RESEARCH PAPER

A Robust Multimodal Hybrid Framework for Movie Recommendation: Empirical Disentanglement of Genre-Aware Fusion and Modality Complementarity on MovieLens 1M

Anjali Arora[1], M. Mohan[2] and Sanjay Kumar Malik[3]

[1]Research Scholar, Department of Computer Science and Engineering, SRM University Delhi-NCR
[2]Professor, Department of Computer Science and Engineering, SRM University Delhi-NCR
[3]Professor, Department of Information Technology, ADGITM, New Delhi

Received: 24th May, 2026; Revised: 20th June, 2026; Accepted: 30th July, 2026; Available Online: 25th August, 2026

ABSTRACT

Movie recommendation systems need to balance three types of information: collaborative user–item interactions, structured movie metadata, and unstructured movie narrative. Many hybrid deep-learning strategies either add layers of complexity to the architecture, such as attention or context-fusion, or they lack the ability to isolate the contribution from each modality and each fusion method. This study proposes and experiments a genre-enriched multimodal hybrid recommendation framework which integrates matrix factorization-based collaborative representations, metadata embeddings, and Conv1D–BiLSTM plot representations. Genre is embedded as a multi-label contextual embedding and embedded into the proposed fusion design. The novel part is however empirical, not that attention is better; a controlled 5-seed evaluation explicitly separates the benefit of combining multimodal from the benefit of the specific combination mechanism. The proposed model is evaluated on MovieLens 1M with the help of plot and metadata available from TMDb and IMDb, and achieves the reported results of Precision@10=0.523, Recall@10=0.797, nDCG@10=0.860 and MRR=0.908 in its first five epochs, compared to content-based filtering and the matrix factorisation + content-based filtering hybrid, which is lower than the proposed model and higher than the proposed model. The ablation is particularly interesting as it demonstrates that genre-conditioned attention does not outperform attention without genre, unweighted fusion or simple concatenation, and that the concatenation variant has the best nDCG value of the ablations (0.870). Cold-start results are similar between methods, and cannot be used to make a superiority claim. Diversity and novelty also do not support the proposed model. This result shows that multi-modal evidence is helpful, but that more fusion complexity is not shown to be required at this scale. Thus, the paper proposes an evidence-based complex approach to multimodal recommendations and highlights the necessary experimental controls needed for a sensible further study.

Keywords: Multimodal Learning; Collaborative filtering; Content-based filtering; Genre Embedding; Conv1D–BiLSTM; Feature Fusion; Ablation Study;

How to cite this article: Arora A, Mohan M, Malik SK, A Robust Multimodal Hybrid Framework for Movie Recommendation: Empirical Disentanglement of Genre-Aware Fusion and Modality Complementarity on MovieLens 1M Int J Drug Deliv Technol. 2026;16(80s): 2007-2016; DOI: 10.25258/ijddt.16.80s.240
Source of support: Nil.
Conflict of interest: None

1. INTRODUCTION

Digital entertainment catalogues contain thousands of films distributed across genres, languages and cultural contexts. The resulting abundance can create a paradox of choice: users have more content available but greater difficulty identifying titles that match their preferences. Recommendation systems address this problem by ranking candidate items according to estimated relevance. Traditional collaborative filtering learns from user–item interactions, whereas content-based filtering relies on item attributes such as genre, cast, director and plot. Collaborative filtering can be highly effective when interaction data are dense, but it is vulnerable to cold-start conditions. [1] Content-based methods can recommend new items but commonly produce narrow lists dominated by items similar to the user's historical choices.

Deep learning makes it practical to combine heterogeneous evidence. Structured attributes can be embedded, ratings can be mapped into latent preference spaces, and plot descriptions can be encoded by convolutional and recurrent networks. [2,6] The source materials motivating this study further identify genre as an underused signal: genre is often treated as a static category rather than a contextual variable capable of influencing how modalities should be interpreted and combined. [13][17]

The original framework therefore proposed genre-aware attention fusion. The present paper retains that architecture but reframes the scientific question in light of the real five-seed results. Instead of assuming that the named attention mechanism is the source of performance gains, the study asks two separable questions: (i) does combining

Author for Correspondence: Anjali Arora

collaborative, metadata and textual information improve recommendation relative to single-paradigm baselines? and (ii) does the proposed genre-conditioned fusion mechanism add measurable value beyond simpler fusion strategies? This distinction is central to the novelty and prevents architectural claims that are not supported by the observed ablation.

