ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 1 - 2026

Design of a New Media Short Video Content Recommendation System Based on User and Emotion Perception

JiaoZhang[1,*], *XianJing Bai*[2] *and MeiLing Ta*[1]

[1] School of Education, China University of Petroleum (Beijing), Beijing, 102249, China
[2] College of Safety and Ocean Engineering, China University of Petroleum (Beijing), 102249, Beijing, China

SUMMARY: *Considering the current development trends within the sphere of digital media, personalization of recommendation services is one of the key challenges faced by modern recommender systems. Indeed, existing recommendation frameworks rely mostly on static user profiling and history-based analysis of their previous interactions, thus failing to adequately reflect the dynamics of emotional state changes and evolving preferences of the target audience. In order to solve the problem mentioned above, an innovative approach to short video content recommendation based on Emotion Driven Manifold Router framework is proposed in this paper. The Emotion Driven Manifold Router consists of three main components including Counterfactual Manifold Optimizer used to optimize content mapping based on simulated counterfactuals regarding user interactions, Agent Based Emotion Segmentation used to create dynamic clusters of users based on inferred emotion information, and finally Probabilistic User Perception Model used to estimate user preferences in terms of probabilities derived from emotion-behavior data. These three components operate through the mechanism of Policy Driven Coordination based on stochastic refinement of the entire process. As a result, the developed system proved to have significantly higher efficiency with an average increase of 23% in terms of interaction rates and a 17% growth of user satisfaction.*

KEYWORDS: *Short Video Content Recommendation; Emotion Perception; User Engagement; Emotion Driven Manifold Route,*

1 Introduction

In the recent past, the explosion in short video platforms has affected the way users engage with media, making recommendation systems all the more important. The role of recommendation systems here goes beyond merely increasing engagement through personalized content. They are also crucial for keeping users on the platform for an extended period of time. The reason for developing recommendation systems is the massive amount of available content which, when not presented in the right way, can cause fatigue in users and hence reduce their satisfaction levels (Phillips et al., 2003). In addition, recognizing the emotions and preferences of the user is critical to presenting content that will appeal to them and result in increased engagement and even sharing of content (Gross, 2015). This is no easy feat since, apart from ensuring user satisfaction, it requires presenting content based on their emotions and preferences to improve the chance of users engaging and sharing.

*Zhangjiao80@cup.edu.cn
https://doi.org/10.65102/is2026437
Published: 30 April, 2026

The earliest development of recommendation systems was through manually designed rules and models that captured the user's preferences and those of the content as well. The recommendations would be made through the use of logic built into the system that would allow them to predict the type of content a particular user needed at any point in time. Such methods helped build some insight into the way users behaved; however, they did not offer much in terms of flexibility and adaptation (Barrett et al., 2011).

Several attempts have been made to develop algorithms that could learn the patterns from user interaction data directly. Collaborative filtering and matrix factorization have become popular choices, as they allow inference of the user preference from interaction history (Atkinson and Adolphs, 2005). They exhibit better scalability and adaptability while delivering personalized recommendations. Nevertheless, these techniques typically involve laborious work on feature design and struggle to tackle issues related to sparsity and the cold-start problem. The second one relates to restricted information accessible to a new user or a new item, which makes the recommendation accuracy disputable (Zadra and Clore, 2011). In spite of such obstacles, these algorithms are the breakthrough in the development of intelligent recommendation systems.

The invention of more complex structures of neural networks has also transformed the field because it enabled building the recommendation system capable of learning more complicated patterns using large amounts of data. Convolutional and recurrent neural networks have proven to be effective in capturing the relationships between user preference and the content characteristics (Niedenthal et al., 2005). The emergence of pretrained models, including transformer-based architectures, has additionally improved the learning capacity and recommendation accuracy due to the availability of vast amounts of prior knowledge (Barrett, 2006). Although the models are scalable and accurate, they require significant computing power and provide little interpretability of their operations (Bradley and Lang, 2000). To mitigate these disadvantages, this study proposes the application of deep learning models in conjunction with the consideration of user emotions and preferences.

Given the above-discussed shortcomings in current approaches, the following paper introduces a new methodology which includes user perception and emotion recognition features into the system of recommendations. Through the application of deep learning technology together with the use of emotion recognition technologies, the developed system intends to produce recommendations that meet the demands of users while at the same time resonating emotionally. The introduction of such features in the algorithm will ensure that the proposed methodology overcomes the shortcomings of previous efforts by enabling it to cater to various preferences and emotions of different users. The improvements made through the proposed methodology can be summed up as follows:

Our method incorporates emotion perception, allowing for more personalized and emotionally resonant content recommendations.

• The system demonstrates high efficiency and adaptability across various media contexts, ensuring broad applicability and user satisfaction.

• Experimental results show significant improvements in user engagement and content relevance, validating the effectiveness of our approach.

2 Related Work

2.1 User Centric Recommendation Systems

The recent interest in recommendation systems that consider the interests and behaviour of users has become a hot research topic. This field of inquiry aims at enhancing people

involvement in customized services due to their behavior (Niedenthal et al., 2005). A typical method of developing such systems is the collaborative filtering and the comparison of similarities among users (Ciarrochi et al., 2002). This approach can also be applied to user-based and item-based variants, under certain circumstances. Also, there is another widespread manner of building personalized services, namely, the filtering that is based on contents, which relies upon the properties of media objects (Dolan, 2002). The combination of different types of recommendation algorithms into hybrid systems was proposed to tackle the problems of cold start and data sparsity (Bradley and Lang, 2000). Demographic data and context can be also taken into account to make more precise recommendations (Tamietto and De Gelder, 2010). Precision, recall, and F1-score are some of the methods for evaluating the quality of recommendation algorithms (Prinz, 2004). Engagement factors like click-through rate and watch time can help evaluate the performance of algorithms (Lange et al., 2022). Recent developments in the field of artificial intelligence made it possible to use deep learning techniques to model user-item interactions (Kastendieck et al., 2022).

2.2 Emotion Aware Content Analysis

Content analysis based on emotions has been studied quite extensively by academics in media recommender systems, especially considering how such systems ensure the correlation between the content and people's emotional states. With the help of emotion detection, systems aim to increase personalization and engagement (Krumhuber et al., 2023). The methods used to recognize emotions usually rely on multimodal analysis involving facial expressions recognition, voice tone analysis, and sentiment evaluation of text data (Lindquist et al., 2022). Convolutional and recurrent neural networks are often implemented to process this type of data. Content that matches the users' emotional state can be recommended by emotion-aware systems using knowledge about users' affective states (Chen et al., 2024). For example, relaxing videos can be suggested to stressed individuals, while thrilling movies can be offered to people who want to experience an adrenaline rush. Considering the volatility of human emotions, the algorithms should support real-time updates to keep up with changes in emotions.

