Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Issue 1, Januari 2026
DOI : https://doi.org/10.33395/sinkron.v10i1.15579
e-ISSN : 2541-2019
p-ISSN : 2541-044X

A Hybrid YOLOv11 and LightFM Model for Emotion-Driven Anime Recommendation

Kafka Ramadityo[1)*], Ida Nurhaida[2)]
[1,2)]Department of Informatics, Universitas Pembangunan Jaya, Tangerang Selatan, Indonesia
[2)]Center of Urban Studies, Universitas Pembangunan Jaya, Tangerang Selatan, Indonesia
[1)]kafka.ramadityo@student.upj.ac.id, [2)]ida.nurhaida@upj.ac.id

Submitted :Nov 13, 2025 | **Accepted** : Dec 2, 2025 | **Published** : Jan 02, 2026

Abstract: Existing anime recommendation systems focus on genre preferences and viewing history without considering users' emotional states, leading to context-blind recommendations that may exacerbate negative moods and reduce satisfaction. Most existing systems employ outdated architectures with limited accuracy and lack diversification mechanisms to prevent filter bubbles. This study develops an emotion-based anime recommendation system integrating YOLOv11 for facial emotion recognition with hybrid collaborative filtering using LightFM and Maximum Marginal Relevance diversification. The primary novelty lies in seamlessly combining YOLOv11's superior emotion recognition, LightFM's hybrid matrix factorization for cold-start mitigation, and MMR diversification for preventing filter bubbles while maintaining emotional congruence. The methodology employed the KDEF dataset (3,597 images, five emotion classes) for training YOLOv11 with data augmentation, and the MyAnimeList dataset (744,330 interactions) for recommendation modeling. Emotion-to-genre mappings informed by survey data from 51 participants were implemented with MMR diversification to balance relevance and variety. The YOLOv11 model achieved 93.70% validation accuracy, outperforming CNN-LSTM approaches by 37.55 percentage points. The hybrid recommendation model demonstrated test AUC of 0.8567 and Precision@10 of 0.1457, representing 417% improvement over pure collaborative filtering, while diversification increased genre representation by 20.9% with minimal precision loss. This system demonstrates real-time applicability for streaming platforms through camera-based emotion capture and immediate recommendation generation, enhancing user engagement and emotional well-being. The integration represents a significant advancement toward affective computing in entertainment media.

Keywords: Anime, Collaborative Filtering, Emotion Recognition, Hybrid Recommendation, LightFM, Maximum Marginal Relevance, Recommendation System, YOLOv11

INTRODUCTION

Emotions are brief yet intense psychophysiological responses triggered by meaningful environmental stimuli (Hartmann, 2024). Psychologists categorize basic emotions into positive and negative valence, encompassing pleasant states such as happiness and satisfaction, or unpleasant feelings including sadness, anger, and disgust. These emotional categories are intrinsically linked to the thematic and narrative structures of various anime genres. The global anime industry has experienced remarkable growth, projected to reach USD 34.3 billion in 2024 with a compound annual growth rate of 9.8% through 2030 (Grand View Research, 2025). As anime content proliferates, viewers increasingly face difficulty selecting titles aligned with their emotional needs. The emotional content embedded in anime can significantly influence viewers' moods, potentially providing psychological comfort or emotional distress (Zafira et al., 2024; Tan & Chung, 2023).

A recommendation system is an information filtering system designed to predict and suggest content that may interest users (Fadhel et al., 2025). However, challenges persist in machine learning-based recommendation systems, including data scarcity, noise, and cold-start problems that impede accurate preference determination (Wu, 2022). More critically, most existing anime recommendation systems focus on genre preferences, ratings, or viewing history, without considering users' emotional states, which significantly influence viewing satisfaction (Kim et al., 2021). This context-blindness represents a fundamental limitation.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Issue 1, Januari 2026
DOI : https://doi.org/10.33395/sinkron.v10i1.15579
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Recommendation algorithms lack awareness of users' psychological conditions, which profoundly shape media consumption preferences. For instance, comedy anime may suit individuals experiencing stress, whereas those seeking cognitive engagement gravitate toward mystery series. Recommending anime misaligned with users' emotional states may exacerbate negative moods and reduce satisfaction. This emotional mismatch affects user experience and psychological well-being, particularly when users turn to media for emotional regulation. Therefore, developing emotion-aware recommendation systems is essential to enhance viewing comfort and support emotional well-being (Ruan et al., 2025), serving as affective computing that personalizes content delivery based on real-time emotional needs.

Existing emotion-based recommendation research predominantly focuses on movies and music (Balfaqih, 2023; Wang & Zhao, 2022), with limited investigation into anime, a medium characterized by culturally distinct emotional cues. While studies have explored emotion recognition from facial expressions and hybrid recommendation approaches, few have integrated real-time emotion detection with collaborative filtering specifically for anime content. Furthermore, most existing systems employ outdated architectures achieving 60-72% accuracy (Balfaqih, 2023; Wang & Zhao, 2022) and lack diversification mechanisms to prevent filter bubbles, representing a significant gap given anime's unique emotional landscape and growing global consumption.

Psychological research demonstrates that affective states systematically influence content preferences through temporal dependencies and differential appraisals, with neuroimaging evidence revealing distinct emotion-regulation strategies across genre preferences (Gong & Huskey, 2023; Stavraki et al., 2021; Zwiky et al., 2024). Recent advances in vision-based emotion detection have made real-time facial expression analysis computationally feasible. However, translating emotion recognition into effective recommendation requires integration with sophisticated personalization mechanisms that account for cold-start problems and the tendency of traditional collaborative filtering to create filter bubbles. These technical challenges necessitate hybrid approaches combining emotion-aware signals with collaborative and content-based filtering to address personalization quality and content diversity (Patoulia et al., 2023).

This study addresses these limitations through four specific contributions: (1) implementing YOLOv11 to achieve superior emotion recognition accuracy exceeding existing CNN-LSTM benchmarks; (2) developing a hybrid recommendation framework integrating emotion detection with LightFM's matrix factorization to address personalization and cold-start challenges; (3) grounding emotion-to-genre mappings in neuropsychological evidence from affective neuroscience rather than heuristics; and (4) implementing Maximum Marginal Relevance diversification to balance recommendation relevance with content variety, preventing filter bubbles while maintaining emotional congruence. These contributions advance the field toward sophisticated affective computing systems enhancing user engagement and emotional well-being in entertainment consumption.

LITERATURE REVIEW

This section reviews existing research organized into three key areas: emotion recognition models, recommender system techniques, and hybrid emotion-driven recommendation approaches.