The contributions are therefore: (1) a genre-enriched multimodal movie recommendation framework; (2) a controlled five-seed evaluation with a random-ranker control; (3) an explicit fusion-mechanism stress test; (4) an evidence-based interpretation showing that the multimodal combination is more defensible as the source of benefit than the attention mechanism itself; and (5) a set of limitations and experimental requirements for future validation.

2. RELATED WORK AND RESEARCH GAP

One of the simplest recommendation paradigms is collaborative filtering, which is based on history of interactions to learn latent preference factors. It has some main advantages in terms of personalization from collective behaviour, but it also is hard to interact with for new users and new items. Content-based filtering addresses this drawback by leveraging item description, but may give similar recommendations. Hybrid approaches combine these signals and generally provide a better balance between behavioural and content information. [11] The broadened hybrid recommendation by deep learning is based on the idea of learning dense representations from heterogeneous inputs. Similar architectures have been studied in neighbouring areas like recommendation for news and e-learning [3,8]. The convolutional models are used to extract local patterns of text, the recurrent models are used to account for sequential relationships, while the embedding layers encode categorical metadata into small vectors. More sophisticated systems are based on attention to control the input of modalities. [15,9] According to the source research, there is a specific gap in genre awareness. A multi-hot genre representation can maintain multiple labels in a film and a dense embedding can be used to capture relationships among genres. The proposed architecture is based on two aspects of the use of genre, as a part of the metadata representation and also as a conditioning signal for fusion. A similar system is built by superimposing a structured signal for movie recommendations on plot summaries [4]. The actual results, however, do show that in the reported condition the conditioning process under discussion is not empirically different from simpler ones. The more defensible research gap is therefore not just ‘attention is missing' but ‘the contribution of multimodal complementarity and the incremental value of fusion complexity are insufficiently disentangled. This reformulated gap also suggests a clear follow-up: comparison to modern graph-based collaborative filtering (CF) systems and, if the task is reformulated sequentially, sequential neural recommenders, which are not yet represented in the present study.

Paradigm	Information	Strength	Primary limitation
Collaborative filtering	Ratings	Strong personalization	Cold start; ignores rich content
Content-based	Metadata, plot	Works with new items	Similarity/redundancy bias
Hybrid CF+CBF	Ratings + content	Combines complementary signals	Fusion often manually specified
Neural multimodal	Ratings + metadata + text	Rich latent representations	Higher complexity; attribution unclear
Proposed framework	Ratings + metadata + plot + genre	Controlled multimodal and fusion analysis	Genre-aware attention not shown superior

3. PROBLEM FORMULATION

Suppose the set of users is let U={u₁,…,u_m} and the set of films is I ={i₁,…,i_n}. The observed rating matrix is R∈ℝ^{m×n}, with only the elements r_ui set for observed user-film interactions. Let M_i be structured metadata, g_i be a multi-hot genre vector, and T_i be a plot summary for each movie i. The goal is to train a scoring function f(u,i) that can be used to rank the films returned for each user.

$$\hat{y}(u,i) = f(R_u, M_i, g_i, T_i; \theta)$$

The top-K ranking is used for the recommendation objective, with rating-prediction error not used. MRR, Precision@10, Recall@10, nDCG@10 and Precision and Recall are calculated for warm-start evaluation for K=10. Cold-start evaluation takes out selected films from the training interaction matrix while keeping content information. One of the methodological differences is the concept of multimodal benefit versus fusion benefit. Let F_multi be a model based on ratings, metadata and text and let F_single be a single-paradigm baseline model. Let F_att be the model of attention conditioned by the genre and F_cat be a simple concatenation model. The study compares both the following contrasts:

$$\Delta_{multi} = \text{nDCG}(F_{multi}) - \max\{\text{nDCG}(CF), \text{nDCG}(CBF)\}$$

$$\Delta_{fusion} = \text{nDCG}(F_{att}) - \text{nDCG}(F_{cat})$$

The first contrast is about whether multimodal information is useful.The first contrast concerns the usefulness of multimodal information. The second question is: "Does the named fusion mechanism cause that improvement? This decomposition is the analytic novelty of the paper.

