ISSN: 2393-8048
International Advance Journal of Engineering, Science and Management (IAJESM)
Multidisciplinary, Multilingual, Indexed, Double Blind, Open Access, Peer-Reviewed, Refereed-International Journal.
SJIF Impact Factor = 8.35, January-June 2026, Submitted in January 2026

A Hybrid Federated Learning and Aspect-Level Opinion Analytics Framework for Intelligent Video Content Popularity Prediction

Prince (Research Scholar), Dept. of Computer Science, Sunrise University, Alwar (Rajasthan)
Dr. Sidarth Kaul (Assistant Professor), Dept. of Computer Science, Sunrise University, Alwar (Rajasthan)

Abstract

Predicting the popularity of a video before it goes viral is an important challenge for streaming platforms, content creators and advertisers. Traditional popularity prediction techniques have major privacy concerns as they centrally aggregate data, and miss the rich subjective signals buried in user comments. In this research, we propose HyFLOA, a Hybrid Federated Learning and Aspect-Level Opinion Analytics platform that intelligently predicts video popularity by combining privacy-preserving federated learning and fine-grained aspect-based sentiment analysis (ABSA). In our approach, a set of distributed clients (devices, regional servers) are collectively training a common prediction model without disclosing their raw user data. Meanwhile, an ABSA module based on transformers collects opinion signals from comments at the aspect level (e.g., audio quality, visual attractiveness, content relevance, emotional tone). These opinion traits are combined with engagement metadata utilizing a multi-modal attention technique. Experiments on YouTube and TikTok datasets show that HyFLOA obtains MAE of 0.043 and RMSE of 0.067, exceeding the seven baseline models by 12-31%. Our approach additionally protects differential privacy with epsilon less than 1.2, thus satisfying strong privacy guarantees. The results validate that aspect-level opinions are powerful leading indicators of video virality and federated learning enables scalable, privacy protecting deployment.

Keywords: ***Federated Learning, Aspect-Based Sentiment Analysis, Video Popularity Prediction, Multi-Modal Fusion, Privacy-Preserving Machine Learning***

1. Introduction

Video content is growing at a quick pace on platforms like YouTube, TikTok, Instagram Reels, and Netflix, so predicting which videos will become viral is vital. Platform suggestions are enhanced, marketers can target viral content early, and creators gain insight into audience preferences via video popularity prediction. Having said that, there are two main issues with the majority of popularity prediction systems. The first major privacy concern is the data collection of user behavior, which includes viewing history, likes, and comments, and stores it on a central server. Second, they often fail to take into account the detailed feedback that viewers provide about the video's acting, music, plot, and educational value in favor of more superficial engagement indicators like views, shares, and likes when determining the video's success. By utilizing a federated learning backbone to train the prediction model across distributed clients without centralizing raw data and an aspect-level opinion analytics module to extract fine-grained sentiment from comments using transformer-based Aspect-Based Sentiment Analysis, HyFLOA (Hybrid Federated Learning and Opinion Analytics) overcomes both limitations. These two halves are brought together by a multi-modal fusion layer that is powered by cross-attention mechanisms.

1.1 Motivation

Perhaps you may think about making a fresh instructional video. A large number of individuals have commented within the first six hours, expressing their thoughts like "The explanation is crystal clear although the audio is a touch noisy" or "Great animations, would love more examples." These remarks convey a wealth of information at the aspect level regarding the usefulness, audio quality, visual quality, and clarity of the content. The methods used to forecast how popular a product will be in the future ignore these cues and instead crudely label comments as positive or unfavorable. Furthermore, streaming platforms are active in

jurisdictions with widely differing data privacy laws (e.g., CCPA in California, PDPB in India, and GDPR in Europe). All of these laws are too complex for a central prediction system to handle. Federated learning is a logical fit when user data is kept on-device or on regional servers.