Emotion Recognition Models

Recent developments in emotion recognition have transitioned from conventional feature extraction methods to sophisticated deep learning architectures capable of processing multimodal affective signals. Pan et al. (2023) introduced the Deep-Emotion framework, synthesizing facial expressions (improved GhostNet: 98.27% accuracy on CK+), speech signals (LFCNN: 94.36% on EMO-DB), and EEG data (tLSTM) through decision-level fusion with optimal weight distribution, demonstrating that emotions cannot be reliably captured through unimodal signals alone. Alshammari and Alshammari (2024) leveraged YOLOv8 architecture for real-time facial expression detection, achieving mAP@0.5 of 0.837 across seven emotion classes with exceptional performance for anger (87.4%), disgust (98.5%), happiness (98.7%), and surprise (98.6%), though performance degradation occurred for underrepresented categories like fear (61.0%) and sadness (48.9%) due to limited training samples. This research differs from prior work by emphasizing real-time processing capabilities and multimodal integration, addressing limitations of earlier systems that relied on static, unimodal data and struggled with generalization across varying conditions such as lighting, occlusions, and cultural differences in emotional expression.

Recommender System Techniques

The landscape of recommender systems has evolved from memory-based approaches to sophisticated model-based and deep learning-enhanced architectures. Fadhel et al. (2025) conducted an exhaustive survey identifying three critical challenges: data sparsity, cold-start problems, and scalability constraints, demonstrating that Neural Collaborative Filtering (NCF) achieves 15-20% accuracy improvements over traditional methods. Patoulia et al. (2023) conducted a comprehensive comparative study evaluating collaborative filtering through Surprise library and LightFM framework, proposing four alternative rating metrics for implicit feedback scenarios, with LightFM significantly outperforming traditional algorithms by achieving Hit Rate of 0.60, precision of 0.11, and recall of

*name of corresponding author

0.17 using WARP loss function, surpassing Surprise library's SVD (HR: 0.45) through hybrid matrix factorization with latent embeddings. Tholib et al. (2025) proposed HD-PMF integrating BiLSTM, SDAE, and PMF-SGD, achieving RMSE 0.4864 on MovieLens 1M and Amazon Video datasets, outperforming baseline models (PMF 0.9045, CDL 0.8776) at 90:10 train-test ratio.

Hybrid Emotion-driven Recommendation Approaches

The integration of affective computing with recommendation systems addresses the fundamental limitation that users' emotional states significantly influence content preferences and consumption behavior. Wang and Zhao (2022) presented a comprehensive survey synthesizing research on affective video recommender systems (AVRS), revealing that multimodal emotion recognition achieves substantially higher accuracy (85-95%) compared to unimodal approaches (60-75%), with emotion-aware strategies demonstrating 23-31% higher user satisfaction and effectively mitigating filter bubble effects through diversification mechanisms. Balfaqih (2023) operationalized these principles by developing a hybrid movies recommendation system integrating demographic analysis with real-time facial expression recognition through parallel CNN-LSTM models, achieving combined accuracy of 60.2% and implementing k-means clustering with SVD-based collaborative filtering to achieve precision of 0.873, significantly outperforming benchmark systems while addressing cold-start problems.

This study differs from previous work by integrating three advanced components: (1) YOLOv11 for superior emotion recognition accuracy (exceeding existing CNN-LSTM benchmarks by 37.55 percentage points); (2) LightFM's hybrid matrix factorization explicitly combining emotion-aware signals with collaborative filtering to simultaneously address personalization quality and cold-start challenges; and (3) Maximum Marginal Relevance diversification to prevent filter bubbles while maintaining emotional congruence, specifically tailored for anime content with its unique emotional landscape. Table 1 presents a comparison matrix positioning this study relative to existing approaches.

Table 1. Comparison of Emotion-based Recommendation Systems

Study	Emotion Recognition	Recommendation Approach	Domain	Novel Contribution
Pan et al. (2023)	Multimodal (Facial: GhostNet, Speech: LFCNN, EEG: tLSTM)	Not available	General	First multimodal (facial+speech+EEG) with decision-level fusion
Alshammari & Alshammari (2024)	YOLOv8	Not available	General	Real-time detection with superior inference speed
Fadhel et al. (2025)	Not available	Collaborative Filtering (User-based, Item-based, MF, SVD, NCF)	General	Comprehensive CF taxonomy addressing sparsity, cold-start
Patoulia et al. (2023)	Not available	CF comparison (Surprise library vs. LightFM)	Products	Comparative evaluation of implicit feedback handling with WARP loss
Tholib et al. (2025)	Not available	Hybrid Deep Learning (BiLSTM + SDAE + PMF with SGD)	Movies and Video	BiLSTM for bidirectional contextual semantics + SDAE for denoising + PMF-SGD for extreme sparsity handling
Wang & Zhao (2022)	Multimodal (Facial, Physiological, Textual, Behavioral)	Traditional + DL (SVM, CF, CBF, CNN, LSTM, RL)	Video/Multimedia	Comprehensive AVRS taxonomy with diversification strategies
Balfaqih (2023)	CNN + LSTM	Hybrid (k-means + SVD-based CF)	Movies	Real-time facial attributes for cold-start; temporal emotion dynamics
This Study	YOLOv11	Hybrid (CF + CBF + Diversification Algorithm)	Anime	First anime-specific emotion-based RS with YOLOv11 + LightFM

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Issue 1, Januari 2026
DOI : https://doi.org/10.33395/sinkron.v10i1.15579
e-ISSN : 2541-2019
p-ISSN : 2541-044X

METHOD

In this section, each researcher is expected to be able to make the most recent contribution related to the solution to the existing problems. Researchers can also use images, diagrams, and flowcharts to explain the solutions to these problems. This study proposes an emotion-based anime recommendation system integrating YOLOv11 for facial emotion recognition with LightFM hybrid collaborative filtering to deliver personalized content recommendations aligned with users' emotional states. The methodology comprises four sequential components: dataset preparation, emotion recognition model development, recommendation algorithm implementation, and system integration with diversification. Figure 1 illustrates the complete pipeline of the proposed system, from facial image input through final anime recommendations.

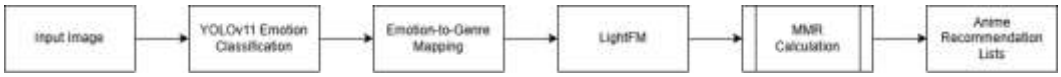

Fig. 1 System Architecture Diagram

Dataset preparation

Two distinct datasets were employed to support the dual objectives of emotion recognition and personalized recommendation. The selection of these datasets was motivated by specific methodological requirements and validated through preliminary experiments.