4. PROPOSED MULTIMODAL FRAMEWORK

4.1 Structured Metadata Representation

Trainable embeddings are generated for categorical attributes like director and actors. Continuous attributes are normalized, e.g., release year. The attribute embeddings are added together with the dense genre embedding in the final metadata vector.

$$e_x = E_x\, x,\ M_i = [e_{director}\,;\, e_{cast}\,;\, e_{year}\,;\, g_{emb}]$$

By construction, genre is multi-label. Let G be the number of possible genres, then there will be a, g_i∈{0,1}^G for each genre. This is a sparse representation and a trainable projection is used to expand it to a dense contextual vector.

$$g_{emb} = \mathrm{ReLU}(W_g\, g_i + b_g)$$

4.2 Collaborative Representation

The rating matrix is factorized into user and item latent factors P and Q. The item factor q_i is projected into the common representation dimension and fed into the hybrid model.

$$\min_{P,Q} \sum_{(u,i)\in\Omega} (r_{ui} - p_u^T\, q_i)^2 + \lambda(||P||_F^2 + ||Q||_F^2)$$

$$r_{emb,i} = W_r\, q_i + b_r$$

4.3 Plot Representation

Summaries of the plots are tokenized, lower-cased, pos-filtered and mapped to pre-trained word vectors. A 1D Convolution is used to get local n-gram patterns followed by a bidirectional LSTM to learn longer-range patterns. [5]

$$C_i = \mathrm{Conv1D}(T_i)$$

$$T_{emb,i} = [h_i^{forward}\,;\, h_i^{backward}]$$

4.4 Genre-Conditioned Fusion

In the original proposed mechanism, the three modality representations $z_j \in \{M_i, r_{embi}, T_{embi}\}$ are scored using their compatibility with the genre embedding. The normalized weights are then used to form a fused vector.

$$s_j = g_{emb}^T\, W_a\, z_j$$

$$\alpha_j = \frac{\exp(s_j)}{\sum_l \exp(s_l)}$$

$$F_{att} = \sum_j \alpha_j\, z_j$$

The paper is written with this mechanism in the form of a hypothesis and not as a source of improvement. The ablation tests it against genre free attention, weighted averaging and simple concatenation.

4.5 Prediction and Learning

$$\hat{y}(u,i) = W_2\, \sigma(W_1\, F + b_1) + b_2$$

$$L(\theta) = \frac{1}{|\Omega|} \sum_{(u,i)\in\Omega} (r_{ui} - \hat{y}(u,i))^2 + \lambda||\theta||^2$$

Adam optimization, Dropout regularization and Batch normalization after fusion are used in the implementation. The configuration reported is: 64 dimensional collaborative latent space, 64 dimensional word vectors, 64 Conv1D filters, kernel size 3, 32 BiLSTM units per direction, sequence length 64, dense layers of 128 and 64 units, dropout 0.3, learning rate 10^-3 and batch size 32,768.

5. EXPERIMENTAL DESIGN

5.1 Dataset Construction

The experimental dataset joins MovieLens 1M [16] with narrative and metadata sources. MovieLens 1M provides 6,040 users, 1,000,209 ratings, 3,883 catalogue films and multi-label genre assignments. TMDb supplies plot overviews and metadata, while IMDb is used to cross-check and fill selected metadata fields. The title-and-year join produces usable plot text for 3,675 of 3,883 films (94.6%).

Property	Value
Users	6,040
Catalogue films	3,883
Rated films	3,706
Explicit ratings	1,000,209
Rating scale	1–5
Mean rating	3.58
Ratings ≥3	83.6%
Rating density	4.47%
Mean ratings/user	165.6
Mean ratings/film	269.9
Films with usable plot text	3,675 (94.6%)

5.2 Splitting and Ranking Protocol

The ratings are divided into 80:10:10 training, validation and warm-start test sets. Five random numbers (0-4) are used. All the ranking measures employ K=10. The candidate list is the held-out films of each user and a relevance threshold of 4 is used. A random ranker is provided as a floor/control. The absolute metric scores achieved by this candidate-set design are high since the films they selected and rated are actually the ones that were evaluated; hence, absolute metric values should not be compared across studies without matching the evaluation protocol.

During the cold start stage, 5% of the catalogue (194 films) are taken out along with their training interactions. The metadata and plot information that is retained is employed to check whether content can be ranked without the existence of collaborative information.