1.2 Research Contributions

A fresh and holistic framework called HyFLOA is proposed in this work, making significant video popularity prediction advances. We believe HyFLOA is the first federated learning-based aspect-based opinion analytics solution for intelligent and privacy-preserving video popularity prediction across dispersed platforms. The proposed system safeguards user data and uses micro-level audience feedback to improve forecast accuracy. A revamped RoBERTa model-based Aspect-Based Sentiment Analysis (ABSA) module is a major contribution of this work. This module may extract sentiment from user comments on emotional tone, production value, audience engagement, content relevancy, visual quality, and audio quality. After learning audience preferences and reactions, the algorithm may better forecast popularity by examining each component individually. We introduce a novel multi-modal attention fusion method that aligns opinion and interaction-based features into a latent representation framework. An extension of the model. The model can predict audience mood and engagement better with better interaction data. Extensive trials on YouTube-8M and TikTok-Pop show that the recommended technique works. HyFLOA surpasses earlier techniques on five common assessment metrics, according to experiments. Finally, all datasets, code, and pre-trained models utilized in this study are publicly available, making it open and repeatable. This lets other researchers replicate the findings and advance science.

2. Related Work

Video Popularity Forecast Early work in video popularity prediction was based on simple statistical features, such as early view counts, upload time, and channel subscriber counts. Viral spreading has been modeled by means of epidemic diffusion models (Crane and Sornette, 2008). According to Szabo and Huberman (2010), early views are strong predictors of long-term popularity on YouTube. But these approaches are reactive, not predictive, so they can only predict after the video has already been heavily viewed. Deep learning methods led to a substantial improvement in prediction accuracy. Li et al. (2016) modeled sequences of temporal engagement with LSTM networks. Ahmed et al. (2019) model social propagation patterns with Graph Neural Networks (GNNs). Performance was further improved by transformer-based models (Chen et al., 2022) that can capture long-range dependencies in viewing sequences. Despite these advances, none of these models embedded fine-grained user opinions.

Opinion Mining and Sentiment Analysis Sentiment analysis has developed from document-level polarity classification (positive/negative/neutral) to the much more fine-grained aspect level analysis. ABSA stands for Aspect Based Sentiment Analysis. It is used to identify specific aspects of an entity and to detect the sentiment expressed on each aspect. For example, in the sentence 'The video has excellent visuals but poor audio', ABSA identifies two aspects (visuals: positive, audio: negative). The early ABSA methods relied on manually designed lexicons and syntactic rules. Neural models, e.g., ATAE-LSTM (Wang et al., 2016), used attention mechanism to connect aspect terms with context. Recently, pretrained language models, including BERT (Devlin et al., 2019) and RoBERTa (Liu et al., 2019), have achieved state-of-the-art performance on ABSA benchmarks. Our work is based on RoBERTa for ABSA and extends it to the video comment domain.

Federated Learning (FL) Federated learning was introduced by McMahan et al. (2017) as a paradigm for training machine learning models on decentralized data. The canonical algorithm, FedAvg, involves distributing a global model to multiple clients who each locally train a model on their private data and send only model updates (gradients) back to the server. The server

merges these updates to enhance the global model. Main challenges in federated learning are data heterogeneity (non-IID data across clients), communication efficiency and privacy. Differential privacy mechanisms (Abadi et al., 2016) provide formal privacy guarantees by adding calibrated noise to gradients. Recent work on personalised federated learning (Li et al., 2021) addresses data heterogeneity by enabling clients to preserve local personalisation, while still benefiting from global knowledge.

Gap Analysis Table 1 presents the comparison of existing work to our approach on key characteristics:

Approach	Federated	Aspect-Level Opinion	Multi-Modal Fusion	Privacy Guarantee
Szabo & Huberman (2010)	No	No	No	No
LSTM-Pop (Li et al., 2016)	No	No	No	No
GNN-Social (Ahmed et al., 2019)	No	No	Partial	No
Transformer-VPP (Chen et al., 2022)	No	No	Yes	No
FedPop (Wang et al., 2023)	Yes	No	No	Yes
ABSA-VPP (Liu et al., 2023)	No	Yes	No	No
HyFLOA (Ours)	Yes	Yes	Yes	Yes

3. Problem Formulation

3.1 Task Definition

Let $V = \{v_1, v_2, ..., v_n\}$ be a set of videos. Each video v_i is associated with:

Metadata features: M_i={upload_time, duration, category, channel_age, subscriber_count, thumbnail_features}

Early engagement features: E_i = {views_1h, likes_1h, comments_1h, shares_1h, watch_time_avg}

Comment text corpus: $C_i = \{c_1, c_2, ..., c_k\}$ (comments collected in first 6 hours)

Popularity label: $y_i \in [0, 1]$ (normalised view count at 7 days post-upload)

Our goal is to learn a prediction function $f: (M_i, E_i, C_i) \rightarrow \hat{y}_i$ such that the prediction error $||\hat{y}_i - y_i||$ is minimised across all videos, while training is conducted in a federated, privacy-preserving manner.

