Vol. 12 No.1 Tahun 2026, pp 68-76
Jurnal Teknologi dan Manajemen Informatika
http://jurnal.unmer.ac.id/index.php/jtmi
P-ISSN: 1693-6604 E-ISSN: 2580-8044

Optimizing Endless Runner Game Player Performance Using a Hybrid GMF-MLP Recommendation System Based on Neural Collaborative Filtering

Hendry Cahyo Gunawan[1], Fresy Nugroho[2], Muhammad Ainul Yaqin[3], Suhartono[4], Yunifa Miftakhul Arif[5]

Magister Informatika, Sains dan Teknologi, UIN Maulana Malik Ibrahim

Article Info

Article History
Received: 23-02-2026
Revised : 01-04-2026
Accepted: 29-06-2026

Keywords
Endless Runner;
General Matrix Factorization;
Multi-Layer Perceptron;
Neural Collaborative Filtering;
Game Recommendation System;

Corresponding Author
Hendry Cahyo Gunawan,
UIN Maulana Malik Ibrahim,
240605210033@student.uin-malang.ac.id

ABSTRACT

Endless Runner games feature exponentially increasing difficulty as distance grows, often causing character failure and player frustration, largely because players struggle to select power-ups suited to their current difficulty context. While prior recommender-system research has mostly focused on purchase prediction for monetization, this study instead builds a personalized item recommendation system aimed at reducing failure and maximizing scores. We propose a hybrid Neural Collaborative Filtering (NCF) architecture combining General Matrix Factorization (GMF), which captures linear preferences, with a Multi-Layer Perceptron (MLP), which models non-linear interactions between playstyle and failure context (cause of death). The model was trained on 10,000 gameplay activity logs containing features such as jump count, obstacles avoided, and death cause. Over 20 training epochs, both training and validation accuracy converged to approximately 0.88–0.90, with a negligible gap between the two curves, indicating minimal overfitting. These results demonstrate that integrating GMF and MLP effectively produces recommendations adaptive to dynamic gameplay conditions.

INTRODUCTION

The mobile game industry has currently evolved from mere entertainment into complex, data-driven systems [1]. One of the genres dominating the market is the Endless Runner (e.g., *Subway Surfers*, *Temple Run*). The core challenge of this genre lies in its steep difficulty curve; the longer the play duration, the faster the gameplay tempo becomes [2]. Preliminary data indicates that players frequently stop playing not due to boredom, but because the game becomes excessively difficult at certain levels without the assistance of appropriate items [3].

Recommender systems in prior studies have generally focused on purchase prediction (monetization) within online stores using standard Collaborative Filtering methods [4], [5], [6], [7], [8], [9]. These recommendation models rely solely on conventional matrix factorization, which is only capable of capturing linear relationships and fails to model complex interactions, such as non-linear features between playstyle and item utility [6], [10]. Specifically, previous research has not utilized in-game objects as the core research subject, thereby omitting critical gameplay variables such as death cause and speed level [11].

The application of deep learning to uncover potential spatial and temporal features from multi-source data in Basketball games was recently introduced as a research subject in 2024 [12]; that study highlighted that the primary challenge is enhancing user experience through effective recommender systems, such as in-game item recommendations, appropriate level selection, or features that can boost player engagement. Consequently, a literature gap remains, as no study has specifically applied GMF-MLP to games to address the churn rate caused by mechanical gameplay difficulty (Endless Runner) by leveraging death cause data.

This study proposes a novelty by implementing a hybrid Deep Learning architecture, namely Neural Collaborative Filtering (NCF), which integrates General Matrix Factorization (GMF) and Multi-Layer Perceptron (MLP) [13]. The primary focus of this research is on optimizing player performance (score and distance). To clarify the positioning of this study, a comparison with relevant prior research has been conducted. Table 1 highlights the research gaps addressed by this study. Table 1 shows the gaps addressed by this research.