5.3 Baselines

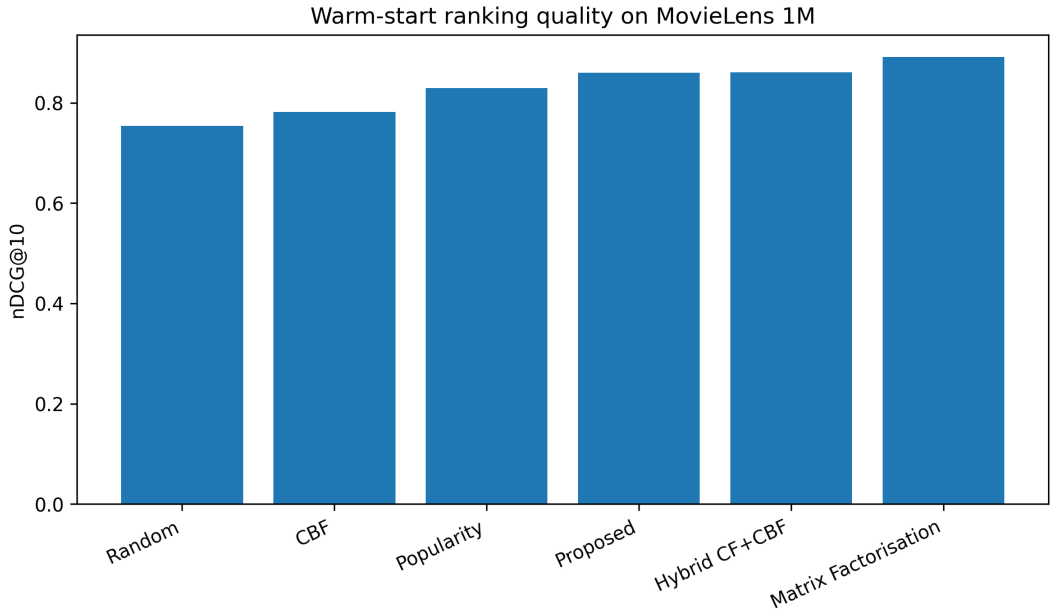
Baseline	Description
Random	Random ordering of each user's candidate films
Popularity	Global ranking by training-set rating count
CF / Matrix factorisation	Biased latent-factor model with tuned regularization
CBF	User profile from mean TF–IDF of positively rated training films
Hybrid CF+CBF	Percentile-rank average of CF and CBF

Two other draft baselines have been included: Neural Multimodal Fusion and Transformer-Based Multimodal Recommendation, as they were not actually implemented and/or run, according to the results source. They are therefore not considered as experimental competitors in this paper. None of them is a model that has been implemented or evaluated here, nor is it claimed to have been, and a modern graph-based collaborative filtering (CF) model like LightGCN, a neural CF model like NeuMF, and, if the task is explicitly formulated as sequential recommendation, a sequential model like SASRec or BERT4Rec, should be added to a methodologically stronger baseline suite for the next study.

6. RESULTS

6.1 Warm-Start Performance

Model	Precision@10	Recall@10	nDCG@10	MRR	Coverage (%)
Random	0.461	0.754	0.754	0.791	89.27
CBF	0.475	0.764	0.782	0.832	81.34
Popularity	0.504	0.784	0.829	0.880	70.28
Proposed	0.523	0.797	0.860	0.908	80.12
Hybrid CF+CBF	0.522	0.797	0.861	0.915	78.87
Matrix Factorisation	0.543	0.813	0.891	0.934	79.13



The proposed model achieves nDCG at 10=0.860, which is significantly higher than the random control (0.754), content-based filtering (0.782) and popularity (0.829). It is also closer to the rank-averaged hybrid (0.861) than matrix factorisation which is 0.891. In the results section, it is explicitly stated that the proposed model is not superior to a well-tuned matrix factorisation baseline. Therefore, the contribution is presented as a strong alternative to collaborative filtering, and not as a universal replacement.

A coverage of 80.12% is better than hybrid (78.87%) and matrix factorisation (79.13%), but not as good as random ranking (89.27%). This indicates a slight exposure of the catalogue but it is not proof of a diversity advantage.