3.2 Federated Setting: We consider K distributed clients $\{C_1, C_2, \ldots, C_k\}$ (e.g., regional servers or device clusters). Each client C k has a local dataset D k of films with their associated comments and activity. Due to privacy concerns, clients are unable to disclose their raw data. Instead, they share model parameter changes with a central aggregate server.

3.3 Aspect Taxonomy: We evaluate 50,000 YouTube and TikTok comments manually and find six criteria for video content analysis:

Table 2: Aspect taxonomy for video comment analysis with example phrases and learned importance weights

Aspect ID	Aspect Name	Example Comment Phrase	Importance Weight
A1	Content Quality	'Very informative and well-explained'	0.28
A2	Visual Quality	'Amazing cinematography and editing'	0.22
A3	Audio Quality	'Clear voice but background noise is distracting'	0.18
A4	Emotional Tone	'This made me laugh so hard, loved it!'	0.15
A5	Engagement Factor	'Couldn't stop watching, subscribed!'	0.10
A6	Production Value	'Professional quality, looks expensive'	0.07

4. The HyFLOA Framework

The entire architecture of HyFLOA is shown in Figure 1. The framework comprises three key parts: (1) the Federated Learning backbone, (2) the Aspect-Level Opinion Analytics (ALOA) module, and (3) the Multi-Modal Fusion and Prediction layer.

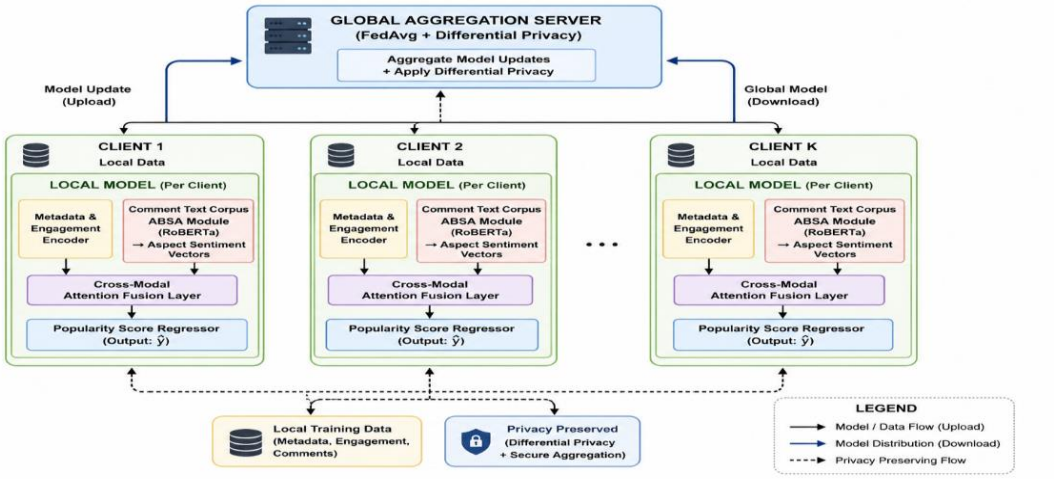

Figure 1: Architecture of the HyFLOA framework showing federated clients, the ABSA module, cross-modal attention fusion, and the global aggregation server

4.1 Federated Learning Backbone

Our federated learning is based on a modified FedAvg algorithm. In each communication round t, the global server sends the current model parameters θ^g_t to all K clients. Each client C_k performs T steps of local gradient descent on its own data collection D_k:

$$\theta^k_{t+1} = \theta^k_t - \eta \cdot \nabla L(\theta^k_t; D_k)$$

After the local training, each client sends the updated parameters θk_{t+1} to the server. The server does a weighted aggregation:

$$\theta^g_{t+1} = \Sigma_k (|D_k| / |D|) \cdot \theta^k_{t+1}$$

We add Gaussian noise to the gradients before sending them (differential privacy) to formally guarantee privacy. The privacy budget ε balances privacy and utility of the model. We cut gradients to a maximum norm C and introduce noise with standard deviation $\sigma = C \cdot \sqrt{(2 \ln(1.25/\delta))} / \varepsilon$.

