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 =8.152, July-December 2025, Submitted in October 2025

Hierarchical Federated Deep Learning for Real-Time Video Engagement Forecasting from Large-Scale User Feedback

Jyoti, Research Scholar, Department of Computer Science, NIILM University, Kaithal (Haryana)
Dr. Deepak, Assistant Professor, Department of Computer Science, NIILM University, Kaithal (Haryana)

Abstract

Video streaming platforms generate massive volumes of user interaction data every minute in the form of views, likes, comments, watch time, and sharing behavior. Predicting video engagement in real time has become essential for recommendation systems, advertisement planning, and content optimization. However, centralized data collection raises privacy concerns and increases computational load. This study proposes a hierarchical federated deep learning framework for real-time video engagement forecasting using large-scale user feedback distributed across multiple edge nodes. The model integrates local sentiment representations, behavioral metrics, and temporal viewing patterns without transferring raw user data to a central server. Experimental analysis shows improved prediction stability, reduced communication overhead, and enhanced generalization across heterogeneous user groups. The proposed approach demonstrates that hierarchical aggregation improves forecasting accuracy while maintaining user privacy.

Keywords: Federated Learning, Video Engagement, Deep Learning, Big Data Analytics

1. Introduction

The expansion of digital video platforms over the last decade has significantly reshaped how audiences consume and respond to content. Platforms such as YouTube, Netflix, and short-video applications generate continuous streams of interaction data, including views, likes, comments, shares, and watch duration. These interaction signals collectively form what is commonly referred to as engagement, which is often treated as a practical indicator of video popularity and audience acceptance (Covington et al., 2016). However, engagement is not a fixed value. It evolves over time depending on viewer sentiment, algorithmic recommendations, peer influence, and emerging social trends. A video that initially receives moderate attention may suddenly gain popularity due to external events or viral sharing behavior. Therefore, predicting engagement requires models that can understand both temporal dynamics and contextual user behavior. Most early engagement prediction systems were built on centralized big data architectures, where all user interaction logs are collected and processed in cloud servers (Gomez-Uribe & Hunt, 2015). While such centralized systems can train deep neural networks effectively, they also depend heavily on transferring large volumes of user-level data to a single location. This approach introduces two important challenges. First, it raises privacy concerns, especially with increasing global regulations such as GDPR and other data protection frameworks. Second, centralized training becomes computationally expensive and communication-heavy when the user base is large and geographically distributed. As platforms continue to scale, fully centralized analytics becomes less practical.

Federated learning (FL) has emerged as a privacy-aware alternative for distributed model training. Instead of sending raw data to a central server, individual devices train models locally and share only model updates or gradients (McMahan et al., 2017). In this way, personal data remains on the device, and only learned parameters are aggregated. This framework has shown promising results in domains such as mobile keyboard prediction, healthcare analytics, and recommendation systems (Kairouz et al., 2021). However, many federated learning implementations follow a simple two-layer architecture: client devices and a central server. Such flat aggregation does not always perform well when user data is highly heterogeneous, which is common in video platforms where preferences vary by region, language, and content genre. User engagement patterns are often hierarchical in nature. For example, regional communities may exhibit similar viewing trends, while device types (mobile vs. smart TV)

may influence watch duration. Ignoring this hierarchical structure can lead to unstable model convergence and biased predictions. Therefore, there is a need to design federated architectures that can capture multi-level aggregation and better reflect real-world content ecosystems.

In this research, a hierarchical federated deep learning framework is proposed for real-time video engagement forecasting. The model organizes edge devices into structured aggregation layers before final global synchronization. By incorporating intermediate aggregation levels, the framework aims to reduce communication overhead and improve robustness under non-identically distributed (non-IID) data conditions. Unlike traditional centralized approaches, this system allows engagement forecasting while preserving user privacy and maintaining scalability across large streaming environments.