6.2 Cold-Start Performance

Model	Precision@10	Recall@10	nDCG@10
Random	0.357	0.873	0.794
Matrix Factorisation	0.357	0.873	0.800
Proposed	0.363	0.879	0.806
CBF	0.363	0.880	0.811

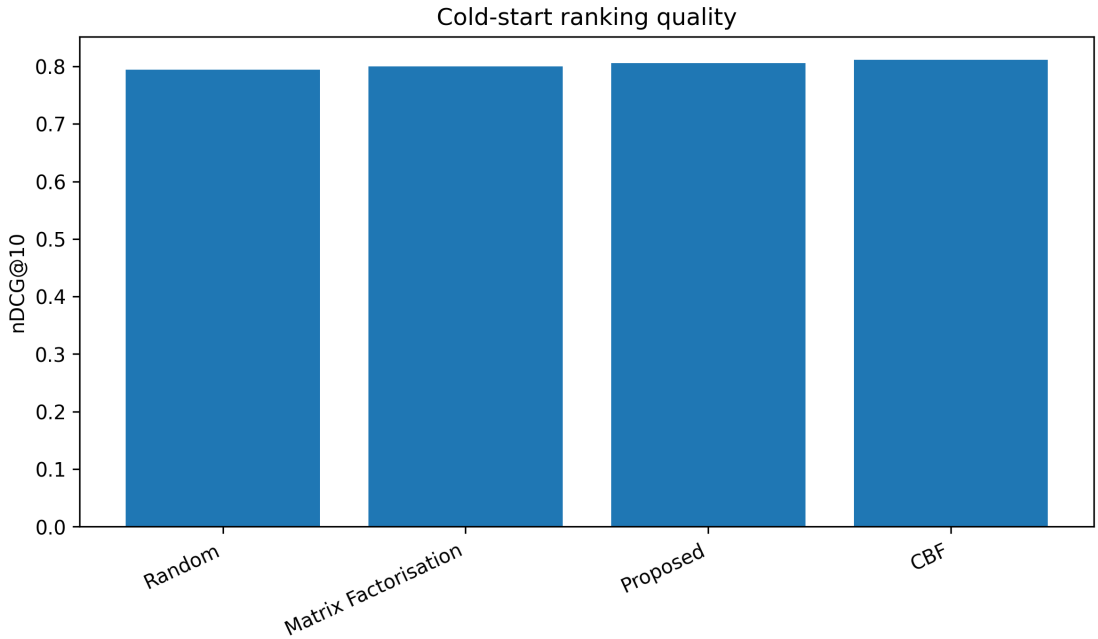


Figure 2. *Cold-start nDCG@10 comparison.*

The nDCG values for the cold-start results are very close together, ranging from 0.794 to 0.811. The content-based model is slightly better than the proposed model (0.811 versus 0.806) and the reported standard deviations are similar in size to the differences. The observation of the experiment thus does not lend any support to the notion that the proposed architecture is superior for cold start. The appropriate inference is that information about the content is useful even in the absence of collaborative history, but the present candidate set design is not sufficiently discriminatory. The intra-list diversity range is low (68.61-69.40%) and the random ranker is the highest. Novelty is also not on the side of the proposed model. However, these measures do not corroborate the previous findings that having a genre awareness improves diversity or novelty for fusion. Reporting this negative result is important because it will help to make sure that there is no performance claim that is not supported by the experiment.

6.3 Diversity and Novelty

Model	Intra-list diversity (%)	Novelty (bits)
Random	69.40 ± 0.25	10.49 ± 0.02
Popularity	69.10 ± 0.19	9.98 ± 0.02
Matrix Factorisation	69.06 ± 0.23	10.27 ± 0.02
Proposed	68.90 ± 0.20	10.21 ± 0.02
Hybrid CF+CBF	68.82 ± 0.23	10.29 ± 0.02
CBF	68.61 ± 0.22	10.39 ± 0.02