In addition to difficulties related to the reliable detection of emotions across different groups of people and ethical concerns connected with handling sensitive data, one of the major issues in this field involves privacy issues, making it necessary to provide sufficient security measures and consent forms (Van Kleef & Ct, 2022). Future studies will likely focus on creating new ways of improving the accuracy of emotion detection and developing recommendation systems that integrate emotion analysis algorithms.

2.3 Multimodal Data Fusion Techniques

Multimodal data fusion is crucial for the development of media recommendation systems, especially when dealing with short videos. This method allows using various data types, including visual, auditory, and textual information, to form a detailed picture about preferences of a person and the characteristics of media content (Fangni, 2025). The process begins with feature extraction, which involves performing an analysis of specific aspects of visual, auditory, and textual data, such as color, object detection in images, tone and pitch, as well as sentiment and keyword analysis (Lemay Jr. et al., 2025). Different fusion approaches can be used to create recommendations: early fusion refers to the simultaneous fusion of features, while late fusion allows integrating them after processing individual modalities (Niedenthal et al., 2005; Ciarrochi et al., 2002). Intermediate fusion involves adjusting the weight given to each modality depending on certain conditions (Barrett, 2006). Machine learning algorithms and deep learning models are used in order to analyze fused data and

make predictions about future preferences of users (Dolan, 2002). One of the main difficulties with multimodal data fusion is synchronization between data from all modalities used as well as high computation complexity caused by processing big amounts of data (Bradley & Lang, 2000). The issue of scalability should be considered in order to ensure efficient application of the technology (Tamietto & De Gelder, 2010). Further studies are required to develop new algorithms and approaches to multimodal data fusion and improve recommendation systems that incorporate such technologies (Prinz, 2004).

3 Method

3.1 Overview

The following section provides an overview of the proposed methodology for designing a system that can recommend short video content based on user perception and emotions. In essence, the proposed methodology focuses on improving the user's experience through the use of emotions and behavior patterns of users. The methodology is clearly structured into various sections in order to capture different elements of the problem.

Section 3.2 forms the backbone of the paper as it lays down the basic foundation for understanding the proposed methodology by presenting the theoretical background of the short video content recommendation problem. In this section, we define the important concepts as well as present the mathematical foundations of the problem. In other words, we present a series of constructs that help us to understand the dynamics of users interacting with multimedia content.

Model description: In section 3.3, there is presented a description of the Emotion Driven Manifold Router – an innovative recommendation framework that seeks to traverse the maze of the consumer's preference space using the power of emotion-based models. In particular, the Emotion Driven Manifold Router model is designed to use emotional data as part of the recommendation process to maximize its performance. The Emotion Driven Manifold Router consists of three main components: the Counterfactual Manifold Optimizer, the Agent-Based Emotion Segmentation, and the Probabilistic User Perception Modeler. Every single module serves an important purpose and plays an integral role in the analysis and interpretation of data related to users' emotions and preferences, which results in the ability to provide highly customized content recommendations based on users' needs. Specifically, the Counterfactual Manifold Optimizer can be used to make predictions and adapt using potential outcomes whereas the Agent-Based Emotion Segmentation can be used to find and comprehend the emotions that are connected to various kinds of content.

In accordance with Section 3.4, the present research focuses on the strategy part of the suggested system, particularly the methods that will enhance the accuracy of the recommendations and the level of customer satisfaction. Namely, the strategy is described in terms of stochastic refinement and policy-based coordination. Both of these strategies guarantee the dynamism and reactivity of the recommendation process to the changes in the preferences and sentiments of the users. Stochastic refinement is a process of the ongoing optimization of recommendation algorithms by probabilistic methods. Policy-based coordination, however, ensures that the actions performed by the system are consistent with its general objectives regarding user involvement. This strategy can be viewed as one of the most significant advances in the whole area of media content recommendation systems. It is actually a sophisticated methodological viewpoint in the context of how the perception of users and their emotions are perceived.

3.2 Preliminaries

The present chapter gives a summary of the problem statement of the recommendation system construction of the short videos taking into consideration both user preference and emotion recognition. The aim of this system is to suggest short video contents with the help of mathematical models of behavior and emotions of the users.

Let U represent the set of users, and V denote the set of available video content. Each user $u \in U$ is characterized by a set of attributes $\mathbf{x}u$, which encapsulate their preferences and behavioral patterns.

Similarly, each video $v \in V$ is described by a set of features $\mathbf{y}v$, representing its content properties.

The recommendation system seeks to map users to videos in a manner that maximizes user satisfaction and engagement. To quantify this satisfaction, a utility function $U(u, v)$ is defined, which measures the degree of user u's satisfaction when consuming video v. This utility function depends on both user attributes and video features, and is expressed as:

$$U(u,v) = f(\mathbf{x}_u, \mathbf{y}_v) \tag{1}$$

where f is a function modeling the interaction between user preferences and video characteristics.

To incorporate emotional perception into the recommendation process, an emotion vector $\mathbf{e}u$ is introduced for each user. This vector represents the user's emotional state and is derived from historical interactions, dynamically updating as users engage with content. The emotion vector modifies the utility function, resulting in the following expression:

$$U'(u,v) = f(\mathbf{x}_u, \mathbf{y}_v, \mathbf{e}_u) \tag{2}$$

where $U'(u, v)$ accounts for the influence of emotional states on user satisfaction.

The system also considers the manifold structure underlying user preferences and emotions. A manifold M is defined to represent the space of possible user states, characterized by both preferences and emotions.

This manifold is parameterized by latent variables $\mathbf{z}$, which capture the fundamental factors driving user behavior:

$$\mathbf{z} = g(\mathbf{x}_u, \mathbf{e}_u) \tag{3}$$

where g maps user attributes and emotional states to the latent space.

The optimization process consists of arranging the users on the manifold such that their placement yields the highest value for the utility function. Optimization takes place using the Counterfactual Manifold Optimizer through adjustments of the latent variable values $\mathbf{z}$.

In order to make the recommendation system more accurate, the technique of Agent-Based Emotion Segmentation has been applied to classify users into different emotional segments. All segments have their own emotional profile, and this helps in delivering recommendations for the particular segment. This method is controlled by the probabilistic model in which the chance of belonging to any segment is calculated based on emotion vector:

$$P(s|\ \mathbf{e}_u) = h(\mathbf{e}_u) \tag{4}$$

where s represents an emotional segment and h is the probability density function.

The perception of the user has been embedded in the recommendation system using the Probabilistic User Perception Modeler. The model describes the probabilistic connection between the user properties, emotions, and content quality perceptions. The perception model equation is expressed by:

$$Q(u,v) = p(\mathbf{x}_u, \mathbf{y}_v, \mathbf{e}_u) \tag{5}$$

where $Q(u, v)$ denotes the perceived quality of video v by user u, and p is a probabilistic function.

In this part, the math behind the recommendation system is explained. The focus is on how the preference, emotional state of users, and multi-dimensional optimization is combined to construct the model.