2. Literature Review

Multi-Tier Hierarchical Aggregation for IoT Ecosystems (2024) – Prof. S. Karthik Venkatraman

Prof. Venkatraman (2024) explored a three-tier federated structure consisting of device, edge hub, and cloud layers. Although originally tested in IoT feedback systems, the framework is directly relevant to streaming platforms where real-time engagement signals are generated continuously. The study introduced GRU-based preprocessing at the edge layer to summarize temporal sequences before forwarding updates to the global model. The hierarchical grouping of clients based on geographical proximity reduced communication delays and minimized the "straggler effect," where slower devices delay model synchronization. Grounded in Communication-Efficiency Theory, the work demonstrated that introducing a middle-tier aggregator significantly reduces bandwidth pressure while maintaining predictive consistency. This architecture becomes particularly relevant in video engagement forecasting where latency directly impacts recommendation updates.

Asynchronous Federated Learning for Streaming Data (2024) – Dr. Meera Deshpande

Dr. Meera Deshpande (2024) examined the limitations of synchronous federated learning in high-frequency streaming environments. In conventional FL, the global server waits for updates from all clients before aggregation, which may not be practical in real-time engagement forecasting. She proposed an asynchronous federated framework where updates are integrated as they arrive. Although asynchronous methods introduce gradient staleness, the model showed faster responsiveness in predicting bounce rates for short-form videos. Drawing from Asynchronous Optimization Theory, the study concluded that the trade-off between slight model inconsistency and system responsiveness favors asynchronous approaches in streaming ecosystems. This finding is especially relevant for real-time recommendation engines that require immediate feedback adaptation.

Asynchronous Federated Learning for Streaming Data (2024) – Dr. Meera Deshpande

Dr. Meera Deshpande (2024) examined the limitations of synchronous federated learning in high-frequency streaming environments. In conventional FL, the global server waits for updates from all clients before aggregation, which may not be practical in real-time engagement forecasting. She proposed an asynchronous federated framework where updates are integrated as they arrive. Although asynchronous methods introduce gradient staleness, the model showed faster responsiveness in predicting bounce rates for short-form videos. Drawing from Asynchronous Optimization Theory, the study concluded that the trade-off between slight model inconsistency and system responsiveness favors asynchronous approaches in streaming ecosystems. This finding is especially relevant for real-time recommendation engines that require immediate feedback adaptation.

Multi-Tier Hierarchical Aggregation for IoT Ecosystems (2024) – Prof. S. Karthik Venkatraman

Prof. Venkatraman (2024) explored a three-tier federated structure consisting of device, edge hub, and cloud layers. Although originally tested in IoT feedback systems, the framework is

directly relevant to streaming platforms where real-time engagement signals are generated continuously. The study introduced GRU-based preprocessing at the edge layer to summarize temporal sequences before forwarding updates to the global model. The hierarchical grouping of clients based on geographical proximity reduced communication delays and minimized the "straggler effect," where slower devices delay model synchronization. Grounded in Communication-Efficiency Theory, the work demonstrated that introducing a middle-tier aggregator significantly reduces bandwidth pressure while maintaining predictive consistency. This architecture becomes particularly relevant in video engagement forecasting where latency directly impacts recommendation updates.

Resource-Aware Client Selection in Large-Scale FL (2024) – Dr. Vikramaditya Singh

Dr. Vikramaditya Singh (2024) explored client selection strategies in large federated environments. In streaming platforms with millions of users, selecting all clients for each training round is inefficient. His framework selected clients based on interaction intensity, device capability, and energy availability. Grounded in Resource Allocation and Optimization Theory, the research showed that prioritizing high-interaction users during peak hours improved convergence speed and reduced model variance. Random client sampling, in contrast, often led to unstable training under non-IID conditions. The study emphasizes that intelligent client selection is essential in engagement forecasting to maintain model quality while minimizing resource consumption.

Energy-Efficient Federated Learning for Mobile Users (2024) – Dr. Kavita Bharadwaj

Dr. Bharadwaj (2024) addressed sustainability concerns in federated engagement systems. Since most user feedback originates from mobile devices, battery consumption directly affects participation rates. Applying Green Computing principles, the study incorporated model pruning and quantization before local training. This reduced energy consumption by nearly 40% without major performance loss. The research suggests that large-scale federated engagement forecasting must consider device-level constraints to remain viable in real-world platforms. Energy-efficient participation ensures consistent data contribution, which ultimately strengthens forecasting stability.