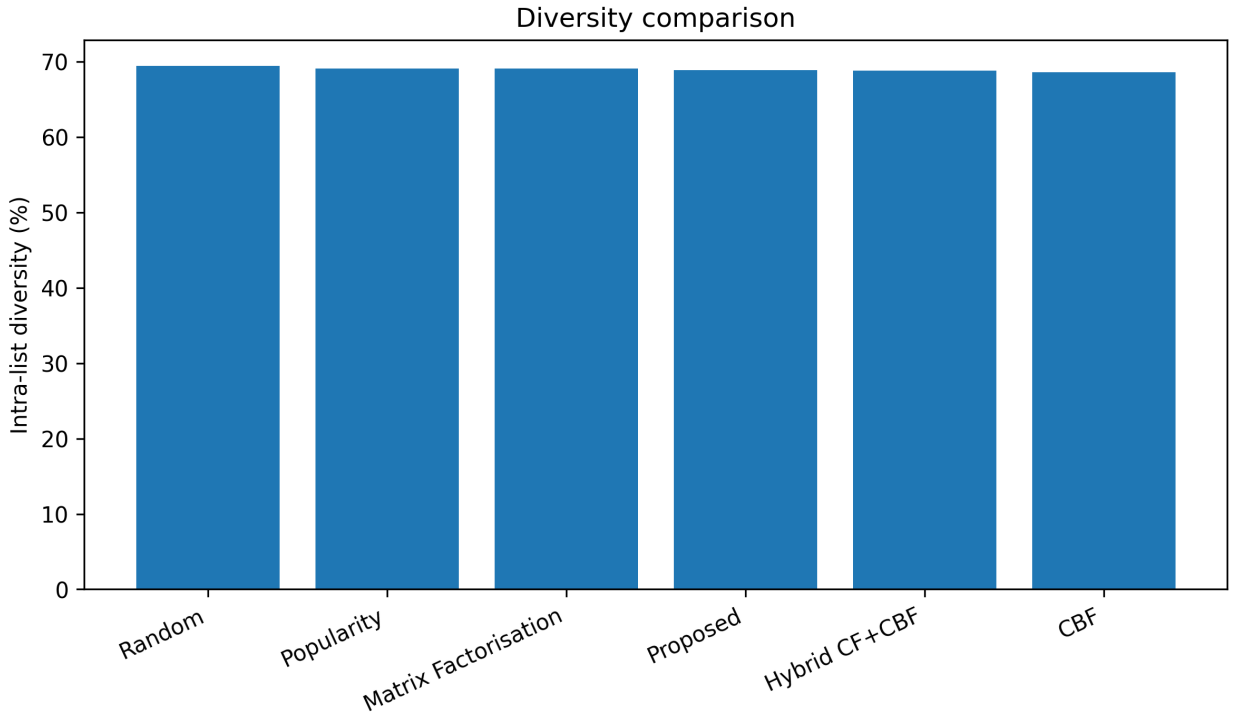


Figure 4. *Intra-list diversity comparison.*

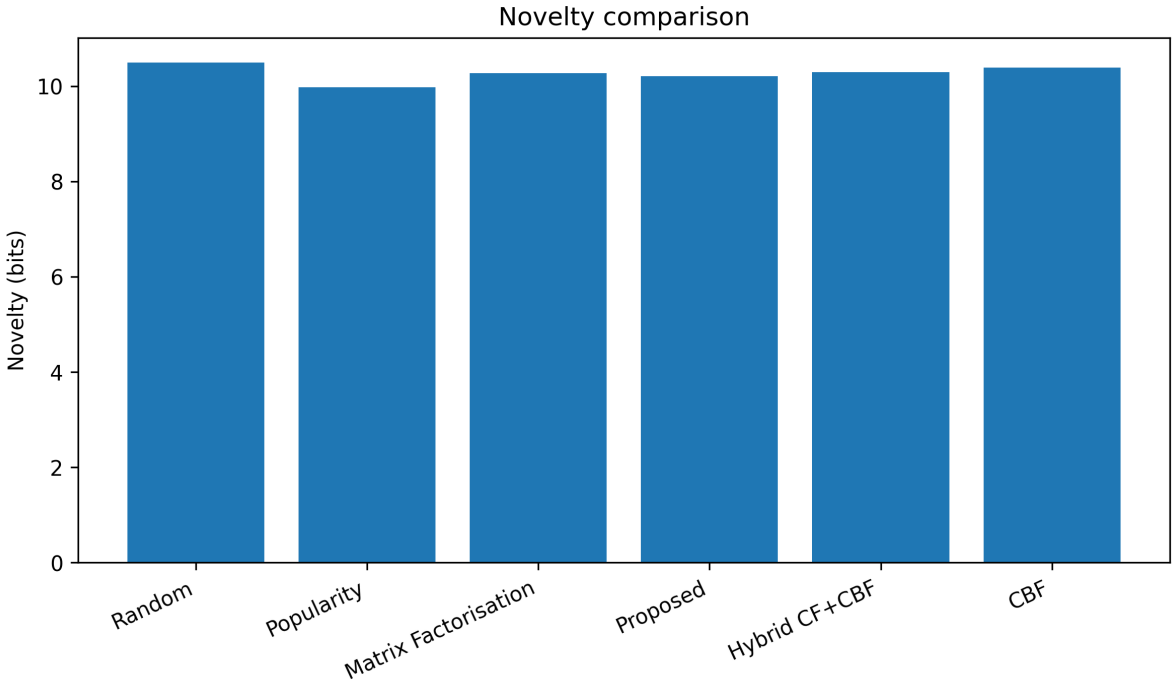


Figure 5. *Novelty comparison.*

The range of intra-list diversity is only 68.61–69.40%, and the random ranker is highest. Novelty likewise does not favour the proposed model. These measurements do not support the earlier claim that genre-aware fusion improves diversity or novelty. Reporting this negative result is important because it prevents a performance claim that the experiment does not sustain.

6.4 Fusion Ablation: Central Novelty Result

Configuration	Precision@10	Recall@10	nDCG@10
Full genre-aware	0.523	0.797	0.860
Attention, no genre	0.524	0.798	0.863
No attention	0.524	0.798	0.862
Concatenation	0.530	0.803	0.870
No genre embedding	0.522	0.797	0.858