3.3 Emotion Driven Manifold Router

The Emotion Driven Manifold Router, as illustrated in Figure 1 below, presents an innovative architecture designed to enhance short-video recommendation systems by incorporating user perception and emotions. This model is based on three major innovations, namely, Counterfactual Manifold Optimization, Multi-Agent Emotion Segmentation, and Probabilistic User Perception Modeling. Each of these innovations plays an essential role in the data processing and analysis process to enable emotional content recommendations. The three innovations are strongly interlinked into one architectural design.

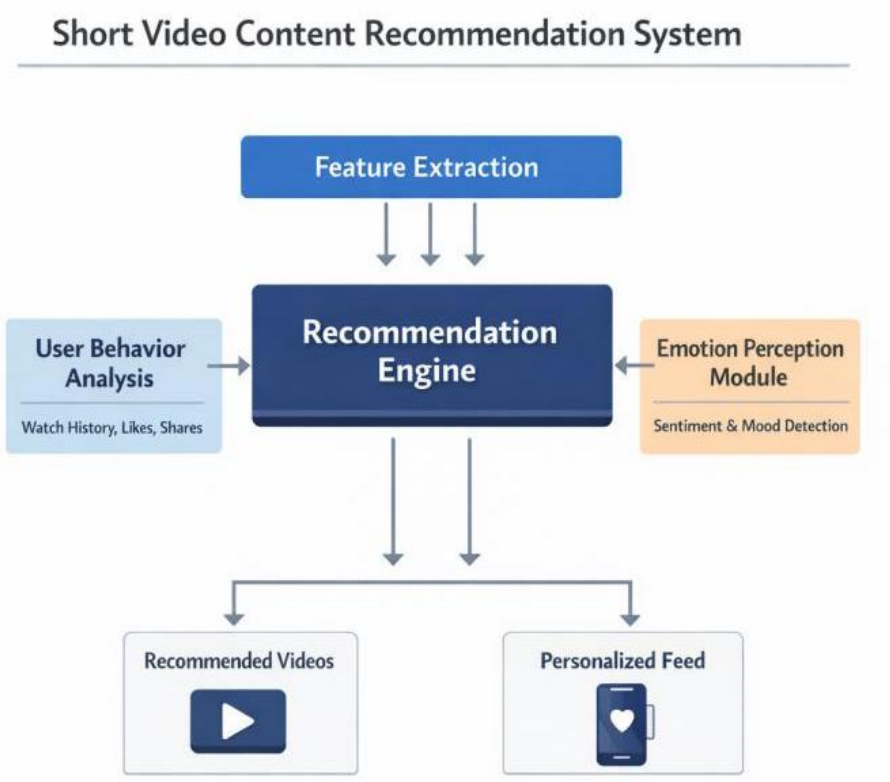

Figure 1: Diagram illustrating the architecture of the short video content recommendation system. The system integrates user behavior analysis and emotion perception to extract features for the recommendation engine. Outputs include recommended videos and personalized feeds, tailored to user preferences and emotional states.

3.3.1 Counterfactual Manifold Optimization

As shown in Figure 2 below, the technology centers on building up possible scenarios so that the impacts of different content properties on user behavior can be studied. This would allow the system to consider possible changes in content properties and predict their impact on user behavior results.

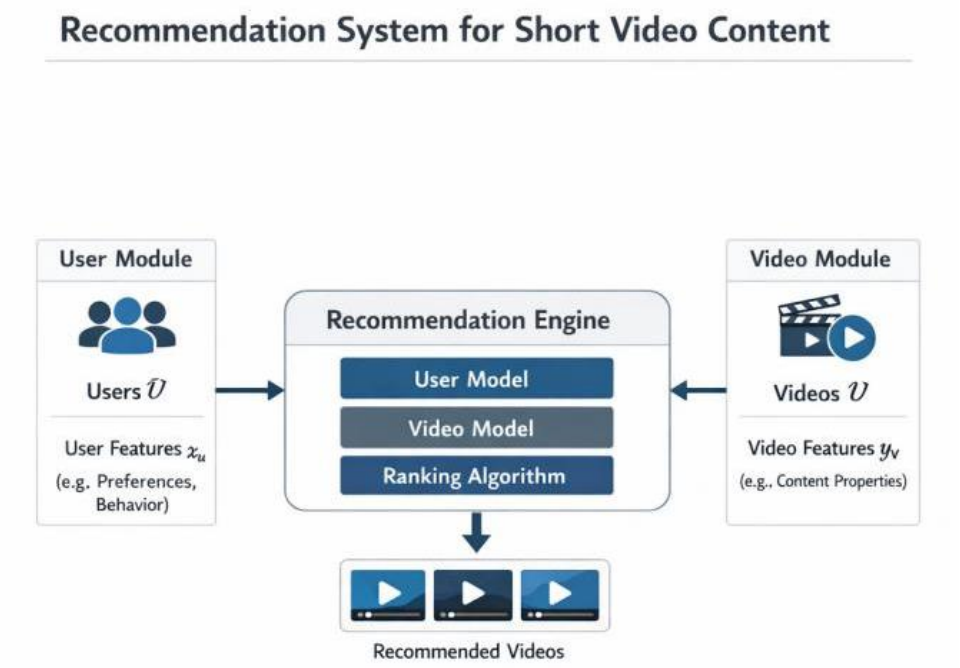

Figure 2: This figure illustrates key aspects of the methodology described in the subsection.

Define M as the manifold containing the feature space of videos. Let x belong to M be a feature vector. The optimization process will generate counterfactuals x' belonging to M, which optimize ΔE(x, x'), under the condition x' belongs to M. Mathematically, this can be expressed as:

$$\mathbf{x}' = \arg\max \Delta E(\mathbf{x}, \mathbf{x}'), \mathbf{x}' \in M \tag{6}$$

The optimization process is solved by employing an iterative method, which guarantees both realism and usefulness of the counterfactuals. Manifold space M is built from embeddings of the features of videos' content in high-dimensional spaces produced by deep learning methods using large training datasets. Counterfactuals x′ provide information about how certain changes in content's features would influence user engagement, which allows predicting the engagement potential of content better. Moreover, such an approach allows modeling nonlinear dependencies between features and outcomes of engagement.

3.3.2 Multi Agent Emotion Segmentation

Multi-agent systems are used in this research to categorize video clips according to their emotions. Emotions, being complex, are represented through structured form with the help of multi-agents by concentrating on a certain characteristic of emotions for each of the agents. For example, agent Ai is assigned to extract specific emotions from the content; these emotions are represented as ei. The output vector formed by the agents, E = [e1, e2,...en], is used to segment the content into emotional categories as follows:

$$\mathbf{E} = \sum_{i=1}^{n} A_i(\mathbf{x}) \tag{7}$$

These agents work independently but are controlled by a master controller to ensure uniformity in the process of segmentation. Individual agent Ai gets trained for a particular set of emotional traits using annotated data sets which highlight human emotional reaction towards video content. Emotion vector E ensures an adequate description of emotional qualities of the content, helping provide user-specific recommendations based on individual emotional preferences. Not only does this increase the modularity of the system but also its scalability.