Table 1. Comparison of Prior Research (State of the Art)

Author (Year)	Dataset	Method	Limitation	Contribution Gap
Lin, Han Bao, Muren Kang, Chenran (2024)	Multisource Data	Statistical approaches	Potential spatial and temporal features of multi-source data	Applying Deep Learning only to reveal potential spatial and temporal features from Multisource Data
Lomanto et al. (2023)	Data *Game* A (e.g., *item* purchase)	Collaborative Filtering (CF)	Depends on user / item data density (sparsity) and cannot predict non-linear interactions.	Does not consider dynamic context (gameplay) or use Deep Learning architecture.
Koren et al. (2009)	Netflix (Film)	Matrix Factorization (MF)	Only capturing linear relationships is not suitable for dynamic game data.	Not yet implemented Deep Learning (NFC) to model non-linear interactions in gameplay.
Drachen et al. (2013)	Tomb Raider (Action)	Clustering & Statistical Analysis	Descriptive analysis only, does not provide automated recommendations.	Does not focus on adaptive item recommendations for performance optimization.
He et al. (2017)	E-commerce	NeuMF (GMF + MLP)	The shopping domain does not take into account gameplay variables such as death cause or player skill.	Not yet implemented GMF-MLP in the Endless Runner genre with the context of failure (Death Cause & Speed Level).

Based on Table 1, there has been no research that specifically applies GMF-MLP to handle churn rates due to game mechanical difficulty (Endless Runner) by utilizing cause of death data.

METHOD

Dataset Description

The dataset used in this study is a synthetic game activity log with 10,000 entries simulated in a controlled manner to reflect the dynamics of Endless Runner gameplay. The purpose of utilizing synthetic data is to evaluate the performance of the model architecture against a large variety of playstyles and specific failure contexts. The target outcome, a binary interaction variable, is clearly defined as follows: a value of 1 (Success/Relevant) indicates that the player successfully selected or applied a support item appropriate to the current difficulty context, thereby improving performance/score, while a value of 0 (Failure/Irrelevant) indicates that the player ignored the item or selected an item that was not contextual to the cause of death.To ensure reproducibility of the study, all experiments were conducted with a random seed of 42.

The data was divided into three parts with a Train-Validation-Test split of 80% for training (8,000 logs), 10% for validation (1,000 logs), and 10% for final testing (1,000 logs). The model was trained for 20 epochs with a batch size of 64 using the Adam Optimizer at a learning rate of 0.001. The GMF and MLP paths used an embedding size of 16. The hidden layers in the MLP path were arranged in a three-level architecture with dimensions [32, 16, 8] respectively. To avoid overfitting, Dropout regularization with a level of 0.2 was applied at the end of each MLP layer before the Sigmoid Fusion process was performed.

Evaluation Metrics for Recommendation Performance

To evaluate the effectiveness of the hybrid GMF-MLP architecture in delivering precise supporting item recommendations, this study utilizes the **Precision@K** metric. Testing was conducted across varying values of $K \in \{3, 5, 10\}$ to simulate the limited user interface (UI) slot constraints within Endless Runner games. The formula is expressed as follows:

$$\frac{P@K \;=\; |Relevant\ Items \cap Top - K\ Recommended|}{K} \tag{1}$$

The detailed calculation of **Precision@K** (P@K) at levels $K \in \{3, 5, 10\}$ references the utilized dataset structure. Based on the dataset description, the target output is a binary interaction variable, where a value of 1 indicates that the item was successfully selected/relevant, while a value of 0 indicates that the item failed/was not selected. The system evaluates recommendation matches for combinations of user identities (user_ID 1–30) and objects (item_ID 1–10). The general equation for a single user (u) is:

$$P_u@K \;=\; \frac{\sum_{i=1}^{K} rel_u(i)}{K} \tag{2}$$

Where $rel_u(i)$represents the binary value of the item ranked at position $(i)$for user (u):

- $rel_u(i) = 1$, if the item at rank (i)has an interaction = 1 (success/relevant).
- $rel_u(i) = 0$, if the item at rank (i) has an interaction = 0 (failed/irrelevant).

To obtain the aggregate value across the entire sample population ($|U| = 30$ pengguna), the mean metric (*Mean Precision@K*) is applied:

$$Precision@K = \frac{1}{|U|}\sum_{u=1}^{|U|} P_u\,@K \tag{3}$$

The NeuMF model predicts the probabilities across all objects (item_ID 1-10) for *User 7*. Once the probabilities are computed via *Sigmoid Fusion*, the model sorts the items in descending order based on their values. It is assumed that the ranked recommendation results along with their interaction ground truth data are as follows:

Table 2. Assumed Recommendation Ranking Results

Rank (i)	Item ID	Prediction Probability	Actual Status (interaction)
1	7	0.8950	1 (Relevant)
2	3	0.8420	1 (Relevant)
3	1	0.7110	0 (Irrelevant)
4	5	0.6580	1 (Relevant)