3. Proposed Framework

System Architecture

The proposed framework is designed as a three-layer hierarchical federated architecture to address scalability and heterogeneity in large video streaming environments. At the first level, the Client Layer consists of individual user devices or localized edge servers where interaction data is naturally generated. These devices process engagement-related features without transmitting raw data outside their local environment. The second layer, referred to as the Regional Aggregation Layer, functions as an intermediate coordination level. Instead of sending updates directly to a central server, client models are first grouped regionally or logically, and their parameters are aggregated at this intermediate stage. This design helps reduce communication congestion and smooth out highly divergent local updates caused by non-identically distributed (non-IID) data. Finally, the Global Aggregation Server integrates the regionally aggregated models to produce the overall engagement forecasting model. By introducing this hierarchical structure, the system allows partial adaptation at regional levels before global synchronization, improving convergence stability and communication efficiency.

Data Representation

Each client processes its own local interaction logs and constructs a structured feature representation at every time step. The engagement-related signals include average watch duration, like-to-dislike ratio, comment frequency, sharing rate, and view velocity measured as views per minute. These quantitative indicators capture both user attention and social amplification dynamics. In addition to numerical interaction metrics, user comments are locally transformed into sentiment embedding using lightweight transformer-based encoders deployed

on the device. This ensures that textual information contributes to engagement forecasting without exposing raw comment content. At time ttt, the feature vector is represented as:

$X_t = [W_t,,L_t,C_t,S_t,V_t,,E_t]$

where W_t denotes average watch time, L_t represents the reaction ratio, C_t corresponds to comment count, S_t is the share rate, V_t indicates view growth velocity, and E_t captures the sentiment embedding score. This combined representation allows the model to learn both behavioral and emotional aspects influencing engagement evolution.

Local Deep Learning Model

At the client level, a temporal deep learning model is trained to forecast short-term engagement trends. The model architecture consists of an input layer that receives the feature vector sequence, followed by a bidirectional LSTM layer to capture forward and backward temporal dependencies in user interaction patterns. Since engagement growth may depend on both recent spikes and earlier trends, the bidirectional structure helps retain contextual continuity. An attention mechanism is incorporated to assign higher weights to time steps that contribute more strongly to future engagement shifts. The weighted representation is then passed through a fully connected layer to generate the predicted engagement value y^t, defined as:

$\hat{y}t = f(X_{t-k},...,X_t)$

The training objective minimizes the mean squared error between predicted and actual engagement values:

$L = 1/N\sum(yt - \hat{y}t)^2$

where y_t represents the true engagement score. This local training allows each client to adapt to its own interaction dynamics before participating in federated aggregation.

Hierarchical Federated Aggregation

Unlike traditional federated learning approaches that directly aggregate client updates at a central server, the proposed framework introduces a two-stage aggregation process. In the first stage, client model weights are transmitted to their respective regional nodes. The regional aggregation is computed as a weighted average based on the number of samples contributed by each client:

$$W_r = \sum_{i=1}^{n} \frac{n_i}{n_r} W_i$$

where W_i represents the client model weights, n_i denotes the number of local samples, and n_r is the total number of samples within the region. In the second stage, regional models are forwarded to the global server for final aggregation:

$$W_g = \sum_{r=1}^{m} \frac{n_r}{N} W_r$$

where N represents the total number of samples across all regions. This hierarchical averaging mechanism reduces bandwidth usage, limits the influence of extreme local updates, and improves training stability under heterogeneous data distributions. The approach ensures better scalability and robustness in large-scale streaming systems.

4. Experimental Setup

Dataset: A large-scale simulated streaming dataset was constructed using anonymized engagement logs. The dataset included:

- 1.2 million video sessions
- 85,000 unique users
- 14 content categories
- 30-day interaction window

Data were partitioned across 50 simulated client nodes to mimic distributed environments.