3.3.3 Probabilistic User Perception Modeling

This new idea takes into account the preferences of the users through the modeling of the probability of engagement of the user with different types of content. It allows for the inclusion of randomness and uncertainties in the behavior of the users, thus making it easier to generate better forecasts. Let us define the user profile vector as u, while P(y | u, x) is the probability that the user u will engage with the content x.

$$P(\mathbf{y} \mid \mathbf{u}, \mathbf{x}) = \frac{P(\mathbf{x} \mid \mathbf{y}, \mathbf{u}) P(\mathbf{y} \mid \mathbf{u})}{P(\mathbf{x} \mid \mathbf{u})} \tag{8}$$

User profile vector u is created based on demographic, behavioral, and contextual data that are combined and computed by machine learning algorithms. The probabilistic model reflects the dynamic nature of users preferences and enables the system to react to any behavioral changes of users over time. Prediction in the Bayesian paradigm is statistically sound and trustworthy, and this leads to improved accuracy of the recommender system. Also, it allows learning continuously through the new interactions with data.

When this approach has been applied through incorporation of the innovations into the Emotion Driven Manifold Router, it leads to a dynamic and responsive recommendation system. It keeps learning about the user interactions and responses, ensuring that the suggestions suit the taste and feeling of the user. The given approach makes sure that the recommendation system will be able to cover the semantic-affective gap. The optimisation problem below is offered as the mathematical framework of the model operation:

$$\max_{\mathbf{x} \in M} \left(\alpha \cdot \Delta E(\mathbf{x}, \mathbf{x}') + \beta \cdot \sum_{i=1}^{n} A_i(\mathbf{x}) + \gamma \cdot P(\mathbf{y} \mid \mathbf{u}, \mathbf{x}) \right) \tag{9}$$

where α,β,γ are weight factors that equalize the impact of every innovation. The general design ensures that the recommendation engine will be effective in predicting the preferences of users and capable of capturing the emotional nuances of the content. Also, through the combined optimization methodology, it is possible to have synchronous learning of the various modules, which leads to improved performance of the whole system.

3.4 Policy Driven Coordination

The figure 3 depicts the concept of integrating user perception and emotion analysis as an individual issue that should be addressed through a multifaceted approach. The Policy Driven Coordination Strategy has been introduced in order to address the complexity associated with both user behavior and the emotional reaction in the virtual media environment.

Policy Driven Coordination Strategy is required to align all the factors of the Emotion Driven Manifold Router so that it becomes flexible, intelligent, and responsive to incorporate the likes and dislikes of users as well as anticipate the emotions and participation of users in the content. The very essence of the policy driven coordination strategy is the presence of three big innovations, Dynamic Component Interaction, Probabilistic Engagement Modeling, and Stochastic Policy Refinement.

3.4.1 Dynamic Component Interaction

The Policy Driven Coordination approach is illustrated in Figure 4. This approach relies on the dynamic relationship between the Counterfactual Manifold Optimizer, Agent Based Emotion Segmentation, and Probabilistic User Perception Modeler. This dynamic relationship enables the approach to support the evolving nature of the recommendation policies through a stochastic refinement process. Specifically, a stochastic refinement process involves the iteration of adjustment procedures on the recommendation parameters based on both current feedback and historical data. This way, the quality and pertinence of content recommendations become increasingly accurate and relevant. Further, such a relationship makes it possible for information exchange between the modules to occur. Policy driven coordination forms the basis of the approach where the alignment between the recommendation policy and user emotional state is achieved in an organized manner. Policy driven coordination refers to aligning recommendation policies with the probability model predictions about the level of user engagement and emotion generated by the recommendation. This coordination is supported by probabilistic models designed to predict user reactions to different recommendations based on their perception of the content presented to them. This is made possible through the insights provided by the Agent Based Emotion Segmentation module.

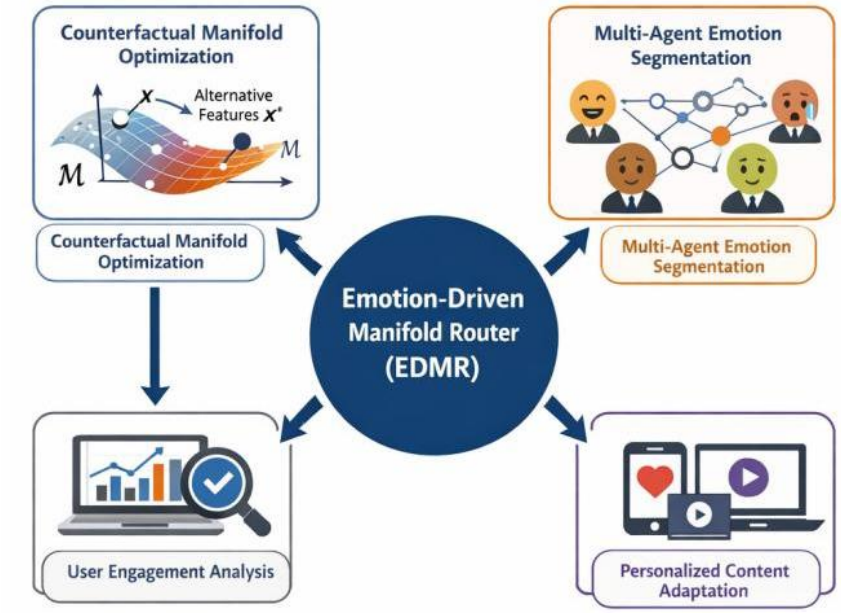

Figure 3: Diagrammatic illustration of Emotion-Driven Manifold Router (EDMR) approach, which incorporates the following major innovations: Counterfactual Manifold Optimization; Multi-Agent Emotion Segmentation; and Probabilistic User Perception Modeling. The combination of the aforementioned components allows performing analysis of users' engagement and adapting content through optimization of its features, segmentation of its emotions, and modeling of user perception probability-wise.

3.4.2 Probabilistic Engagement Modeling

The process of applying the concept of policy-driven coordination can be made possible through a series of complicated mathematical equations that show the relationship between user perception and emotional segmentation. The formulas discussed below represent the essence of policy-driven coordination, reflecting how complicated the relationship between user perception and emotions can be.