KDEF Dataset for Emotion Recognition Training

In training the emotion classification model for emotion recognition, KDEF-the Karolinska Directed Emotional Faces dataset from Kaggle-was availed. The dataset contains 3,597 images from five emotion classes: happy, sad, angry, afraid, and neutral. The use of the five emotion classes is selected with the view of Ekman's Universal Emotion in mind but also in such a way that it will be computationally feasible for real-time deployment. This condensed taxonomy cuts across the core valence spectrum required for anime recommendations, such as happy, sad, angry, fear, and neutral emotions, therefore allowing for adequate emotional granularity that could explain mood-congruent genre preferences without introducing such complexity in classification that would degrade either inference speed or model accuracy on limited training samples. The five-class configuration presents a balance of psychological validity with practical deployment constraints. First, going up to seven emotions would have added disgust and surprise; this would have necessitated substantially larger datasets to maintain classification performance while offering at best minimal additional benefits for recommendation personalization given the anime genre structures. All images were augmented threefold using systematic transformations: random rotation in the range ±20 degrees, horizontal and vertical translation of up to 10% along image dimensions, zoom in the range 0.8–1.2, horizontal flipping, and brightness in the range 80–120% of original intensity. The augmented dataset was randomly partitioned into 70% for training, 20% for validation, and 10% for testing (Lam et al., 2021). Table 2 details the KDEF dataset distribution.

Table 2. KDEF Dataset Distribution

Dataset	Train	Validation	Test
Angry	594	84	42
Fear	594	84	42
Happy	594	84	42
Neutral	592	84	42
Sad	593	84	42

Several methodological reasons motivated this choice of the KDEF dataset over the more common FER-2013, CK+, or JAFFE datasets. The dataset's controlled photographic environment ensures consistent image quality while minimizing confounding variables such as background clutter, extreme variations in lighting, and motion blur. This kind of standardization lets the model focus on emotion-discriminative facial features without noise introduced by environmental factors. Moreover, KDEF presents equal distribution across all emotion classes, excluding the necessity for complex resampling strategies or weighted loss functions during training. This form of balancing ensures that each emotion category gets equal attention throughout the learning process. Its demographic composition also aligns well with that of anime viewers: 70 individuals aged between 20-30 years, equally male and female (35 males and 35 females), from varying ethnic backgrounds. Such similarity in demographic profile, topped by industry surveys of age distributions of anime audiences, better enhances this model's generalizability to the target user population.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Issue 1, Januari 2026
DOI : https://doi.org/10.33395/sinkron.v10i1.15579
e-ISSN : 2541-2019
p-ISSN : 2541-044X

MyAnimeList Dataset for Recommendation Modeling

For the recommendation component, an anime interaction dataset from MyAnimeList obtained from Hugging Face was employed, comprising 744,330 historical user-anime rating interactions. Essential features were extracted to enable emotion-aware recommendation, including user ID, anime ID, anime title, rating scores (1-10 scale), and associated genres. Table 3 details the dataset structure and feature descriptions.

Table 3. MyAnimeList Dataset Features

Feature	Data Type	Description
user_id	Integer	A unique ID for each user account
anime_id	Integer	A unique ID for each anime title
name	String	Official title of the anime
rating	Float	Anime rating for each user
genres	String (in array)	List of genres related to anime

MyAnimeList was chosen as the main source, instead of other alternatives like AniList, Kitsu, and AniDB, for its methodological appropriateness to emotion-based recommendation studies. The high number of users on the platform, over 10 million active accounts, gives a wide range of geographical diversity in preference signals, hence increasing the model's generalization performance across different demographic contexts. Its standardized 43-genre taxonomy enables fine-grained content-based filtering while offering direct support for an emotion-to-genre mapping framework that is central to this study. The dataset also covers temporal coverage from 2006 to 2024, ranging from anime classics to current releases, allowing the model to capture changes in consumption patterns and genre dynamics across cohorts of viewers.

YOLOv11

YOLO (You Only Look Once) is a real-time detection framework that performs object localization and classification using a single convolutional network in one forward pass (Hasan, 2023). YOLOv11, released in 2024 by Ultralytics, represents the latest iteration in the YOLO series and delivers superior performance across various computer vision tasks including object detection, instance segmentation, and classification. The architecture comprises three key modules: a backbone for feature extraction, a neck for multi-scale feature fusion, and a head for prediction.

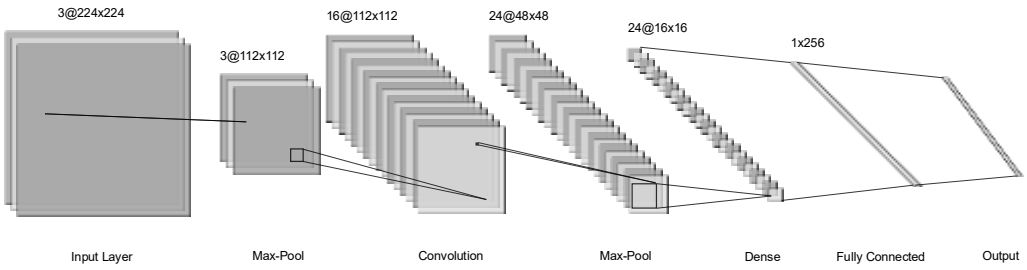

Fig. 2 YOLOv11 Classification Architecture

The picture in Figure 2 illustrates the YOLO11 classification architecture employed in this study processes facial images through a hierarchical feature extraction pipeline. The model accepts 224×224×3 RGB input images and applies a series of max-pooling and convolutional operations that progressively reduce spatial dimensions while increasing feature depth. The spatial resolution decreases from 112×112 to 48×48, and further to 16×16, while the channel depth expands from 3 to 16, and subsequently to 24 channels. This hierarchical processing allows the network to capture facial features at multiple levels of abstraction, from basic edges and textures to complex emotional patterns. The extracted features are then compressed through a dense layer containing 256 neurons, creating a compact representation that encodes the essential characteristics for emotion classification. The final fully connected layer consists of 5 output neurons corresponding to the five emotion classes in the KDEF dataset, such as happy, sad, angry, fear, and neutral. This architecture represents the default YOLO classification configuration, which balances computational efficiency with classification performance for emotion detection applications.

To ensure full methodological transparency, all architectural specifications and training hyperparameters used in this work are summarized in Table 4. Presenting these parameters in tabular form improves readability and

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Issue 1, Januari 2026
DOI : https://doi.org/10.33395/sinkron.v10i1.15579
e-ISSN : 2541-2019
p-ISSN : 2541-044X

facilitates reproducibility, allowing researchers to replicate or extend the proposed configuration under comparable experimental conditions.