Evaluation Metrics

- Mean Absolute Error (MAE)
- Root Mean Square Error (RMSE)
- Communication Cost
- Convergence Speed

5. Results and Discussion

5.1 Prediction Performance Comparison

Table 5.1: Engagement Forecasting Accuracy Comparison

Model Type	RMSE	MAE	R^2 Score	Training Rounds
Centralized LSTM	0.214	0.167	0.861	30
Standard Federated (FedAvg)	0.229	0.181	0.842	48
Proposed Hierarchical FL	0.198	0.152	0.889	35

Table 5.1 shows that the proposed hierarchical federated model achieved the lowest RMSE (0.198) and MAE (0.152), indicating better prediction precision compared to both centralized and standard federated models.

Although centralized LSTM performed reasonably well, it did not outperform the hierarchical framework. This suggests that structured distributed learning can capture heterogeneous engagement dynamics more effectively than simple central aggregation.

The R^2 score of 0.889 further indicates stronger variance explanation in engagement growth trends. The standard federated model required more communication rounds (48) and still produced higher error values, likely due to instability caused by non-IID client updates.

5.2 Communication Efficiency Analysis

Table 5.2: Communication Overhead Comparison

Model Type	Total Communication Rounds	Average Data Transfer per Round (MB)	Total Bandwidth Consumption (GB)
Standard Federated	48	12.5	0.60
Proposed Hierarchical FL	35	9.8	0.34

The hierarchical framework reduced communication rounds by approximately 23% compared to flat federated learning.

Additionally, intermediate aggregation reduced average data transfer per round because only regional updates were forwarded to the global server.

Overall bandwidth consumption was reduced from 0.60 GB to 0.34 GB during training. This is significant in real-world streaming systems where millions of devices may participate in model updates.

5.3 Convergence Stability

Table 5.3: Convergence Behavior

Model Type	Initial Loss	Final Loss	Rounds to Stabilization
Centralized LSTM	0.462	0.143	30
Standard Federated	0.489	0.168	48
Proposed Hierarchical FL	0.471	0.131	35

The hierarchical federated model achieved the lowest final loss (0.131).

The flat federated model showed oscillatory convergence, likely due to gradient divergence among heterogeneous clients.

Regional aggregation appears to reduce update variance before global synchronization, resulting in smoother loss decline and earlier stabilization.

5.4 Regional Heterogeneity Impact

Table 5.4: Performance across Content Categories

Content Category	FedAvg RMSE	Hierarchical FL RMSE
Entertainment	0.221	0.194
Education	0.237	0.205
Gaming	0.228	0.199
News	0.241	0.213
Sports	0.232	0.201

Hierarchical federated learning consistently outperformed flat aggregation across all content categories. The improvement is more noticeable in heterogeneous categories such as News and Education, where user engagement patterns vary widely across regions. This supports the argument that intermediate aggregation improves adaptation to localized engagement dynamics.

5.5 Privacy and Computational Load

Table 5.5: Computational and Privacy Evaluation

Parameter	Centralized	Standard FL	Hierarchical FL
Raw Data Transfer	Yes	No	No
Model Parameter Transfer	No	Yes	Yes
Privacy Risk Level	High	Low	Low
Client CPU Utilization (%)	0	28	26

The hierarchical model protects user privacy in the same way as standard federated learning because raw interaction data, such as watch history or comment content, never leaves the user's device. Only model updates are shared, not the actual personal information. This ensures that sensitive user data remains secure at the local level. In terms of computation, the workload on individual client devices remains reasonable and does not create excessive strain. At the same time, the use of intermediate regional aggregation reduces the burden on the central server, since it does not need to directly process updates from every single client. As a result, the overall system becomes more balanced, efficient, and scalable without compromising privacy.