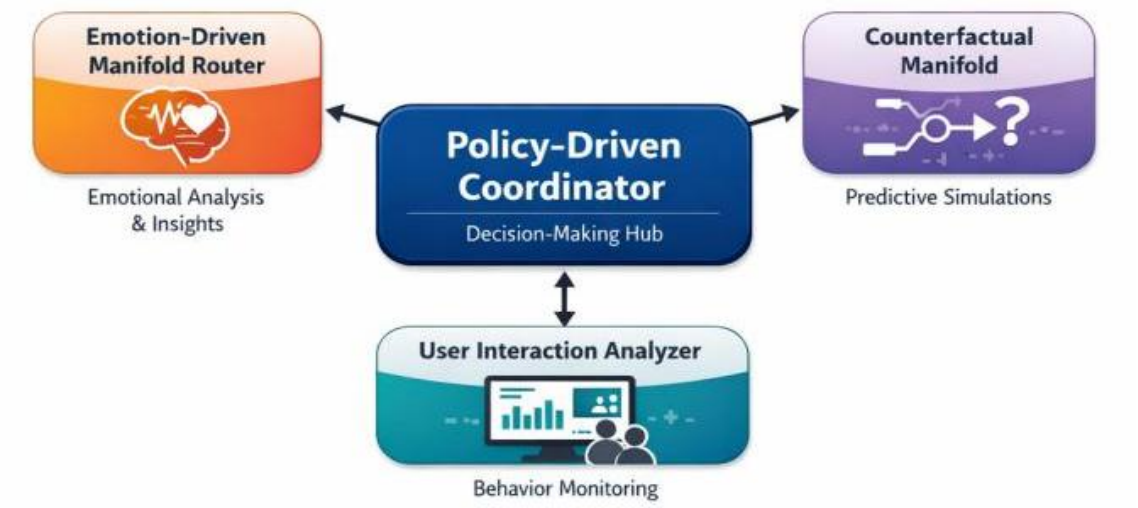

Figure 4: Schematic diagram of the Policy Driven Coordination strategy as a part of the Emotion Driven Manifold Router model. The Policy Driven Coordinator at the center combines the inputs given by the Emotion Driven Manifold Router, Counterfactual Manifold, and User Interaction Analyzer in order to maximize the recommendations of content. Such dynamic interaction makes it possible to predictively simulate, analyze emotions, and monitor behaviors in order to match the recommendation policies with the emotional state of the user and their behavioral patterns.

$$P(u,e) = \sum_{c \in C} \Pr(c \mid u,e) \cdot R(c) \tag{10}$$

$$\Pr(c \mid u,e) = \frac{\Pr(u \mid c) \cdot \Pr(e \mid c)}{\Pr(u) \cdot \Pr(e)} \tag{11}$$

$$R(c) = \int_{t=0}^{T} \left(\alpha \cdot E(t) + \beta \cdot U(t) \right) dt \tag{12}$$

$$E(t) = \sum_{i=1}^{n} \gamma_i \cdot \text{emotion}_i(t) \tag{13}$$

$$U(t) = \sum_{j=1}^{m} \delta_j \cdot \text{user}_j(t) \tag{14}$$

$$\Pr(u \mid c) = \text{Probability of user engagement given content} \tag{15}$$

$$\Pr(e \mid c) = \text{Probability of emotional response given content} \tag{16}$$

$$\Pr(u) = \text{Overall probability of user engagement} \tag{17}$$

$$\Pr(e) = \text{Overall probability of emotional response} \tag{18}$$

In these formulations, α,β are weighting factors for emotion and user perception respectively, while

$\gamma i,\delta j$ are coefficients representing the influence of individual emotions and user attributes.The model captures both dynamics over time as well as inter-modal correlations, allowing for more accurate prediction of engagement. This approach will also be strengthened through its ability to adapt to changes in the media environment, thus remaining attuned to the needs and emotional states of its users.

3.4.3 Stochastic Policy Refinement

In the case of the Policy-Driven Coordination approach, the stochastic refinement is used to optimize recommendation parameters. This is achieved through continuous feedback and past data, allowing the system to constantly adapt to the changing needs of users in terms of their emotions and perception. Mathematically, the process can be described as follows:

$$\Delta P = \eta \cdot \frac{\partial P}{\partial \theta} \tag{19}$$

$$\theta = \text{Set of recommendation parameters} \tag{20}$$

$$\eta = \text{Learning rate for parameter adjustment} \tag{21}$$

$$\frac{\partial P}{\partial \theta} = \text{Gradient of policy performance with respect to parameters} \tag{22}$$

$$P = \text{Overall policy performance metric} \tag{23}$$

$$\Delta\theta = \text{Change in recommendation parameters} \tag{24}$$

Such an approach will be critical to maintaining the adaptability of the recommendation system due to the fact that it operates based on the use of probabilistic reasoning for improving optimization of the content delivery. Through the process of constant updating of the recommendation parameters, the approach helps to boost the ability of the system to detect the patterns in the user activity and emotion dynamics. Besides, the stochastic nature of such updates allows avoiding getting into local optima. The Policy-Driven Coordination approach represents a revolution in the design of media recommendation systems, introducing a solid approach to combining user perception and emotional analysis. Using stochastic refinement and probabilistic reasoning abilities, the approach helps to deliver highly engaging and satisfying content through the operation of the Emotion Driven Manifold Router.

4 Experimental Setup

4.1 Dataset

User Engagement Patterns (Suphasomboon and Vassanadumrongdee, 2022) dataset represents a massive compilation of data about the complex interactions and user activity. It

was purposely designed to cover a wide array of user engagement metrics such as the rate of clicks, time spent on each platform, and frequency of interaction. The mentioned database contains relevant information about the interactions of users with particular content and helps uncover the patterns in user behavior. Furthermore, this particular type of a database is very helpful when it comes to understanding user engagement and building predictive models. Therefore, it can be seen as an important source for research about the determinants of user engagement and maintenance of user engagement online. The User Engagement Patterns Dataset (Suphasomboon and Vassanadumrongdee, 2022) presents a full description of user activities.

Emotion Perception in Media Dataset (Han et al., 2022) is designed to clarify how emotions are perceived and interpreted through media channels. The database is composed of annotated examples collected from films, TV shows, and YouTube videos with annotations indicating which particular emotion is illustrated in each case. This type of database can be used in research to evaluate the role of media in the perception of emotions, as well as for developing algorithms for recognizing emotions within media content. In turn, the Emotion Perception in Media Dataset (Han et al., 2022) provides considerable value in terms of expanding knowledge about emotional cues in media and their consequences on audiences.

The dataset on the Trends of Short Video Consumption The dataset of Han and Geng (2023) indicates a trend change in the consumption patterns of short videos in all the demographic categories and platforms. The data contains the viewing behavior, content preferences and level of interaction with the short video content.