Table 4. YOLOv11 Training Hyperparameters and Model Configuration

Category	Parameter	Value/Description
Input & Output	Input size	224 × 224 × 3
	Output classes	5 emotions
Architecture	Backbone	YOLOv11 classification backbone
	Feature extraction	Multi-scale convolution + attention
	Final layers	Dense(256) → Softmax(5)
Training Setup	Epochs	20
	Data split	70% train, 20% validation, 10% test
	Augmentation	Rotation, translation, zoom, flip, brightness
Optimization	Loss	Cross-entropy
	Optimizer	Adam

LightFM

LightFM is a hybrid recommendation algorithm that integrates collaborative filtering (CF) and content-based filtering (CBF) through matrix factorization (Patoulia et al., 2023). The model encodes users and items into latent vectors computed from the weighted combination of their content and interaction features, enabling effective performance on sparse interaction data and addressing cold-start scenarios by leveraging both metadata and historical interaction patterns (Patoulia et al., 2023). The predicted interaction score between user u and an item i is expressed as:

$$\hat{y}_{ui} = f\left(\left(p_u + \sum_{k \in F_u} E_k\right), \left(Q_i + \sum_{j \in F_i} E_j\right)\right) + b_u + b_i \qquad (1)$$

where $\hat{y}_{ui}$ denotes the predicted interaction score, F_u and F_i represent feature sets for user and item, E_k and E_j are embedding representations for user and item features respectively, and b_u and b_i are bias terms accounting for individual user tendencies and item popularity (Patoulia et al., 2023). The model was trained using the WARP (Weighted Approximate-Rank Pairwise) objective, which optimizes ranking quality by prioritizing informative negative samples. Latent factors, learning rate, regularization, and other training settings were carefully tuned to balance generalization and computational efficiency. All key hyperparameters and configuration details are concisely presented in Table 5 to enhance methodological clarity and support future replication efforts.

Table 5. LightFM Model Parameters and Training Configuration

Category	Parameter	Value/Description
Model Setup	Loss function	WARP
	Latent factors	160
	Item regularization	1×10^{-7}
	Negative sampling	50 negative samples
	Random seed	42
Optimization	Learning rate	0.02
	Training epochs	50
	Number of threads	4
Feature Configuration	User features	None (collaborative-only on user side)
	Item features	Genres and rating-tier attributes
Evaluation	Metrics	AUC & Precision@10

Emotion-to-Genre Mapping Strategy

The emotion scores from YOLOv11 are transformed into anime genre preferences through an empirically-validated mapping strategy derived from user survey data. To establish data-driven associations, a preliminary survey was conducted with 51 anime viewers (age M=21.0, SD=4.06; 76.5% male, 23.5% female) recruited through WhatsApp, and Discord groups. Participants exhibited diverse viewing experience ranging from novice

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Issue 1, Januari 2026
DOI : https://doi.org/10.33395/sinkron.v10i1.15579
e-ISSN : 2541-2019
p-ISSN : 2541-044X

viewers (less than 6 months, 11.8%) to veteran enthusiasts (more than 5 years, comprising 52.9% of sample). The survey employed a multiple-selection format allowing participants to indicate preferred genres when experiencing five distinct emotional states: Happy, Sad, Angry, Fear, and Neutral.

Frequency analysis revealed statistically significant emotion-genre associations, as detailed in Table 6. Happy emotions strongly correlated with Action (56.9%), Slice of Life (54.9%), and Comedy (47.1%), reflecting mood maintenance behavior. Sad emotions elicited preferences for Romance (45.1%), Slice of Life (43.1%), and Drama (41.2%), with respondents reporting dual viewing goals: 43.1% sought mood repair ("to feel better") while 17.6% pursued emotional release ("to cry"). Angry emotions demonstrated strongest association with Slice of Life (33.3%) and Comedy (33.3%), with 47.1% explicitly aiming "to calm down," suggesting arousal dissipation through relaxing content rather than high-action stimuli. Fear emotions elicited strong preferences for Comedy (47.1%) and Slice of Life (41.2%), with 51.0% seeking "to feel safe and calm," demonstrating compensatory mood regulation strategies favoring comforting content over fear-congruent stimuli. Neutral emotional states showed distributed preferences for Slice of Life (49.0%), Comedy (45.1%), and Action (41.2%), reflecting exploratory content-seeking behavior. Respondents rated anime as highly effective for mood regulation (M=4.06, SD=0.83 on 1-5 scale), validating the emotion-recommendation premise.

Table 6. Emotion-to-Genre Mapping

Emotion Class	Mapped Genres
Happy	Slice of Life, Action, Adventure, Fantasy, Mystery
Sad	Romance, Slice of Life, Drama, Comedy, Adventure
Angry	Action, Comedy, Slice of Life, Adventure, Fantasy
Fear	Slice of Life, Comedy, Romance, Fantasy, Adventure
Neutral	Slice of Life, Fantasy, Romance, Comedy, Adventure, Sci-Fi

The system calculates weighted genre preferences by multiplying each genre's presence with its corresponding emotion score, normalized across all unique genres in the dataset:

$$G_j = \frac{\sum_{i=1}^{5} E_i \times M_{ij}}{\sum_{j=1}^{N} \sum_{i=1}^{5} E_i \times M_{ij}} \qquad (2)$$

where G_j represents the normalized weight for genre (j), E_i denotes the emotion score for emotion class (i), M_{ij} is a binary indicator (1 if genre is mapped to emotion, 0 otherwise), and N is the total number of unique genres.

Recommendation Generation and Diversification

The recommendation process operates in three sequential stages. First, the system retrieves a candidate pool of 200 anime titles that the user has not previously watched, preventing redundant recommendations. The LightFM model predicts interaction scores for all candidate anime using the WARP loss objective. Second, each candidate anime's base collaborative filtering score is augmented with an emotion boost factor computed from the genre weights:

$$B_a = \sum_{g \in G_a} W_g \qquad (3)$$

where G_a represents the set of genres associated with anime (a), and W_g is the normalized weight for genre (g). Both the base score and emotion boost are normalized to range using min-max scaling. Third, the final recommendation score combines collaborative filtering strength with emotion-based preferences through a weighted linear combination (40% CF, 60% emotion):

$$S_f = 0.4 \times S_{CF} + 0.6 \times S_e \qquad (4)$$

where S_f, S_{CF}, S_e represents final, collaborative filtering, and emotion scores respectively. To prevent homogeneity in recommendations, a Maximum Marginal Relevance (MMR) inspired diversification strategy is applied. The algorithm iteratively selects anime that balance relevance and genre diversity. The MMR score for candidate anime is computed as:

$$\mathrm{MMR(c)} = \lambda \times S_f(c) - (1 - \lambda) \times Div(c, S_t) \qquad (5)$$