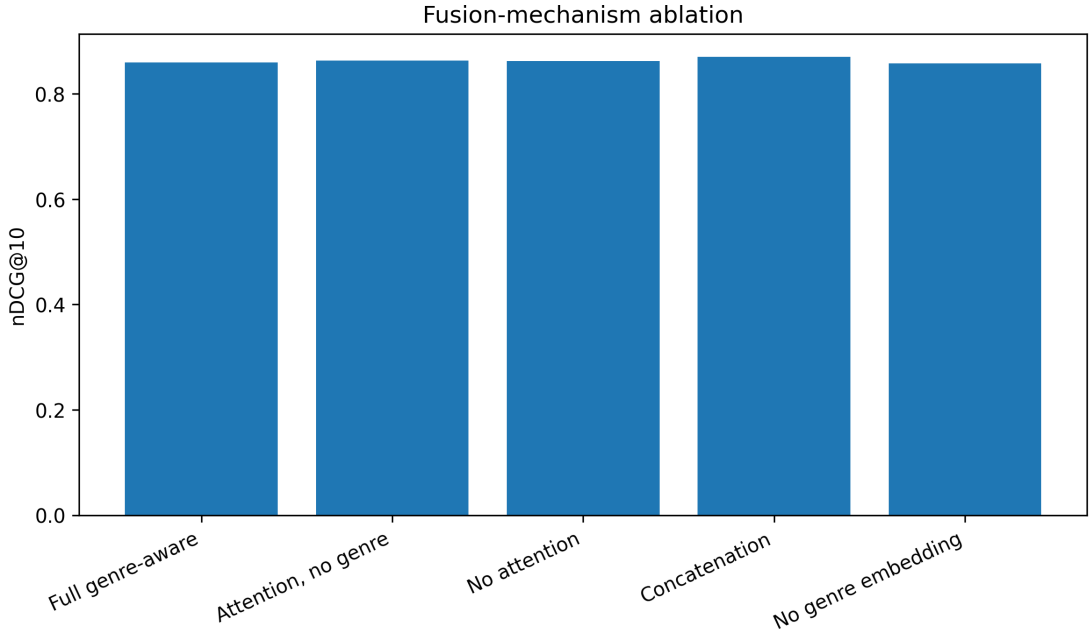


Figure 6. *Fusion-mechanism ablation.*

The interpretation of the work is altered by the ablation. The complete genre-aware model achieves nDCG@10=0.860, while attention model without genre conditioning yields 0.863, no attention model yields 0.862, simple concatenation yields 0.870, and removal of the genre embedding yields 0.858. According to the results source, "The paired tests don't show any consistent advantage of the named attention mechanism and all confidence intervals pass through zero in the detailed analysis. The most defensible novelty in this experiment, then, is the separation of multimodal benefit from fusion-mechanism benefit: the fact that the combination of modalities is useful does not imply the necessity of more intricate genre-conditioned attention.