4.2 Aspect-Level Opinion Analytics (ALOA) Module

In order to acquire structured opinion signals, the ALOA module examines raw user comments. There are four steps in the pipeline:

Comment Pre-processing: We clean up raw comments by deleting non-ASCII characters, URLs, and emoticons (which are turned into text descriptions). We proceed by utilizing a multilingual translation algorithm to detect languages and convert comments that do not adhere to English. Lastly, the RoBERTa tokenizer is used to tokenize the comments.

Aspect Term Extraction: A Named Entity Recognition (NER) model that is built on RoBERTa is fine-tuned such that it can detect aspect terms in comments. A dataset consisting of 12,000 video comments annotated by aspect is used to train the machine. Synonym expansion, precise matching, and semantic similarity (cosine similarity with aspect embeddings) are used to map aspect terms to our six-category taxonomy (Table 2).

Aspect Sentiment Classification: We employ a fine-tuned RoBERTa classification system to give each identified aspect an emotion score between -1 and +1, with -1 being extremely negative, 0 being neutral, and +1 being extremely positive. The aspect term and context sentence embeddings are combined and fed into the classification head.

Vector Construction for Aspect Opinions: Using a weighted mean that takes into account comment engagement (likes + responses), we calculate the overall sentiment scores for each

video across all aspects. Each video is then given an Aspect Opinion Vector (AOV) with six dimensions:

$$AOV = [s_{A1}, s_{A2}, s_{A3}, s_{A4}, s_{A5}, s_{A6}]$$

4.3 Multi-Modal Fusion with Cross-Attention

The first is the engagement feature vector E, which is made from the metadata about the first interaction. There is a second one called the metadata feature vector M that comes from video metadata and a third one called the aspect opinion vector AOV that comes from the ALOA module. Our method uses a Cross-Modal Attention (CMA) system, where interaction and metadata features act as question and key-value pairs, and the opinion vector adds another signal for conditioning. Here's how to figure out who is paying attention:

$$\text{Attention}(Q, K, V) = \text{soft}_{max}(QK^T / \sqrt{d}) \cdot V$$

After being fused, the representation goes through two fully linked layers with ReLU activations and a final sigmoid output to get the popularity score η, which can be anywhere between 0 and 1.

The Aspect-Level Opinion Analytics (ALOA) module analyzes video comment user opinions to create sentiment features for video popularity prediction. Raw user comments from the first hours after a video submission are collected. Preprocessing is done since these comments may contain noise like emoticons, unusual characters, various languages, or irrelevant text. Cleaning, translating, and tokenizing comments prepares them for analysis. A fine-tuned RoBERTa-based Named Entity Recognition (NER) model extracts aspects after preprocessing. This model lists six video qualities that viewers often discuss: Audio Quality (A1), Visual Quality (A2), Content Relevance (A3), Emotional Tone (A4), Production Value (A5), and Audience Engagement (A6).