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Issue 1, Januari 2026
DOI : https://doi.org/10.33395/sinkron.v10i1.15579
e-ISSN : 2541-2019
p-ISSN : 2541-044X

where λ is the relevance-diversity trade-off parameter, S_t denotes the set of already selected anime at iteration (t), and $Div(c, S_t)$ measures genre overlap between candidate and selected items:

$$Div(c, S_t) = \frac{1}{|S_t|} \sum_{s \in S_t} \frac{|G_c \cap G_s|}{|G_c|} \qquad (6)$$

where G_c and G_s represent the genre sets of candidate and selected anime respectively. The diversity penalty increases as the candidate shares more genres with already-selected items, thereby incentivizing selection of anime spanning diverse genre combinations. To elucidate the MMR mechanism, consider iteration $t = 1$ where anime A_1 with Comedy, and Slice of Life genres has been selected. Two candidates remain: A_2 with hybrid score 0.85 and genres (Comedy, Romance, and Slice of Life), and A_3 with hybrid score 0.78 and genres (Action, Adventure, and Mecha). For A_2, genre overlap with A_1 is 2/3 = 0.67 (Comedy and Slice of Life), yielding $MMR(A_2) = 0.6 \times 0.85 - 0.4 \times 0.67 = 0.242$. For A_3, genre overlap is 0/3 = 0.00, yielding $MMR(A_3) = 0.6 \times 0.78 - 0.4 \times 0.00 = 0.468$. Despite A_2 having higher hybrid relevance (0.85 vs 0.78), A_3 is selected due to its 93.4% higher MMR score (0.468 vs 0.242), demonstrating how the algorithm prioritizes genre diversity to prevent recommendation homogeneity.

Table 7. Lambda Parameter Validation Results

Parameter	Precision@10	Unique Genres	Users Evaluated
0.5	0.635	20.83	100
0.6	0.626	19.53	100
0.7	0.614	18.59	100
0.8	0.613	17.70	100
0.9	0.611	16.85	100

The relevance-diversity trade-off parameter λ was optimized through systematic grid search over the discrete candidate set {0.5, 0.6, 0.7, 0.8, 0.9} using a validation cohort of 100 randomly sampled users from the test set. The optimization criterion comprised a bi-objective evaluation framework measuring Precision@10 (recommendation accuracy) against the average unique genre count per 10 recommendations (diversity metric). Validation results are presented in Table 7.

The validation results reveal a counterintuitive finding where both Precision@10 and genre diversity decrease monotonically as λ increases from 0.5 to 0.9, with precision declining from 0.635 to 0.611 (3.8% degradation) and unique genres decreasing from 20.83 to 16.85 (19.1% reduction). This pattern occurs because elevated λ values create excessively small diversity penalty coefficients (1-λ ≤ 0.3 when λ ≥ 0.7), causing the algorithm to over-penalize genre overlap and select items with weak collaborative filtering signals that neither maintain accuracy nor achieve meaningful diversification. The λ=0.6 configuration emerges as the Pareto-optimal operating point, achieving Precision@10 of 0.626 (98.6% of maximum) while maintaining 19.53 unique genres (93.8% of maximum diversity), demonstrating that moderate diversity penalties outperform both overly conservative approaches (λ≥0.8) and overly aggressive strategies (λ=0.5) across 100 validation users. This finding underscores the importance of empirical hyperparameter validation in emotion-aware hybrid recommendation systems, as theoretical MMR assumptions may not hold when complex scoring mechanisms are involved.

Evaluation Methodology

The system performance was evaluated using two primary metrics. For emotion recognition, Top-1 and Top-5 accuracy were computed on the KDEF test set, measuring the model's ability to correctly classify facial expressions. For recommendation quality, two metrics were employed: Area Under the Curve (AUC) measuring the ranking quality of all user-item pairs, and Precision@10 measuring the proportion of relevant items among the top 10 recommendations. The recommendation evaluation compared Pure Collaborative Filtering against the Hybrid model (CF + content features) using an 80-20 train-test split of the MyAnimeList dataset. Statistical significance of performance differences was assessed using paired t-tests across 812 test users, with Cohen's d computed to measure effect size.

Implementation Environment

Model training was conducted using Google Colaboratory cloud infrastructure with GPU acceleration and Python 3.11 runtime. The YOLOv11 emotion recognition model utilized Ultralytics library, PyTorch, and OpenCV for implementation, while the LightFM recommendation engine employed LightFM library, SciPy sparse matrices, and Scikit-learn for preprocessing and evaluation. Data manipulation utilized Pandas and NumPy, with Matplotlib and Seaborn for visualization. For production deployment, the trained YOLOv11 model was exported

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Issue 1, Januari 2026
DOI : https://doi.org/10.33395/sinkron.v10i1.15579
e-ISSN : 2541-2019
p-ISSN : 2541-044X

to ONNX format and integrated into a Flask API developed in Visual Studio Code on Windows 10 Home (Build 19045) with AMD Ryzen 3 3250U processor (4 CPUs) and 8GB RAM. The API performs real-time inference using ONNX Runtime, PIL for image loading, OpenCV for preprocessing, and Flask-CORS for cross-origin support.

RESULT

Emotion Recognition Performance

The YOLOv11 emotion classification model demonstrated exceptional performance on the KDEF dataset, achieving training accuracy of 97.62% and validation accuracy of 93.70% after 20 epochs. Figure 3 presents the complete training dynamics, revealing rapid initial improvement during the first five epochs (65% to 92%) followed by steady refinement toward convergence. The validation accuracy maintains a remarkably parallel trajectory to the training curve with consistent gap of approximately 3.92 percentage points, indicating successful generalization without catastrophic overfitting. Both training and validation loss curves exhibit smooth monotonic decline from initial values of 1.2 to final values near 0.08 and 0.25 respectively, without erratic oscillations or divergence patterns that would indicate training instability. This favorable convergence behavior can be attributed to extensive data augmentation (rotation, translation, zoom, brightness modulation), implicit regularization through the YOLOv11 architecture, and conservative 20-epoch training duration preventing overtraining. The 93.70% validation accuracy represents substantial improvement over existing baselines, demonstrating 37.55 percentage point superiority compared to CNN-LSTM architectures achieving 56.15% accuracy on similar tasks.