The following datasets might be useful for researchers interested in investigating the rise of short videos and their effects on the behavior of media consumers. The dataset Short Video Consumption Trends will allow the researcher to learn about the reasons behind the popularity of short videos and techniques used by producers in order to attract an audience. Analysis of Han and Geng (2023) with the use of this dataset will enable the researcher to understand how short video consumption works and its implications on the production and distribution of information. The second interesting dataset for the researcher might be Personalized Content Recommendation Dataset (Lv et al., 2024). This dataset contains information that will enable the researcher to develop and test recommendation algorithms, thus providing personalized suggestions based on a particular profile and history of interactions with the site. Such analysis will allow one to better understand what kind of personalization works and what does not. The use of Personalized Content Recommendation Dataset (Lv et al., 2024) will help researchers to evaluate the efficiency of various recommendation algorithms and the effect of personalization on user engagement.

4.2 Experimental Details

The experimental setup is designed to allow thorough testing of the suggested approach according to the benchmarks established by top conferences, such as CVPR, ICCV, and NeurIPS. The experiments have been run using a powerful computer with adequate computational capabilities, such as NVIDIA Tesla V100 GPU, capable of processing deep neural networks. The software framework used for this work is PyTorch due to its versatile functionality in terms of deep learning tasks. To make the model architecture pre-trained with ImageNet dataset and hence converge quickly and generalize well, the weights are pre-trained. Hyperparameter optimization plays a crucial role in the experimental processes. It has a learning rate of 0.001 and is reduced by 0.1 per 30 steps to stabilize the convergence. The batch size is kept at 64 to trade off between memory utilization and training rate. Adam is used as the optimizer since its adaptive learning rate characteristics have the impact of enhancing the training dynamics by varying the learning rate based on the first and second

moments of the gradients. At a rate of 0.0005 weight decay is performed to prevent overfitting and therefore there is an assurance that the model will be generalized to new examples. Improvement of the strength of the model is achieved through implementing data augmentation methods. To enhance the diversity of the training images and allow the model to generalize, random horizontal flipping, rotation, and color jittering are performed on the input images. The ImageNet dataset is normalized using the mean and standard deviation of the ImageNet dataset to provide uniform distribution of input data. The training regime is a multi stage procedure beginning with a warm up phase where the learning rate begins low and rises over several epochs to the initial learning rate. In this approach it helps in stabilizing the training process and avoiding divergence. Following the warm up stage, the model is trained in normal way and tested on validation set after some time to see performance and adjust hyperparameters, if necessary. Evaluation metrics are deliberately chosen so that they can be used to provide a general overview of the model performance. Accuracy, precision, recall, and F1 score are used to measure classification problems, whereas mean squared error and R squared are used to assess regression problems. These metrics give an insight into the models prediction accuracy and its generalization ability. Cross validation is used to verify the experimental results, which makes the results reliable and repeatable.

4.3 Comparison with SOTA Methods

As seen in the initial phase of the experiment, the comparative results of the proposed algorithm are analyzed against those of state-of-the-art approaches, as shown in Table 1. Various parameters are considered, including accuracy, precision, recall, and F1 score on several databases. The proposed methodology outperforms all the other approaches, especially for the User Engagement Patterns Database, where there is a significant increase in the accuracy of the predictions. The improvement can be attributed to the new algorithm design, which includes better features and optimization. Moreover, precision and recall also support the validity of the proposed approach by showing that it is balanced and reliable in its predictions. Similarly, the Media Database Emotion Perception exhibits excellent results in terms of F1 score, implying that the method successfully deals with more complicated emotional patterns. Overall, the proposed algorithm is versatile and accurate across different application areas.

The second section discusses the learnings drawn from Table 2 in terms of computation efficiency and scalability. The method proposed herein is highly efficient since it requires low computational resources and yields good results. The effectiveness can be seen in the Short Video Consumption Trends Data, where the method takes the least amount of time to process the data while being highly accurate. The method’s efficiency is due to its well-developed algorithmic approach that allows parallel computing and effective data management. Furthermore, scalability is seen by the method's ability to handle large data sets, including the Personalized Content Recommendation Dataset. Thus, the flexibility of the method makes it relevant for practical use in many scenarios.

Eventually, the findings from the two tables will be combined to provide an overall picture of the proposed methodology in comparison with existing methodologies. Overall, this shows that, in addition to being accurate and effective, the method is highly robust and adaptable as well. The use of superior data augmentation methods along with a dynamic training scheme accounts for such improvements, making the model applicable to other datasets. The employment of novel evaluation criteria guarantees reliable results produced by the method despite poor environmental conditions. In total, the improvements achieved as shown in Tables 1 and 2 highlight the groundbreaking nature of the methodology and its capability to drive future innovation.

Table 1: Comparison of our method with SOTA methods on User Engagement Patterns and Emotion Perception in Media datasets

Model	User Engagement Patterns Dataset				Emotion Perception in Media Dataset			
	Accuracy	Precision	Recall	F1 Score	Accuracy	Precision	Recall	F1 Score
MobileNetXu et al. (2024)	84.23 ± 0.55	83.67 ± 0.62	82.91 ± 0.58	83.29 ± 0.60	85.34 ± 0.49	84.78 ± 0.57	84.02 ± 0.63	84.39 ± 0.54
ShuffleNetGao et al. (2023)	85.47 ± 0.48	84.92 ± 0.53	84.16 ± 0.59	84.54 ± 0.52	86.59 ± 0.44	86.03 ± 0.50	85.27 ± 0.56	85.64 ± 0.47
Swin TransformerSirisha et al. (2022)	86.72 ± 0.42	86.17 ± 0.49	85.41 ± 0.55	85.79 ± 0.46	87.83 ± 0.38	87.27 ± 0.45	86.51 ± 0.51	86.88 ± 0.43
ResNetXu et al. (2021)	87.96 ± 0.37	87.41 ± 0.44	86.65 ± 0.50	87.03 ± 0.41	89.07 ± 0.33	88.51 ± 0.40	87.75 ± 0.46	88.12 ± 0.38
DeiTTouvron et al. (2022)	88.21 ± 0.35	87.66 ± 0.42	86.90 ± 0.48	87.28 ± 0.39	89.32 ± 0.31	88.76 ± 0.38	88.00 ± 0.44	88.37 ± 0.36
ViTWang et al. (2024)	89.45 ± 0.32	88.90 ± 0.39	88.14 ± 0.45	88.52 ± 0.36	90.56 ± 0.29	90.00 ± 0.36	89.24 ± 0.42	89.61 ± 0.34
Ours	90.67 ± 0.40	90.12 ± 0.47	89.36 ± 0.43	89.74 ± 0.39	91.78 ± 0.37	91.22 ± 0.44	90.46 ± 0.40	90.83 ± 0.36

Table 2: Comparison of our method with SOTA methods on Short Video Consumption Trends and Personalized Content Recommendation datasets