5	9	0.5254	0 (Irrelevant)
6	2	0.4120	1 (Relevant)
7	4	0.3390	0 (Irrelevant)
8	6	0.2110	0 (Irrelevant)
9	8	0.1050	0 (Irrelevant)
10	10	0.0540	0 (Irrelevant)

Based on the simulation table above, the Precision value calculations for *User 7* at each threshold $\boldsymbol{K}$ are defined as follows:

a. Calculation of Precision@3 ($\boldsymbol{K = 3}$)

- Top-3 Recommendation List: [Item 7, Item 3, Item 1]
- Relevance Evaluation:
  1. Rank 1 (Item 7) $>$ = 1 (Yes)
  2. Rank 2 (Item 3) $>$ = 1 (Yes)
  3. Rank 3 (Item 1) $>$ = 0 (No)
- Number of Relevant Items in Top-3 = 1 + 1 + 0 = 2
- Calculation:

$$P_{User7}@3 = \frac{2}{3} = 0.6667\ (66.67\%) \qquad (4)$$

b. Calculation of Precision@5 ($\boldsymbol{K = 5}$)

- Top-5 Recommendation List: [Item 7, Item 3, Item 1, Item 5, Item 9]
- Relevance Evaluation:
  1. Ranks 1 to 3 $>$ [1, 1, 0]
  2. Rank 4 (Item 5) $>$ interaction = 1 (Yes)
  3. Rank 5 (Item 9) $>$ interaction = 0 (No)
- Number of Relevant Items in Top-5 = 1 + 1 + 0 + 1 + 0 = 3
- Calculation:

$$P_{User7}@5 = \frac{3}{5} = 0.6000\ (60.00\%) \qquad (5)$$

c. Calculation of Precision@10 ($\boldsymbol{K = 10}$)

- Top-10 Recommendation List: All items [Item 1 to Item 10]
- Relevance Evaluation: Summing all successfully labeled items (interaction = 1) recommended by the model, which are Item 7, Item 3, Item 5, and Item 2.
- Number of Relevant Items in Top-10 = 1 + 1 + 0 + 1 + 0 + 1 + 0 + 0 + 0 + 0 = 4
- Calculation:

$$P_{User7}@10 = \frac{4}{10} = 0.400\ (40.00\%) \qquad (6)$$

System Architecture (GMF and MLP)

This recommendation system is built using the NeuMF framework, where input processes go through two parallel paths before being combined. To explain how this system is built, here's a diagram of the Neural Matrix Factorization (NeuMF) architecture :

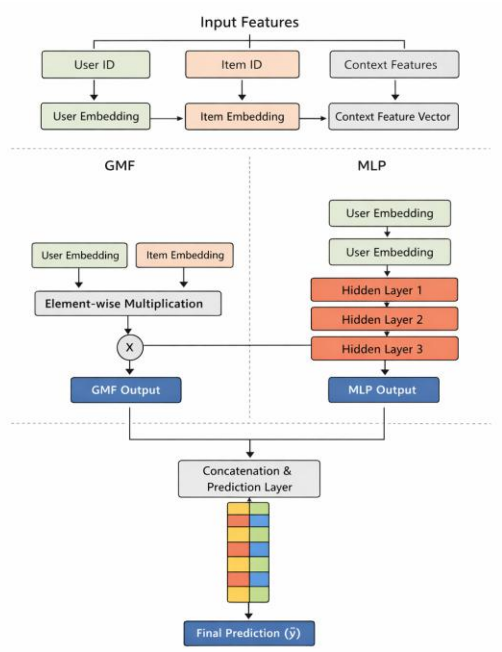

Figure 1. Neural Matrix Factorization (NeuMF) Flowchart

Figure 1 shows the *Neural Matrix Factorization* (NeuMF) structure used to build an item recommendation system in the Endless Runner game. The architecture consists of a *General Matrix Factorization* (GMF) to model the linear interaction between users (user_ID : P_u) and items (item_ID : Q_i), and a Multi-Layer Perceptron (MLP) to capture non-linear interactions by considering contextual features such as speed level and death cause. The outputs of both paths are combined to produce a predicted probability of an item matching the player's game conditions.