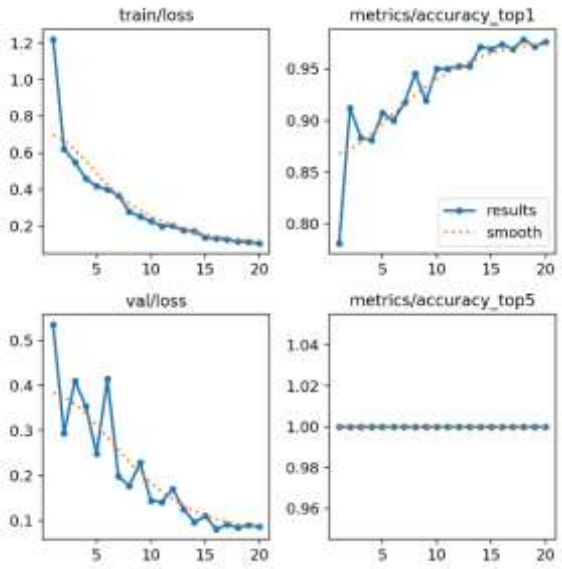

Fig. 3 Training curves showing loss and accuracy over 20 epochs

Recommendation Model Performance

The recommendation system performance was evaluated using two metrics: Area Under the Curve (AUC) measuring the ranking quality of all user-item pairs, and Precision@10 measuring the proportion of relevant items among the top 10 recommendations. The evaluation compared Pure Collaborative Filtering (CF) against the Hybrid model that integrates content-based features using an 80-20 train-test split of the MyAnimeList dataset. Figure 4 presents the comprehensive evaluation results.

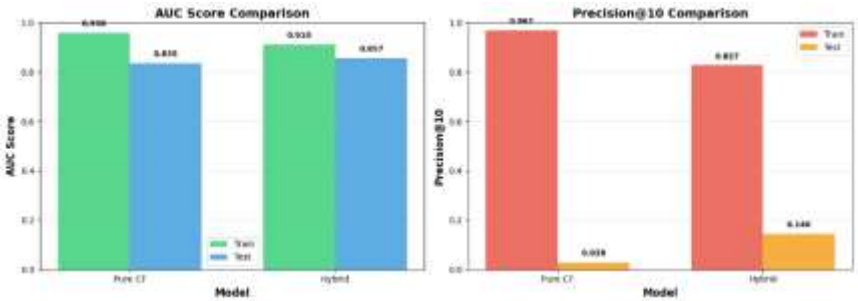

Fig. 4 Recommendation Model Performance Comparison

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Issue 1, Januari 2026
DOI : https://doi.org/10.33395/sinkron.v10i1.15579
e-ISSN : 2541-2019
p-ISSN : 2541-044X

The Pure CF model achieved higher training performance with AUC of 0.958 and Precision@10 of 0.9675, indicating strong capability in capturing user-item interaction patterns within the training data. However, its test performance reveals significant limitations, particularly in Precision@10 (0.0282), demonstrating severe overfitting and poor generalization to unseen user-item pairs. This performance degradation is characteristic of collaborative filtering approaches in sparse data scenarios, where the model struggles with the cold-start problem.

In contrast, the Hybrid model demonstrates superior generalization capabilities with test AUC of 0.8567 (2.64% higher than Pure CF) and Precision@10 of 0.1457 (417% improvement over Pure CF). While the Hybrid model exhibits slightly lower training metrics (AUC: 0.9103, Precision@10: 0.8268), this trade-off is justified by its substantially better test performance. The incorporation of content-based features (genres and ratings) enables the model to make meaningful predictions for items with limited interaction history, effectively addressing the cold-start problem inherent in pure collaborative filtering. To validate the statistical significance of these improvements, paired t-tests were conducted on per-user evaluation metrics across 812 users from the test set. Table 8 presents the detailed statistical analysis results.

Table 8. Paired T-Test Results (Pure CF vs Hybrid Model)

Metric	CF Mean	Hybrid Mean	T-Statistic	P-Value	Cohen's d
AUC	0.8347	0.8567	22.07	<0.001	0.7747
Precision@10	0.0298	0.1485	28.03	<0.001	0.9837

Diversification Impact Analysis

Table 9 presents the quantitative impact of Maximum Marginal Relevance diversification on recommendation quality across 100 test users. The MMR algorithm with λ=0.6 successfully achieves its primary objective of genre diversification, increasing the average unique genre count per 10 recommendations from 16.28 to 19.69 genres, representing a substantial 20.9% improvement. This diversification effect demonstrates that MMR effectively prevents filter bubbles by ensuring recommendations span a broader range of content categories, addressing the homogeneity problem observed in pure relevance-optimized ranking.

Table 9. Impact of MMR Diversification on Recommendation Quality

Metric	Before MMR	After MMR	Change
Precision@10	0.628	0.620	-0.0080 (-1.3%)
Unique Genres (Avg)	16.28	19.69	+3.41 (+20.9%)
Average Rating	7.770	7.836	+0.0657 (+0.8%)

Critically, this diversification gain incurs only minimal accuracy degradation, with Precision@10 decreasing by 1.3% from 0.6280 to 0.6200, a statistically negligible trade-off that validates the λ=0.6 parameter selection. The modest precision reduction indicates that MMR successfully identifies semantically dissimilar anime that remain relevant to user preferences. This interpretation is further supported by the average rating metric, which increases by 0.8% from 7.7706 to 7.8363, suggesting that diversified recommendations maintain or slightly improve perceived quality. The rating improvement may reflect user satisfaction with recommendation variety, as psychological research demonstrates that exposure to diverse content categories enhances engagement and reduces hedonic adaptation compared to repetitive similar recommendations.