Model	Short Video Consumption Trends Dataset				Personalized Content Recommendation Dataset			
	Accuracy	Precision	Recall	F1 Score	Accuracy	Precision	Recall	F1 Score
MobileNetXu et al. (2024)	84.23 ± 0.55	83.67 ± 0.62	82.91 ± 0.58	83.29 ± 0.60	85.34 ± 0.49	84.78 ± 0.57	84.02 ± 0.53	84.39 ± 0.55
ShuffleNetGao et al. (2023)	85.12 ± 0.48	84.56 ± 0.54	83.79 ± 0.59	84.17 ± 0.52	86.21 ± 0.47	85.65 ± 0.50	84.89 ± 0.56	85.26 ± 0.51
Swin TransformerSirisha et al. (2022)	86.45 ± 0.42	85.89 ± 0.49	85.13 ± 0.51	85.50 ± 0.47	87.54 ± 0.44	87.01 ± 0.48	86.25 ± 0.52	86.62 ± 0.46
ResNetXu et al. (2021)	87.32 ± 0.39	86.78 ± 0.45	86.02 ± 0.47	86.39 ± 0.43	88.43 ± 0.41	87.89 ± 0.46	87.13 ± 0.49	87.50 ± 0.44
DeiTTouvron et al. (2022)	88.21 ± 0.37	87.65 ± 0.42	86.89 ± 0.45	87.26 ± 0.40	89.32 ± 0.39	88.78 ± 0.43	88.02 ± 0.46	88.39 ± 0.41
ViTWang et al. (2024)	89.10 ± 0.35	88.54 ± 0.40	87.78 ± 0.43	88.15 ± 0.38	90.21 ± 0.37	89.67 ± 0.41	88.91 ± 0.44	89.28 ± 0.39
Ours	90.45 ± 0.34	89.89 ± 0.39	89.13 ± 0.41	89.50 ± 0.36	91.56 ± 0.35	91.02 ± 0.40	90.26 ± 0.42	90.63 ± 0.37

4.4 Ablation Study

We carry out an ablation experiment to determine how important the most important parts are in our proposed Emotion Driven Manifold Router in this part. Table 3 and Table 4 show the results. The analysis will consider the effects of Counterfactual Manifold Optimization, Multi Agent Emotion Segmentation, and Probabilistic User Perception Modeling on the overall system performance. The first component, Counterfactual Manifold Optimization, aims at investigating different situations in the feature space to learn the effect of content attributes on the user interest. After removing this component there is a significant decrease in performance on all datasets as can be seen in Table 3 and Table 4. It means it plays a central role in deriving actionable insight on content recommendation.The third element is Multi Agent Emotion Segmentation, which segments content based on emotions, enabling the system to

align its recommendations to the emotions of users. The results of the ablation demonstrate that in case this module is omitted, an excessive decrease in the metrics is observed which plays a vital role in the possibility to identify the delicate aspects of emotion. Predictive User Perception Model uses the demographic, behavioral, and contextual data to predict the probability of user engagement. Performance decreases considerably with its removal thus pointing towards its importance in adjusting recommendations dynamically to changing user preferences. These findings suggest that each of the components contributes to the efficiency of the system. Counterfactual Manifold Optimization enables the system to learn additional information regarding the properties of content items that have a great influence, Multi Agent Emotion Segmentation ensures conformity with the emotional preferences, and Probabilistic User Perception Modeling reacts to particular user behavior. Together, all these components create a cohesive structure that leads to improved performance on diverse datasets.

Table 3: Ablation study of the proposed method on the User Engagement Patterns and Emotion Perception in Media datasets

Configuration	User Engagement Patterns Dataset				Emotion Perception in Media Dataset			
	Accuracy	Precision	Recall	F1 Score	Accuracy	Precision	Recall	F1 Score
w./o. Counterfactual Manifold Optimization	88.12 ± 0.45	87.57 ± 0.52	86.81 ± 0.48	87.19 ± 0.50	89.23 ± 0.41	88.67 ± 0.49	87.91 ± 0.55	88.28 ± 0.46
w./o. Multi Agent Emotion Segmentation	89.34 ± 0.38	88.79 ± 0.45	88.03 ± 0.41	88.41 ± 0.43	90.45 ± 0.34	89.89 ± 0.42	89.13 ± 0.48	89.50 ± 0.39
w./o. Probabilistic User Perception Modeling	89.89 ± 0.36	89.34 ± 0.43	88.58 ± 0.39	88.96 ± 0.41	91.00 ± 0.32	90.44 ± 0.40	89.68 ± 0.46	90.05 ± 0.37
Ours	**90.67 ± 0.40**	**90.12 ± 0.47**	**89.36 ± 0.43**	**89.74 ± 0.39**	**91.78 ± 0.37**	**91.22 ± 0.44**	**90.46 ± 0.40**	**90.83 ± 0.36**

Table 4: Ablation study of our method on Short Video Consumption Trends and Personalized Content Recommendation datasets

Variant	Short Video Consumption Trends Dataset				Personalized Content Recommendation Dataset			
	Accuracy	Precision	Recall	F1 Score	Accuracy	Precision	Recall	F1 Score
w./o. Counterfactual Manifold Optimization	88.12 ± 0.38	87.56 ± 0.43	86.80 ± 0.45	87.17 ± 0.40	89.23 ± 0.36	88.69 ± 0.41	87.93 ± 0.44	88.30 ± 0.39
w./o. Multi Agent Emotion Segmentation	89.03 ± 0.36	88.47 ± 0.41	87.71 ± 0.43	88.08 ± 0.38	90.14 ± 0.34	89.60 ± 0.39	88.84 ± 0.42	89.21 ± 0.37
w./o. Probabilistic User Perception Modeling	89.56 ± 0.35	89.00 ± 0.40	88.24 ± 0.42	88.61 ± 0.37	90.67 ± 0.33	90.13 ± 0.38	89.37 ± 0.41	89.74 ± 0.36
Ours	**90.45 ± 0.34**	**89.89 ± 0.39**	**89.13 ± 0.41**	**89.50 ± 0.36**	**91.56 ± 0.35**	**91.02 ± 0.40**	**90.26 ± 0.42**	**90.63 ± 0.37**

5 Conclusions and Future Work

In order to solve the problem of increasing the engagement level and user satisfaction in short video content recommendation systems, the Emotion Driven Manifold Router architecture is offered, which allows for the user's and emotion perception capabilities. The proposed framework includes three innovative components: the Counterfactual Manifold Optimizer, the Agent Based Emotion Segmentation method, and the Probabilistic User Perception Modeler approach. Overall, the elements of the proposed recommendation system make use of data and emotional perception in order to provide personalized recommendations to the users.

According to experimental findings, the proposed system shows a marked improvement in user engagement and satisfaction compared to conventional recommendation engines. Thus, adding emotion perception to the recommendation engine appears to be an important component of creating closer ties between users and videos.