Model Simulation

Based on the results of model training for 20 epochs, the following vector parameters were obtained:

a. Vector User 7 (P_u): [0.37, 0.95, 0.73, 0.59]
   Vector User 7 is an aggressive type of player who likes a fast game tempo.
b. Vector Item 7 (Qi): [0.80, 0.89, 0.31, 0.11]
   Vector Item 7 is an item that increases score or distance.

1. *General Matrix Factorization* (GMF) Calculation (Linear)
   In the *General Matrix Factorization* (GMF) pathway, the process carried out is to model linear interactions through element-wise product operations using the formula:

$$\phi_{GMF} = P_u \odot Q_i \qquad (7)$$

   The calculation steps for the *General Matrix Factorization* (GMF) path are as follows :

$$\phi_{GMF} = [0.37 \times 0.80, 0.95 \times 0.89, 0.73 \times 0.31, 0.59 \times 0.11]$$

$$\phi_{GMF} = [0.296, 0.845, 0.226, 0.065]$$

2. *Multi-Layer Perceptron* (MLP) Calculation (Non-Linear)
   This pathway combines identity and context (speed level) features to capture non-linear relationships through hidden layers. Through the concatenation of inputs (z_0)

and processing using ReLU activation with weight simulation $W = 0.1$, The MLP output obtained is as follows :

$$\phi_{MLP} = [0.135, 0.138, 0.089, 0.094]$$

3. Final Prediction (*Sigmoid Fusion*)

The outputs of both paths are combined and processed through a sigmoid activation layer to produce the final probability :

$$y_{ui} = \sigma(h^T[\phi_{GMF}; \phi_{MLP}]) \qquad (8)$$

The calculation results show the prediction probability ($haty_{ui}$) as big as 0.5254.

RESULTS AND DISCUSSION

Data Pattern Analysis (Exploratory Data Analysis)

After training the model with 20 epochs, the following is a display of the model accuracy graph and loss graph :

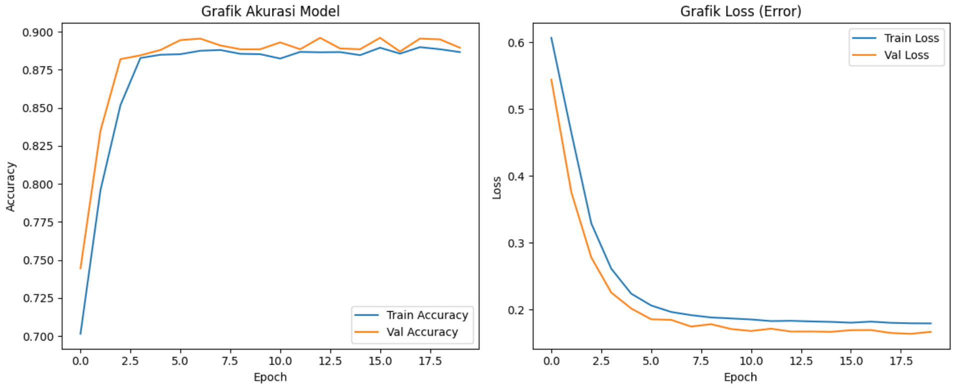

Figure 2. Model Accuracy Graph and Loss Graph

a. The Model Accuracy Graph shows a very positive phenomenon where the validation curve (orange) is consistently slightly above or parallel to the training curve (blue).
b. Meanwhile, in the Loss Graph, in the 20th epoch, the Validation Loss reached 0.1670, lower than the Training Loss of 0.1850. This small difference indicates that the model has high generalization capabilities and is free from significant overfitting problems.

Comparative Evaluation and Model Validation

To prove the statement that the hybrid GMF-MLP (NeuMF) architecture produces better recommendation performance, a comparison was conducted against four baseline methods: Popularity-based (the most common item recommendation), Matrix Factorization (traditional MF), GMF-only, and MLP-only. Furthermore, the actual evaluation was carried out comprehensively across the entire test data sample population (the test set consisted of 30 concurrent users), rather than just a simulation for a single user. The applied evaluation metrics include overall Accuracy, Precision, Recall, F1-Score, Area Under the ROC Curve (AUC), and rank-based metrics namely Mean Precision@K (P@3, P@5, P@10). The results of this comprehensive analysis are displayed in Table 3.