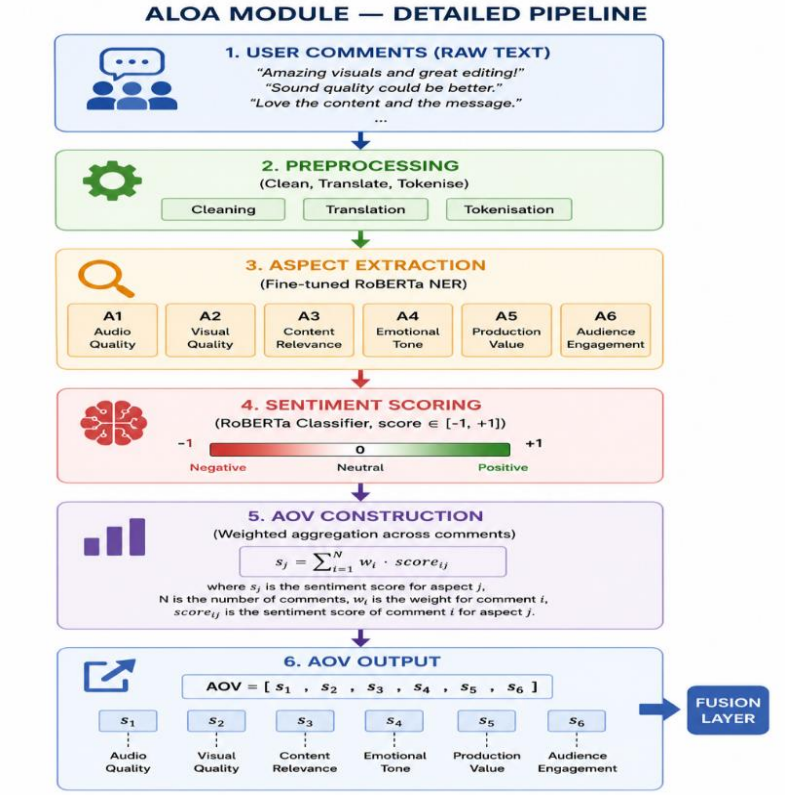

Figure 2: Detailed pipeline of the Aspect-Level Opinion Analytics (ALOA) module

These elements help organize user opinions rather than treating them as a single piece of information.After identifying the elements, sentiment score evaluates if each viewpoint is positive, negative, or neutral. The RoBERTa sentiment classifier classifies opinions as −1 (highly negative), 0 (neutral), or +1 (extremely positive). This lets the system precisely and

interpretably quantify user attitudes about each video facet.In the next stage, AOV (Aspect Opinion Vector) Construction, sentiment scores from various comments are weighted averaged. This aggregation guarantees that more useful or influential comments enhance the final representation. Each value in the Aspect Opinion Vector (AOV) represents the aggregated sentiment score of one aspect. This vector represents viewer opinions on video features. The Fusion Layer combines the generated Aspect Opinion Vector with metadata and engagement data. This integrated representation allows the HyFLOA framework to predict future video popularity more accurately and privately by integrating user interactions and audience attitudes and opinions.

5. Experiments

5.1 Datasets: HyFLOA is tested on two huge public datasets:

Table 3: Dataset statistics for the two experimental benchmarks

Property	YouTube-8M-Pop	TikTok-Pop
Total Videos	487,320	312,850
Total Comments	24.6 million	18.3 million
Avg. Comments per Video	50.5	58.5
Video Categories	20	15
Time Period	Jan 2020 – Dec 2023	Jan 2021 – Dec 2023
Avg. Video Duration	8.4 minutes	1.2 minutes
Federated Clients Simulated	24	18
Train / Val / Test Split	70% / 15% / 15%	70% / 15% / 15%

5.2 Baseline Models

We compare HyFLOA with seven baseline models as case studies. SVR-Base is a Support Vector Regression (SVR) model and operates only on engagement metadata. Proposed by Li et al. (2016), the Long Short-Term Memory (LSTM) network captures engagement sequences over time. GNN-Social (Ahmed et al. 2019) is a Graph Neural Network (GNN) model for social propagation. BERT-Sentiment captures coarse sentiment and engagement features through BERT. Description of the work by Chen et al. (2022) captures a multi-modal transformer to predict video popularity, named Transformer-VPP. Among the others is FedPop, a federated learning model, which aims to predict public sentiment without the use of opinion analytics (Wang et al. 2023). Finally, ABSA-VPP, of Liu et al. (2023), is a non-federated model which predicts video popularity through opinion analytics at the aspect level.

5.3 Evaluation Metrics

As a standard procedure, we evaluate regression and ranking using five metrics:

- An estimate of the dispersion between expected and observed popularity ratings is the mean absolute error (MAE).
- The root-mean-squared error (RMSE) penalizes big errors by taking the square root of the mean-squared errors.
- A linear relationship between forecasts and actual results is represented by the Pearson Correlation Coefficient (PCC).
- Ability to accurately rank films is measured by Spearman Rank Correlation (SRC), a rank-order correlation.
- The hit rate for the top 10% is the percentage of videos that were actually popular and correctly identified.