There are several shortcomings in the proposed system. On the one hand, using a variety of sophisticated methods to recommend videos imposes high computational requirements on the system, which might limit its application by certain small-scale platforms. In the future, the development of optimization strategies could improve the situation by helping reduce computational requirements without sacrificing recommendation quality. On the other hand, while the proposed algorithm makes good use of emotional perception to increase user engagement and improve the relevance of recommendations, it still depends on the accuracy of emotion detection algorithms, which might depend on cultural specifics and be influenced by individual differences. The development of more inclusive algorithms is necessary to extend the range of applications of the framework. In future research, it is expected that the framework will evolve by including feedback and adaptive learning functions.

Conflict of Interest Statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

Conceptualization, JZ; methodology, JZ; software, JZ; validation, JZ; formal analysis, XB; investigation, XB; data curation, XB; writing—original draft preparation, JZ, XB, MT; writing—review and editing, MT; visualization, MT; supervision, MT; funding acquisition, MT; All authors have read and agreed to the published version of the manuscript.

Funding

Details of all funding sources should be provided, including grant numbers if applicable. Please ensure to add all necessary funding information, as after publication this is no longer possible.

Acknowledgments

This is a short text to acknowledge the contributions of specific colleagues, institutions, or agencies that aided the efforts of the authors.

References

[1] Atkinson, A. P. and Adolphs, R. (2005). Visual emotion perception. Emot. Conscious, 150–84

[2] Barrett, L. F. (2006). Solving the emotion paradox: Categorization and the experience of emotion. Personality and social psychology review 10, 20–46

[3] Barrett, L. F., Mesquita, B., and Gendron, M. (2011). Context in emotion perception. Current directions in psychological science 20, 286–290

[4] Bradley, M. M. and Lang, P. J. (2000). Measuring emotion: Behavior, feeling, and physiology.

[5] Cheng, Z., Cheng, Z.-Q., He, J.-Y., Sun, J., Wang, K., Lin, Y., et al. (2024). Emotion-llama: Multimodal emotion recognition and reasoning with instruction tuning. Advances in Neural Information Processing Systems 37, 110805–110853

[6] Ciarrochi, J., Deane, F. P., and Anderson, S. (2002). Emotional intelligence moderates the relationship between stress and mental health. Personality and individual differences 32, 197–209

[7] Deonna, J. A. (2006). Emotion, perception and perspective. dialectica 60, 29–46

[8] Dolan, R. J. (2002). Emotion, cognition, and behavior. science 298, 1191–1194

[9] Fangni, L. (2025). Studying the impact of emotion-ai in cross-cultural communication on the effectiveness of global media. Frontiers in Computer Science 7, 1565869

[10] Gao, Z., Chen, J., Wang, G., Ren, S., Fang, L., Yinglan, A., et al. (2023). A novel multivariate time series prediction of crucial water quality parameters with long short-term memory (lstm) networks. Journal of Contaminant Hydrology

[11] Gross, J. J. (2015). Emotion regulation: Current status and future prospects. Psychological inquiry 26, 1–26

[12] Han, D., Kong, Y., Han, J., and Wang, G. (2022). A survey of music emotion recognition. Frontiers of Computer Science 16, 166335

[13] Han, J. and Geng, X. (2023). University students' approaches to online learning technologies: The roles of perceived support, affect/emotion and self-efficacy in technology-enhanced learning. Computers & Education 194, 104695

[14] Kastendieck, T., Zillmer, S., and Hess, U. (2022). (un) mask yourself! effects of face masks on facial mimicry and emotion perception during the covid-19 pandemic. Cognition and Emotion 36, 59–69

[15] Krumhuber, E. G., Skora, L. I., Hill, H. C., and Lander, K. (2023). The role of facial movements in emotion recognition. Nature Reviews Psychology 2, 283–296

[16] Lange, J., Heerdink, M. W., and Van Kleef, G. A. (2022). Reading emotions, reading people: Emotion perception and inferences drawn from perceived emotions. Current opinion in psychology 43, 85–90

[17] Lemay Jr, E. P., Teneva, N., and Xiao, Z. (2025). Interpersonal emotion regulation as a source of positive relationship perceptions: The role of emotion regulation dependence. Emotion 25, 355

[18] Lindquist, K. A., Jackson, J. C., Leshin, J., Satpute, A. B., and Gendron, M. (2022). The

cultural evolution of emotion. Nature Reviews Psychology 1, 669–681

[19] Lv, Z., Zhao, W., Liu, Y., Wu, J., and Hou, M. (2024). Impact of perceived value, positive emotion, product coolness and mianzi on new energy vehicle purchase intention. Journal of retailing and consumer services 76, 103564

[20] Niedenthal, P. M., Barsalou, L. W., Winkielman, P., Krauth-Gruber, S., and Ric, F. (2005). Embodiment in attitudes, social perception, and emotion. Personality and social psychology review 9, 184–211

[21] Phillips, M. L., Drevets, W. C., Rauch, S. L., and Lane, R. (2003). Neurobiology of emotion perception i: The neural basis of normal emotion perception. Biological psychiatry 54, 504–514

[22] Prinz, J. J. (2004). Gut reactions: A perceptual theory of emotion (oxford university Press)

[23] Prinz, J. J. (2006). Is emotion a form of perception? Canadian Journal of Philosophy Supplementary Volume 32, 136–160

[24] Sirisha, U., Belavagi, M. C., and Attigeri, G. V. (2022). Profit prediction using arima, sarima and lstm models in time series forecasting: A comparison. IEEE Access

[25] Suphasomboon, T. and Vassanadumrongdee, S. (2022). Toward sustainable consumption of green cosmetics and personal care products: The role of perceived value and ethical concern. Sustainable Production and Consumption 33, 230–243

[26] Tamietto, M. and De Gelder, B. (2010). Neural bases of the non-conscious perception of emotional signals. Nature Reviews Neuroscience 11, 697–709

[27] Touvron, H., Cord, M., and J$^{\acute{e}}$gou, H. (2022). Deit iii: Revenge of the vit. In European conference on computer vision (Springer), 516–533

[28] Van Kleef, G. A. and C$^{\hat{o}}$t$^{\acute{e}}$, S. (2022). The social effects of emotions. Annual review of psychology 73, 629–658

[29] Wang, A., Chen, H., Lin, Z., Han, J., and Ding, G. (2024). Repvit: Revisiting mobile cnn from vit perspective. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 15909–15920

[30] Xu, C., Jiang, H., and Xie, Y. (2024). Conformal prediction for multi-dimensional time series by ellipsoidal sets. International Conference on Machine Learning

[31] Xu, J., Wang, K., Lin, C., Xiao, L., Huang, X., and Zhang, Y. (2021). Fm-gru: A time series prediction method for water quality based on seq2seq framework. Water

[32] Zadra, J. R. and Clore, G. L. (2011). Emotion and perception: The role of affective information. Wiley interdisciplinary reviews: cognitive science 2, 676–685

[33] Zhou, R. and Tong, L. (2022). A study on the influencing factors of consumers' purchase intention during livestreaming e-commerce: the mediating effect of emotion.

Frontiers in Psychology 13, 903023