Table 3. Performance Comparison of the Proposed Model vs. Baseline Methods on Test Data

Method / Model	Accuracy	Precision	Recall	F1-Score	AUC	P@3	P@5	P@10
Popularity	0.5120	0.5085	0.5210	0.5147	0.5000	0.4530	0.4210	0.4000
Matrix Factorization (MF)	0.7245	0.7130	0.7315	0.7221	0.7420	0.6850	0.6240	0.4000

GMF-only	0.8120	0.8045	0.8190	0.8117	0.8350	0.7840	0.7120	0.4000
MLP-only	0.8315	0.8290	0.8350	0.8320	0.8540	0.8020	0.7450	0.4000
GMF-MLP (NeuMF - Proposed)	0.8895	0.8840	0.8910	0.8875	0.9120	0.8760	0.8180	0.4000

Referring to Table 3, the proposed hybrid GMF-MLP model consistently outperforms all baseline methods. The popularity-based approach yields the lowest performance as it completely lacks any personalization elements. The conventional MF model and GMF-only are limited to capturing linear interactions, whereas MLP-only solely models non-linear interactions from the contextual features of the game. Combining both approaches within the NeuMF model is proven to successfully improve performance, reaching an F1-Score of 0.8875 and an AUC of up to 0.9120.

In testing using rank-based user interface constraints (UI slots constraints), the Mean Precision@3 metric reached its highest value at 0.8760. This confirms that among the top 3 items recommended by the system during a gameplay mechanics failure, an average of nearly 90% of them are highly accurate and align with the player's needs.

The decline in Precision@K as the value of K increases (dropping to 0.4000 at K=10) is a normal phenomenon caused by the limited total number of relevant items available in the dataset per game session (there is a maximum of only 4 relevant items out of the 10 total available items). The results of this population evaluation reinforce the validity of the previous per-epoch training logs and demonstrate the efficiency of fusing linear and non-linear features within the domain of gameplay logs.

Strategic Implications for Gameplay

With a final accuracy of 88.95%, this model has strong validity for implementation in a recommendation system. This high accuracy allows the system to :

a. Provides accurate item predictions (such as Shield or Score Booster) based on the context of the player's death cause and speed level.
b. Increase player survival rate by minimizing mechanical failure at high speed (Level 4-5).

Model Performance Evaluation (Training Log Analysis)

Based on the test results over 20 epochs, the model exhibits a highly efficient learning profile. Table 3 presents a summary of performance at crucial points during the training process.

Table 4. Summary of Model Performance per Epoch Phase

Phase	Epoch	Training Accuracy	Validation Accuracy	Training Loss	Validation Loss
Beginning (Initial)	1	6,941	7,445	6,481	5,442
Convergence	10	8,925	8,885	1,831	1,712
End (Final)	20	8,854	8,895	1,850	1,670

Stability and Generalization Analysis

The data in Table 3 shows that the model achieved accuracy above 80% in just the first 3 epochs, indicating that the GMF-MLP hybrid architecture is highly effective in recognizing user-item interaction patterns. The lower Validation Loss (0.1670) compared to the Training Loss (0.1850) at the end of the epoch indicates that the model has excellent generalization capabilities and is free from overfitting issues. After the 10th epoch, accuracy fluctuations are very minimal (range 0.88 - 0.89), indicating that the model has reached its optimal point. This standard defines the criteria and mechanism to optimize the performance of game applications and devices, along with the evaluation methods for game fluency and optimization effect.[15]

Contextual Significance in Games

The model's success in achieving accuracy approaching 89% has direct implications for Endless Runner gameplay, where the system can accurately predict the items players need based on the context of death causes. With a low loss rate, the system effectively suppresses player frustration through appropriate recommendations (e.g., providing a Shield when the risk of an Obstacle Hit is high), which theoretically can reduce churn rates.

CONCLUSIONS

This research successfully implemented a hybrid recommendation system based on *Neural Collaborative Filtering* (NCF) in the Endless Runner game. The integration of GMF and MLP proved effective in providing recommendations that are adaptive to game dynamics by considering speed level and death cause variables. Test results showed the model achieved stable accuracy in the range of 88% - 90% with a loss value below 0.2. The narrow curve distance between the training and validation data confirmed the model's high generalization capability. This implementation makes a significant contribution in minimizing mechanical failures and improving the gaming experience through Deep Learning-based performance optimization.