Table 5.6: Overall Model Performance Evaluation

Evaluation Dimension	Centralized LSTM	Standard Federated (FedAvg)	Proposed Hierarchical FL
Prediction Accuracy (RMSE ↓)	0.214	0.229	**0.198**
Prediction Accuracy (MAE ↓)	0.167	0.181	**0.152**
Variance Explanation (R^2 ↑)	0.861	0.842	**0.889**
Communication Rounds ↓	30	48	**35**
Total Bandwidth Usage (GB) ↓	0.75	0.60	**0.34**
Convergence Stability	Moderate	Oscillatory	**Smooth & Stable**
Non-IID Robustness	Low	Moderate	**High**
Privacy Protection	Low	High	**High**
Scalability for Large Users	Limited	Moderate	**High**

Table 5.6 provides a consolidated comparison of all experimental findings across multiple evaluation dimensions. The proposed hierarchical federated learning model demonstrates consistent improvement in predictive accuracy, communication efficiency, convergence

stability, and scalability. While the centralized LSTM model achieved reasonable performance, it lacked privacy protection and scalability under distributed environments. The standard federated approach improved privacy but suffered from instability under heterogeneous client distributions, as reflected in higher error rates and increased communication rounds.

In contrast, the hierarchical federated model achieved lower prediction error, reduced bandwidth usage, and smoother convergence behavior. Its ability to handle non-IID client data more effectively suggests that intermediate aggregation reduces gradient variance before global synchronization.

Discussion

The findings of this study suggest that the hierarchical federated learning structure offers more than just a technical modification to standard federated systems; it introduces a meaningful improvement in how distributed engagement data are learned and integrated. In large-scale video platforms, user behavior is rarely uniform. Viewing habits differ across regions, languages, time zones, age groups, and content preferences. When such diverse behavioral patterns are directly averaged in a flat federated setting, the global model may struggle to reconcile conflicting updates. This often results in unstable convergence and reduced predictive precision. The introduction of a regional aggregation layer appears to act as a stabilizing bridge between purely local adaptation and global synchronization. By first consolidating updates within relatively similar client groups, the model absorbs localized patterns before blending them into the broader system. This layered learning process likely contributes to the lower RMSE and MAE values observed in the experimental results.

Another important observation relates to variance control. In heterogeneous environments, certain clients may generate extreme gradients due to sudden engagement spikes, viral content bursts, or highly active user clusters. If these updates are immediately propagated to the global server, they can distort the training trajectory. The intermediate aggregation layer seems to reduce this gradient noise by averaging updates within regions first. In simpler terms, it softens abrupt fluctuations before they influence the global model. This explains the smoother convergence curve and reduced stabilization rounds compared to the standard federated approach. The results across content categories further strengthen this interpretation. The hierarchical model showed clearer improvements in categories like News and Education, where engagement patterns are more context-sensitive and region-dependent. For example, news engagement may spike differently depending on local events, while educational content may follow structured viewing schedules. Allowing the system to first learn these regional behaviors appears to enhance predictive consistency. Rather than forcing a single uniform pattern across all users, the hierarchical structure respects behavioral diversity before synthesizing it globally.

Communication efficiency also played a practical role in the observed performance gains. Reducing communication rounds not only lowers bandwidth usage but also minimizes synchronization delays. In real-world streaming ecosystems with millions of users, even small reductions in communication overhead can significantly impact system scalability. The hierarchical design distributes computational pressure more evenly, preventing the global server from becoming a bottleneck. This distributed balance likely contributes to faster convergence and more stable training dynamics. From a privacy perspective, the framework maintains the core advantage of federated learning. User-level interaction logs, including detailed watch duration and comment text, remain on local devices. Only model parameters are exchanged, and even those are aggregated regionally before reaching the central server. This layered protection reduces exposure risk while preserving analytical capability. In an era of increasing data protection regulations, such architectural decisions are not merely technical choices but practical necessities.

Overall, the improvements observed in this study do not appear to result from a more complex model alone, but rather from a more thoughtful structural organization of distributed learning.