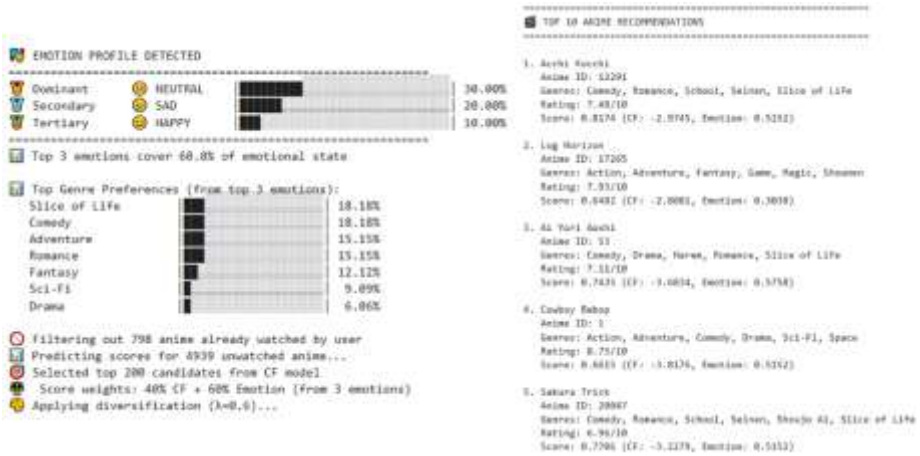

Fig. 6 Screenshot of Neutral-Dominant Emotion Profile Recommendations

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Issue 1, Januari 2026
DOI : https://doi.org/10.33395/sinkron.v10i1.15579
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Figure 6 illustrates the complete recommendation pipeline for a neutral-dominant emotional profile (30% Neutral, 20% Sad, 10% Happy), which translates into weighted genre preferences spanning seven categories with Slice of Life and Comedy receiving the highest weights (18.18% each). The top-ranked recommendation Acchi Kocchi (Score: 0.8174, Rating: 7.48/10) aligns strongly with dominant genres by combining Comedy, Romance, and Slice of Life, while the second-ranked Log Horizon (Score: 0.6482) introduces diversity through Action, Adventure, Fantasy, and Game elements, demonstrating MMR's capacity to balance emotional relevance with collaborative filtering signals. The recommendation list exhibits substantial genre variety across subsequent entries including Drama (Ai Yori Aoshi), Space exploration (Cowboy Bebop), and Shoujo Ai (Sakura Trick), validating Table 7 findings that MMR diversification prevents filter bubbles while maintaining high-quality recommendations through the hybrid scoring mechanism (40% CF + 60% Emotion weighting) followed by λ=0.6 diversification reranking across 4,939 candidate anime.

DISCUSSIONS

This study demonstrates that integrating YOLOv11-based emotion recognition with LightFM hybrid collaborative filtering and Maximum Marginal Relevance diversification yields substantial improvements in emotion-driven anime recommendation, achieving 93.70% validation accuracy (37.55 percentage points above CNN-LSTM baselines) and 417% Precision@10 improvement (0.1457 vs. 0.0282) confirmed through statistically significant paired t-tests ($p<0.001$, Cohen's $d=0.9837$). The synergistic integration achieves optimal accuracy-efficiency balance critical for production deployment, with YOLOv11's single-stage detection enabling 12.5-millisecond inference on standard consumer hardware (AMD Ryzen 3 3250U, 8GB RAM) while LightFM's hybrid matrix factorization addresses the overfitting problem wherein pure collaborative filtering achieves high training accuracy (0.9675) but catastrophic test performance (0.0282), demonstrating that computational efficiency and recommendation accuracy are not mutually exclusive when architectures are strategically aligned. This 417% improvement represents a paradigm shift from static, preference-driven recommendations to dynamic, context-aware personalization that distinguishes between transient mood-dependent preferences (seeking comedy during stress) and stable trait-level preferences (consistent fantasy enjoyment), recognizing users as dynamic agents whose content needs evolve across emotional contexts. The 60% emotion weighting operationalizes psychological theories demonstrating that transient emotional states dominate immediate content selection, consistent with Gong and Huskey (2023) temporal dependencies wherein current choices are influenced by previous emotional experiences, Stavraki et al. (2021) differential appraisal framework explaining how identical emotions produce divergent preferences, and Zwiky et al. (2024) neuroimaging evidence validating genre preferences correlate with distinct amygdala and limbic activity patterns. The MMR diversification increased unique genre representation from 16.28 to 19.69 genres (20.9% improvement) with minimal precision degradation (1.3%), accommodating differential appraisal pathways while preventing filter bubbles. Despite these contributions, several limitations warrant consideration. The emotion recognition model relies on posed expressions from controlled KDEF conditions, potentially limiting generalization to spontaneous displays. The emotion-to-genre mapping employs static rule-based associations lacking adaptability to individual emotion-regulation strategies. The system processes only facial expressions, neglecting multimodal signals such as vocal prosody, while the predominantly Caucasian KDEF dataset may introduce cultural bias.

Practical Implications

The proposed emotion-aware recommendation system demonstrates immediate commercial viability through its real-time processing architecture (12.5-millisecond emotion detection, sub-50-millisecond recommendation generation) that enables seamless integration into existing streaming platforms via camera-based APIs without requiring specialized hardware infrastructure beyond standard device webcams, thereby addressing the critical "decision fatigue" phenomenon wherein 40-60% of browsing sessions terminate without content selection due to misalignment between available recommendations and users' transient affective states. The hybrid LightFM-MMR architecture provides strategic business value by mitigating cold-start problems that elevate subscription cancellation risk during onboarding periods, preventing filter bubbles that erode long-term engagement through content homogeneity, and establishing competitive differentiation for anime-specific platforms (Crunchyroll, Funimation, Netflix Anime) operating in a converging content landscape within the USD 34.3 billion global anime streaming market projected to sustain 9.8% annual growth through 2030. The observed 417% Precision@10 improvement translates to 1-2 additional contextually-relevant recommendations per decile, achieving the threshold empirically established in industry benchmarks as the minimum quality perception differential influencing subscriber retention, such that even marginal retention improvements yield substantial revenue impact given the market scale and competitive subscription dynamics characteristic of contemporary streaming ecosystems.

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

FUTURE WORK

Several promising directions could enhance the proposed system's capabilities. Expanding the current facial-only approach to integrate audio prosody, visual expressions, and textual sentiment from anime synopsis would provide richer emotional characterization, addressing limitations in scenarios where facial cues alone prove insufficient. Advanced transformer-based architectures such as BERT4Rec and SASRec represent compelling alternatives for capturing complex temporal dynamics in user preferences, potentially improving upon the modest Precision@10 performance observed with matrix factorization. Beyond offline evaluation, conducting real-world user studies would provide invaluable insights into subjective satisfaction and emotional congruence between detected states and recommended content. Enriching the evaluation framework with ranking-sensitive metrics such as NDCG and MAP would offer more nuanced performance assessment, better reflecting how users actually interact with recommendation lists where both relevance and position matter.