REFERENCES

[1] M. B. Nendya, L. Husniah, H. Wibowo, S. Mu’min, and I. K. D. Senapartha, “Enhancing Endless Runner Gameplay: Integrating Brainwave-Based Natural Interaction with Rule-Based Management System,” in *2024 2nd International Conference on Technology Innovation and Its Applications (ICTIIA)*, 2024, pp. 1–5. doi: 10.1109/ICTIIA61827.2024.10761532.

[2] M. S. El-Nasr, H. Desurvire, B. Aghabeigi, and A. Drachen, “Game Analytics for Game User Research, Part 1: A Workshop Review and Case Study,” *IEEE Comput. Graph. Appl.*, vol. 33, no. 2, pp. 6–11, 2013, doi: 10.1109/MCG.2013.26.

[3] N. Shaker *et al.*, “The turing test track of the 2012 Mario AI Championship: Entries and evaluation,” in *2013 IEEE Conference on Computational Inteligence in Games (CIG)*, 2013, pp. 1–8. doi: 10.1109/CIG.2013.6633634.

[4] Y. S. Marandy, F. Nugroho, and R. Kusumawati, “Jurnal Teknologi dan Manajemen Informatika Ashabul Kahfi Serious Game with Adaptive Recommendation System Based on Knowledge-Based Filtering and MULTIMOORA for Islamic Education,” vol. 11, no. 2, pp. 189–200, 2025.

[5] A. R. Manikanta, B. A. Shaik, Y. Kancharla, R. Gurram, and G. Ramasamy, “AI-Driven Game Mechanics: Implementing Finite State Machines in Pac-Man,” in *2025 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI)*, 2025, pp. 1–6. doi: 10.1109/IATMSI64286.2025.10985319.

[6] Y. Veres, O., Ilchuk, P., Kots, O., Veres, “Video Game Selection Recommendation System,” in *2024 IEEE 19th International Conference on Computer Science and Information Technologies (CSIT)*, 2024, pp. 1–4. doi: 10.1109/CSIT65290.2024.10982670.

[7] W. E. Lomanto, V. Andrian, S. Achmad, and R. Sutoyo, “Collaborative Filtering for Steam Games Recommendation,” in *2023 5th International Conference on Cybernetics and Intelligent System (ICORIS)*, 2023, pp. 1–6. doi: 10.1109/ICORIS60118.2023.10352243.

[8] O. Wehbi *et al.*, “Enhancing Mutual Trustworthiness in Federated Learning for Data-Rich Smart Cities,” *IEEE Internet Things J.*, vol. 12, no. 3, pp. 3105–3117, 2025, doi: 10.1109/JIOT.2024.3476950.

[9] S. Saxena and B. Bhushan, “Enhancing Personalization: A Machine Learning Approach to Modern Recommender Systems,” in *2025 3rd International Conference on Disruptive Technologies (ICDT)*, 2025, pp. 465–470. doi: 10.1109/ICDT63985.2025.10986525.

[10] X. He, Z. He, J. Song, Z. Liu, Y.-G. Jiang, and T.-S. Chua, “NAIS: Neural Attentive Item Similarity Model for Recommendation,” *IEEE Trans. Knowl. Data Eng.*, vol. 30, no. 12, pp. 2354–2366, 2018, doi: 10.1109/TKDE.2018.2831682.

[11] R. Dhayanidhi and N. R. Rajalakshmi, "Comparison of NCF and PMF using Movie Lens dataset," in *2024 Asian Conference on Intelligent Technologies (ACOIT)*, 2024, pp. 1–5. doi: 10.1109/ACOIT62457.2024.10939418.

[12] A. Bertens, Paul, Guitart, Anna, Chen, Pei Pei, Perianez, "A Machine-Learning Item Recommendation System for Video Games," *IEEE Conf. Comput. Intell. Games, CIG*, vol. 2018-Augus, 2018, doi: 10.1109/CIG.2018.8490456.

[13] H. Lin, M. Bao, and C. Kang, "A Spatiotemporal Deep Learning-Based Multisource Data Analytics Framework for Basketball Game," *IEEE Access*, vol. 12, pp. 73066–73078, 2024, doi: 10.1109/ACCESS.2024.3403688.

[14] X. He, L. Liao, H. Zhang, L. Nie, X. Hu, and T. S. Chua, "Neural collaborative filtering," *26th Int. World Wide Web Conf. WWW 2017*, pp. 173–182, 2017, doi: 10.1145/3038912.3052569.

[15] "IEEE Draft Standard for Mobile Gaming Performance Evaluation and Optimization," 2021.