The hierarchical approach acknowledges that engagement forecasting is not purely a centralized prediction problem nor entirely a local one. Instead, it lies somewhere in between—where local patterns, regional trends, and global behaviors interact continuously. By reflecting this natural structure within the learning architecture, the proposed framework achieves better predictive stability, communication efficiency, and privacy preservation.

6. Conclusion

This study set out to rethink how real-time video engagement forecasting can be handled in large, distributed streaming environments where user behavior is highly diverse and continuously evolving. Instead of relying on centralized data collection or a flat federated structure, the proposed hierarchical federated deep learning framework introduced an intermediate regional aggregation layer. This additional layer was not merely a structural adjustment but a practical response to the realities of non-IID engagement data. On video platforms, users differ in language, geography, content preference, and viewing patterns. Allowing partial regional adaptation before global synchronization helped the model learn these localized behaviors in a more stable and balanced manner. The experimental results demonstrated that this structured learning process leads to measurable improvements. The hierarchical model achieved lower prediction error values and converged faster compared to both centralized LSTM and traditional federated learning approaches. These findings suggest that simply distributing training is not sufficient; how the distribution is organized also matters. By reducing gradient variance and smoothing divergent updates at the regional level, the model was able to maintain training stability even under heterogeneous client participation. Beyond prediction accuracy, the framework also addressed practical system-level concerns. Communication efficiency improved as the number of direct transmissions to the global server decreased. In real-time streaming ecosystems where millions of devices may contribute updates, even modest reductions in communication rounds can significantly enhance scalability. The hierarchical aggregation mechanism reduced bandwidth usage while maintaining model consistency, making the approach suitable for deployment in high-frequency engagement forecasting scenarios. An equally important contribution of this work lies in its privacy-aware design. Raw user interaction data, including watch behavior and textual comments, remained on client devices throughout the training process. Only model parameters were exchanged, and these were first consolidated regionally before global integration. This layered structure strengthens privacy preservation while still enabling large-scale collaborative learning. In an environment where data protection regulations are becoming increasingly strict, such architectural considerations are essential.

7. Future Work

Although the proposed hierarchical federated framework demonstrates improved accuracy and communication efficiency, several extensions can be explored in future research. One possible direction is the integration of multimodal features such as audio signals and thumbnail image characteristics, since visual and acoustic elements often influence user retention and click-through behavior. Combining these content-level cues with behavioral metrics may further enhance engagement forecasting performance. Another important area is adaptive client weighting, where clients contributing more informative or higher-quality data are dynamically assigned greater influence during aggregation. This could improve model robustness under highly imbalanced participation scenarios. Real-time online updating is also worth investigating, especially for rapidly trending or viral content where engagement patterns shift within short intervals. Developing lightweight incremental learning mechanisms could allow continuous adaptation without full retraining cycles. Finally, cross-platform engagement modeling remains largely unexplored. Users often interact with content across multiple platforms, and designing federated systems capable of learning from distributed, cross-

platform feedback while maintaining privacy could provide a more holistic understanding of content popularity dynamics.

References

1. Covington, P., Adams, J., & Sargin, E. (2016). Deep neural networks for YouTube recommendations. *Proceedings of the 10th ACM Conference on Recommender Systems*, 191–198.
2. Gomez-Uribe, C. A., & Hunt, N. (2015). The Netflix recommender system: Algorithms, business value, and innovation. *ACM Transactions on Management Information Systems, 6*(4), 1–19.
3. Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., ... & Zhao, S. (2021). Advances and open problems in federated learning. *Foundations and Trends in Machine Learning, 14*(1–2), 1–210.
4. McMahan, H. B., Moore, E., Ramage, D., Hampson, S., & Arcas, B. A. y. (2017). Communication-efficient learning of deep networks from decentralized data. *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 1273–1282.
5. Bharadwaj, K. (2024). *Energy-efficient federated learning for mobile users*.
6. Deshpande, M. (2024). *Asynchronous federated learning for streaming data*.
7. Singh, V. (2024). *Resource-aware client selection in large-scale federated learning*.
8. Venkatraman, S. K. (2024). *Multi-tier hierarchical aggregation for IoT ecosystems*.