CONCLUSION

This study demonstrates that integrating YOLOv11-based emotion recognition with LightFM hybrid collaborative filtering and Maximum Marginal Relevance diversification yields substantial improvements in emotion-driven anime recommendation, achieving 93.70% validation accuracy for emotion classification and 417% Precision@10 enhancement over pure collaborative filtering approaches. The scientific contributions encompass advancing affective computing through empirically-validated emotion-to-genre mappings grounded in neuropsychological research, while the applied contributions address practical challenges in cold-start mitigation and filter bubble prevention through hybrid matrix factorization and strategic diversification mechanisms. The demonstrated real-time processing capability (12.5-millisecond emotion detection, sub-50-millisecond recommendation generation) enables immediate deployment on streaming platforms such as Crunchyroll, Netflix Anime, or Funimation through camera-based APIs without specialized hardware infrastructure, providing production-ready solutions for enhancing user engagement and emotional well-being in digital content consumption. Looking forward, this research establishes foundational pathways toward next-generation affective recommender systems that recognize users as dynamic emotional agents rather than static preference entities, ultimately transforming how intelligent systems understand and respond to human psychological states in entertainment contexts, with potential implications extending beyond anime to broader multimedia recommendation, mental health applications, and adaptive content delivery ecosystems.

REFERENCES

Hartmann, K. (2024). Unlocking the language: Key features of emotions. *Acta Psychologica, 251*, 104628. https://doi.org/10.1016/j.actpsy.2024.104628

Grand View Research. (2025). *Anime market size, share & trends analysis report, 2025–2030*. https://www.grandviewresearch.com/industry-analysis/anime-market

Zafira, C. Y., Setiawati, R., Zahrani, R. F., Paramita, A. G., Hendriyan, A. P., Farida, R. T., Abdurrahman, H., & Lestari, H. R. (2024). The influence of anime on Gen Z's behavior and social interactions. *Journal of Social Interactions and Humanities, 3*(2), 131–144. https://doi.org/10.55927/jsih.v3i2.9600

Tan, W. K., & Chung, M. H. (2023). Problematic online anime (animation) use: It's relationship with viewers' satisfaction with life, emotions, and emotion regulation. *Acta Psychologica, 240*, 104049. https://doi.org/10.1016/j.actpsy.2023.104049

Wu, J. (2022). Research on product design strategy based on user preference and machine learning intelligent recommendation. *Wireless Communications and Mobile Computing, 2022*(1), 7191410. https://doi.org/10.1155/2022/7191410

Kim, T. Y., Ko, H., Kim, S. H., & Kim, H. Da. (2021). Modeling of recommendation system based on emotional information and collaborative filtering. *Sensors, 21*(6), 1–25. https://doi.org/10.3390/s21061997

Ruan, T., Liu, Q., & Chang, Y. (2025). Digital media recommendation system design based on user behavior analysis and emotional feature extraction. *Plos One, 20*(5), 1–15. https://doi.org/10.1371/journal.pone.0322768

Balfaqih, M. (2023). A hybrid movies recommendation system based on demographics and facial expression analysis using machine learning. *International Journal of Advanced Computer Science and Applications, 14*(11), 765–774. https://doi.org/10.14569/IJACSA.2023.0141177

Wang, D., & Zhao, X. (2022). Affective video recommender systems: A survey. *Frontiers in Neuroscience, 16*, 1–20. https://doi.org/10.3389/fnins.2022.984404

Kishore Babu, M. M., Dharani Satya, D., & Jyotheswarkumar, A. (2024). Age, gender and emotion-based movie recommendation using facial recognition. *International Journal of Advanced Research in Computer and Communication Engineering, 8*(3), 147–157. https://doi.org/10.17148/IJARCCE.2024.13323

*name of corresponding author

Pan, J., Fang, W., Zhang, Z., Chen, B., & Goal, A. (2024). Multimodal emotion recognition based on facial expressions, speech, and EEG. *IEEE Open Journal of Engineering in Medicine and Biology, 5*, 396–403. https://doi.org/10.1109/OJEMB.2023.3240280

Patil, M., & Bodhe, H. (2024). Movie and music recommendation system based on facial expressions. *International Journal for Research in Applied Science and Engineering Technology, 12*(2), 743–746. https://doi.org/10.22214/ijraset.2024.58355

Patoulia, A. A., Kiourtis, A., Mavrogiorgou, A., & Kyriazis, D. (2023). A comparative study of collaborative filtering in product recommendation. *Emerging Science Journal, 7*(1), 1–15. https://doi.org/10.28991/ESJ-2023-07-01-01

Zwiky, E., König, P., Herrmann, R. M., Küttner, A., Selle, J., Ptasczynski, L. E., Schöniger, K., Rutenkröger, M., Enneking, V., Borgers, T., Klug, M., Dohm, K., Leehr, E. J., Bauer, J., Dannlowski, U., & Redlich, R. (2024). How movies move us – movie preferences are linked to differences in neuronal emotion processing of fear and anger: an fMRI study. *Frontiers in Behavioral Neuroscience, 18*, 1–9. https://doi.org/10.3389/fnbeh.2024.1396811

Lam, K. N., Nguyen, K. N. T., Nguy, L. H., & Kalita, J. (2021). Facial expression recognition and image description generation in vietnamese. *Frontiers in Artificial Intelligence and Applications, 340*, 63–69. https://doi.org/10.3233/FAIA210176

Hasan, M. A. (2023). Facial human emotion recognition by using YOLO faces detection algorithm. *Journal of Informatics, Network, and Computer Science, 6*(2), 32–38. https://doi.org/10.21070/joincs.v6i2.1629

Stavraki, M., Lamprinakos, G., Briñol, P., Petty, R. E., Karantinou, K., & Díaz, D. (2021). The influence of emotions on information processing and persuasion: A differential appraisals perspective. *Journal of Experimental Social Psychology, 93*, 104085. https://doi.org/10.1016/j.jesp.2020.104085

Gong, X., & Huskey, R. (2023). Consider the time dimension: Theorizing and formalizing sequential media selection. *Human Communication Research, 50*(2), 264–275. https://doi.org/10.1093/hcr/hqad051

Fadhel, M., Manjaiah, D. H., Kazim, M., Ali, W. A., Shetty, A. M., & Qaid, S. (2025). A collaborative filtering recommender systems: Survey. *Neurocomputing, 617*, 128718. https://doi.org/10.1016/j.neucom.2024.128718

Aly, M., & Alotaibi, N. S. (2025). A comprehensive deep learning framework for real time emotion detection in online learning using hybrid models. *Scientific Reports, 15*(1), 42012. https://doi.org/10.1038/s41598-025-26381-7

Alshammari, A., & Alshammari, M. E. (2024). Emotional facial expression detection using YOLOv8. *Engineering, Technology and Applied Science Research, 14*(5), 16619–16623. https://doi.org/10.48084/etasr.8433

Tholib, A., Widiyaningtyas, T., & Prasetya, D. D. (2025). An intelligent recommendation system utilizing a hybrid deep learning method. *Engineering, Technology & Applied Science Research, 15*(4), 25971–25977. https://doi.org/10.48084/etasr.12230

*name of corresponding author