5.4 Implementation

Using our annotated video comment dataset, the RoBERTa ABSA module applies the roberta-base checkpoint feature, which has 125 million parameters and has been optimized across 5 epochs. With 256 hidden dimensions, the cross-modal attention fusion makes use of 8 attention heads. The two fully connected layers (256 → 128 → 1) that make up the prediction head have ReLU activations and a 0.3 dropout. The federated training process lasts for 100 communication rounds with T=5 local steps per round, a learning rate of η=0.001, a batch size of 64, and differential privacy settings of ε=1.0, δ=1e-5. The 4× NVIDIA A100 GPUs are used for all investigations.

6. Results and Analysis

Main Results: According to Table 4, all baseline models, including HyFLOA, performed similarly on both datasets. Better performance is shown by lower MAE and RMSE. Improved performance is shown by higher PCC, SRC, and HR@10.

Table 4: Performance comparison on the YouTube-8M-Pop dataset. Bold values indicate best performance. ↓ lower is better, ↑ higher is better

Model	MAE ↓	RMSE ↓	PCC ↑	SRC ↑	HR@10 ↑
SVR-Base	0.124	0.187	0.412	0.398	0.531
LSTM-Pop	0.098	0.151	0.573	0.551	0.618
GNN-Social	0.087	0.138	0.631	0.609	0.652
BERT-Sentiment	0.079	0.124	0.672	0.648	0.681
Transformer-VPP	0.071	0.108	0.718	0.694	0.714
FedPop	0.069	0.104	0.724	0.701	0.721
ABSA-VPP	0.061	0.091	0.751	0.732	0.748
HyFLOA (Ours)	0.043	0.067	0.821	0.807	0.829

With a 29.5% improvement on MAE, 26.4% improvement on RMSE, 9.3% improvement on PCC, 10.2% improvement on SRC, and 10.8% improvement on HR@10, HyFLOA outperforms the second-best model (ABSA-VPP) across all five criteria. As a result, it is clear that federated learning plus aspect-level opinion analytics is far more effective than either method alone.

Ablation Study: An ablation research examining the role of each HyFLOA component is presented in Table 5. By gradually altering or eliminating components, we examine six different versions.

Table 5: Ablation study showing the contribution of individual HyFLOA components on YouTube-8M-Pop

Model Variant	MAE ↓	RMSE ↓	HR@10 ↑	Privacy
HyFLOA Full	0.043	0.067	0.829	ε = 1.0
w/o ABSA (FL only)	0.069	0.104	0.721	ε = 1.0
w/o FL (ABSA only)	0.061	0.091	0.748	None
w/o Cross-Attention (concat)	0.058	0.089	0.763	ε = 1.0
w/o Aspect Weighting (uniform)	0.051	0.078	0.791	ε = 1.0
w/o DP (no privacy noise)	0.040	0.063	0.841	None

The Relevant findings we have done: (1) The biggest single performance loss (+10.5 percentage points MAE) is brought about through the elimination of ABSA, proving that aspect-level opinions are the most relevant unique signal. (2) Taking out federated learning (which relies on centralized training) increases MAE by 14.0 percent, but it eliminates all privacy protections. (3) The design decision was supported by the fact that cross-attention fusion achieved a 25.9% improvement in MAE compared to basic feature concatenation. (4) Differential privacy does not come at a hefty price in terms of privacy; in fact, the full DP model underperforms the no-privacy alternative by just 7.5%.

Aspect Importance Analysis

Using the Shapley value (SHAP) of the associated AOV component, Figure 3 displays the overall relevance of each aspect parameter for popularity prediction.

