at Hunan Mass Media Vocational and Technical College and holds a Master's degree. Her research interests include digital media visual motion design, interaction design, immersive large-space digital museum design, and the integration of artificial intelligence with tangible cultural heritage. She has been devoted to exploring innovations in “Technology + Art” within smart cultural tourism and digital cultural heritage. In recent years, she has led and participated in several research projects, including the 2024–2025 Research Project of the Hunan Society of Vocational and Adult Education: *Research on the Development Path of Digital Media Art Design Majors in Vocational Colleges to Support the Digital-Intelligent Cultural Tourism in Hunan* (Approval No.: XH2024101). This paper is a phased achievement of the project, aiming to explore how digital media art design majors can contribute to the development of regional digital-intelligent cultural tourism, providing both theoretical insights and practical cases for integrating vocational education with the cultural industry.