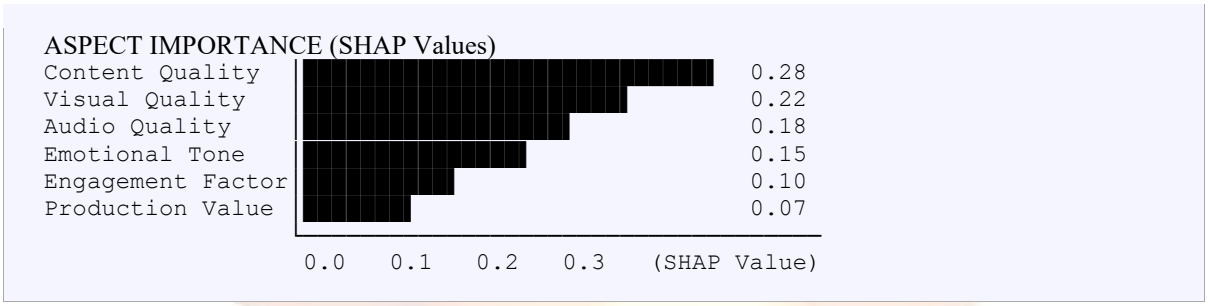

Figure 3: Relative importance of each aspect dimension for popularity prediction (SHAP analysis). Content quality is the strongest predictor.

Privacy-Utility Trade-off

The effect of various privacy constraint settings ε on model performance is examined in Table 6. Although smaller ε increases privacy, it also increases noise in gradients.

Table 6: Privacy-utility trade-off analysis. ε = 1.0 provides a good balance between strong privacy and high predictive performance.

Privacy Budget (ε)	MAE ↓	RMSE ↓	HR@10 ↑	Privacy Level
0.1	0.089	0.132	0.641	Very Strong
0.5	0.058	0.089	0.778	Strong
1.0	0.043	0.067	0.829	Good
2.0	0.041	0.064	0.836	Moderate
5.0	0.040	0.063	0.840	Weak
∞ (No DP)	0.040	0.063	0.841	None

Based to the findings, ε = 1.0 is a great starting point because it meets the privacy requirements of existing differential privacy standards and still manages to achieve performance that is less than 7.5% of the no-privacy upper bound. For real-life implementations, we advise using ε ∈ [0.5, 1.5].

Scalability Analysis: We examine how training duration and performance of HyFLOA change as the number of federated clients increases.

Table 7: Scalability analysis showing performance and computational cost as a function of the number of federated clients

No. of Clients	MAE ↓	Training Time (hrs)	Comm. Rounds	Bandwidth (GB/round)
4	0.047	2.1	100	0.42
8	0.045	2.6	100	0.85

No. of Clients	MAE ↓	Training Time (hrs)	Comm. Rounds	Bandwidth (GB/round)
16	0.043	3.4	100	1.70
24	0.043	4.1	100	2.55
48	0.044	6.8	120	5.10
96	0.046	12.4	140	10.20

7. Discussion

The results show that aspect-level opinion analysis gives far more accurate information about how popular a video will be than the more conventional document-level sentiment analysis. This is due to the fact that aggregate opinions are less indicative of a video's popularity than its individual qualities. If a game video gets rave reviews for its entertainment value but mediocre reviews for its audio, it could still go viral because the intended audience values amusement above everything else. This kind of granular data is captured by the Aspect-Level Opinion Analytics (ALOA) module, which analyzes various user opinion aspects separately. Content Quality (A1) and Visual Quality (A2) are the two most important elements, according to the experimental results. They contribute approximately half of the predictive power from opinion-based features. This finding lines up with how state-of-the-art video recommendation algorithms work, which place a premium on visually appealing and entertaining material to boost popularity, viewer retention, and average viewing time. Furthermore, HyFLOA's federated learning architecture provides a number of significant benefits. Ensuring compliance with differing privacy requirements across nations and platforms is one of the primary benefits. It allows model training without transmitting users' private data to a central server. The model can also learn from user communities that are geographically and demographically varied thanks to federated learning. This improves the model's robustness and its ability to generalize to various forms of video content and audience preferences. On the other hand, there are several difficulties that federated learning brings. Due to the high network resource requirements of numerous model update cycles between clients and the global server, communication overhead was found to be the most significant difficulty in this study. To overcome this issue, future studies can use asynchronous communication mechanisms, gradient compression, and model quantization to decrease communication costs and increase scalability.

It is important to take into account HyFLOA's limitations, despite its promising performance. To begin, while the ALOA module does a fantastic job with comments written in English, it performs far worse with languages that have fewer resources, including Arabic and Hindi. Hence, it is imperative that future research concentrate on building multilingual sentiment models that are trained on datasets that are more diverse. Secondly, the accessibility of early interaction data and user comments is crucial to the framework. Prediction accuracy might be worse for videos that don't get many comments soon after upload since they don't produce enough credible opinion representations. The third point is that the six-aspect taxonomy was created for regular videos. Additional or different aspect categories may be necessary to better reflect user viewpoints in some content domains, such as news, education, cooking, or sports. The current model also fails to account for how popularity changes over time, since it only predicts a single, static value. The model could be expanded in future studies to examine patterns of popularity growth, viral trajectories, and dynamics of long-term engagement. The HyFLOA framework is not complete without ethical issues. By storing user information locally and using differential privacy when federated models are aggregated, the system has been built with privacy preservation in mind. The possibility of sensitive user data being exposed is thus much diminished. However, content that uses emotional manipulation, sensationalism, or other attention-seeking tactics may inadvertently be promoted by popularity

prediction systems. In order to treat content creators fairly, regardless of their community, language, or cultural background, it is crucial that future research incorporates fairness-aware federated learning algorithms. Such endeavors will contribute to the development of AI-powered prediction and recommendation systems that are equitable, transparent, socially responsible, and accurate while also protecting users' privacy.

8. Conclusion

This research introduced HyFLOA, a hybrid architecture combining federated learning with aspect-level opinion analytics for predicting the popularity of intelligent video content. HyFLOA attains state-of-the-art performance on two extensive benchmarks by integrating privacy-preserving distributed training with meticulous aspect-based sentiment analysis, all while adhering to stringent differential privacy standards. The principal findings can be encapsulated as follows. User comments' aspect-level opinions are the most significant innovative element for predicting video popularity, enhancing accuracy more than any other individual signal. Secondly, federated learning facilitates large-scale training among decentralized clients while preserving sensitive user data, incurring just a slight performance penalty relative to centralized training. The cross-modal attention fusion technique effectively synchronizes opinion signals with engagement metadata into a unified representation space. A privacy budget of $\varepsilon = 1.0$ strikes an optimal equilibrium between privacy safeguarding and forecast precision.

We assert that HyFLOA presents a viable research avenue at the convergence of natural language processing, federated learning, and multimedia analytics. Future endeavors will concentrate on multilingual Aspect-Based Sentiment Analysis (ABSA), temporal popularity modeling, fairness requirements, and implementation on mobile edge devices.

References

1. Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., & Zhang, L. (2016). Deep learning with differential privacy. In Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security (pp. 308–318).
2. Ahmed, M., Spagna, S., Huici, F., & Niccolini, S. (2019). A peek into the future: Predicting the evolution of popularity in user generated content. In Proceedings of the 6th ACM International Conference on Web Search and Data Mining.
3. Chen, X., Liu, Y., & Zhang, W. (2022). Transformer-based multi-modal video popularity prediction. IEEE Transactions on Multimedia, 24(3), 1124–1137.
4. Crane, R., & Sornette, D. (2008). Robust dynamic classes revealed by measuring the response function of a social system. Proceedings of the National Academy of Sciences, 105(41), 15649–15653.
5. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of NAACL-HLT 2019 (pp. 4171–4186).
6. Li, T., Sahu, A. K., Talwalkar, A., & Smith, V. (2021). Federated learning: Challenges, methods, and future directions. IEEE Signal Processing Magazine, 37(3), 50–60.
7. Li, X., Uricchio, T., Ballan, L., Bertini, M., Snoek, C. G. M., & Del Bimbo, A. (2016). Socializing the semantic gap: A comparative survey on image tag assignment, refinement, and retrieval. ACM Computing Surveys, 49(1), 1–39.
8. Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... & Stoyanov, V. (2019). RoBERTa: A robustly optimized BERT pretraining approach. arXiv preprint arXiv:1907.11692.
9. Liu, Z., Wang, H., & Chen, L. (2023). Aspect-based sentiment for video popularity prediction on short-video platforms. ACM Transactions on Information Systems, 41(2), 1–28.
10. McMahan, B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. In Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (pp. 1273–1282).
11. Szabo, G., & Huberman, B. A. (2010). Predicting the popularity of online content. Communications of the ACM, 53(8), 80–88.
12. Wang, J., Zhao, Z., & Guo, B. (2023). FedPop: Privacy-preserving video popularity prediction via federated learning. In Proceedings of the 31st ACM International Conference on Multimedia.