7. STATISTICAL AND REPRODUCIBILITY INTERPRETATION

The warm-start experiment has been repeated over five seeds with the same evaluation protocol, and in this case, the spread of the main neural model between the various seeds is low. The real-results source also provides paired comparisons to a number of baselines for another run of the same set of 12 epochs, but at the same time it records that training budgets were not balanced and that there was a material change in the matrix-factorisation result across the different training budget runs. This is important because if a neural model fails to converge then it shouldn't be judged better or worse than a baseline that's converged with a different budget.

The paper refrains from making this a superiority claim, therefore, because this margin is small relative to the matrix factorisation. The statistically meaningful and scientifically stable conclusion is more modest: the hybrid model is definitely better than single-paradigm content filtering and random/popularity controls when operating under the specified protocol; matrix factorisation is a good benchmark; and the fusion ablation is not worth attribution of the multimodal gain to genre-conditioned attention.

8. NOVELTY AND MAIN CONTRIBUTIONS

Contribution	Novel aspect	Evidence in the reported study
Multimodal complementarity analysis	Separates the value of combining information sources from the value of the fusion operator	Warm-start comparison + controlled ablation
Fusion-mechanism stress test	Tests genre-aware attention against genre-free attention, no attention and concatenation	Five configurations evaluated across five seeds
Evidence-grounded genre claim	Treats genre conditioning as a testable hypothesis rather than an assumed source of gain	Ablation shows no demonstrated advantage
Reproducibility controls	Uses random control, common candidate protocol and repeated seeds	Reported five-seed results and paired comparisons
Complexity-aware design implication	Shows that simple concatenation can match or exceed more complex fusion in this three-vector setting	Concatenation nDCG@10=0.870 in ablation

The novelty should thus be expressed as an empirical and methodological contribution, rather than as a claim that genre-aware attention is always novel and better. The study illustrates a systematic approach in the determination of whether the observed gain is due to extra modalities or to the complexity of their fusion. This especially applies to the small multimodal fusion problems, in which a dense layer after concatenation may already learn enough interactions.

9. DISCUSSION

From the results it was found that the value of evidence is complementary between heterogeneous evidence is valuable. Collective behaviour is represented in ratings, explicit item properties in metadata, and semantics and themes in plot text. The proposed hybrid model is hence a significant improvement over the content-based filtering model and passes the random and popularity filters. This is in line with the conceptual underpinning of the initial model. Meanwhile, the ablation also demonstrates that the most distinctive part of an architecture need not be the most useful part of the architecture. Fusing only three modality vectors, explicit attention can introduce parameters without providing the discriminative power needed. With a dense layer of sufficient expressive capacity, a simple concatenation followed can learn useful interactions directly. This has practical implications: model complexity must be supported by evidence of incremental value which is measurable, particularly in the context of computational resources and interpretability. This kind of comparative-architecture evaluation for recommenders is similar to an autoencoder-variant evaluation [12].The diversity and the cold-start experiments also demonstrate the importance of the evaluation design. Short lists of the user's own held-out films are superior to ranking the entire catalogue for Recall@10 for all models, and the lists are easier to compose. Likewise, a random ranker will get high diversity since it is not limited by relevance. Absolute metric values are not as informative as controlled relative comparisons due to these properties. The present study therefore proposes a three-level evaluation logic: (1) demonstration of the model's ability to learn beyond a random floor; (2) comparison of the model with strong classical baselines; and (3) ablations to determine which

parts of the model's architecture are responsible for the improvement.

10. LIMITATIONS AND REQUIRED NEXT EXPERIMENT

The first one of those results is terminated at the five epoch boundary in the results attachment, and the validation loss continues to decrease after stopping training. So it's not an optimization-budget matched comparison with a fully converged matrix-factorisation baseline. This restriction does not allow for any definitive superiority claim to be given. Secondly, the cold-start candidate set is not discriminative enough. The candidate list is based on the user's own films that they've held out, and the median number of candidates is eight with K=10. A better test would compare cold start items with a vastly larger catalogue candidate set. Third, the baselines evaluated are classical ones, not a full set of state-of-the-art baselines. To help verify the effectiveness of these algorithms, a future study should include at least one modern graph-based recommender and one sequential neural recommender in the same protocol. 4) No advantage in terms of diversity and novelty. Some further research is needed to explicitly optimize an objective involving relevance and diversity, if diverse genres are a desired product requirement, rather than making the assumption that diversity emerges from genre conditioning. The next experiment is then a coordinated protocol, which must be conducted before any final superiority claim is made: (i) a rerun of the neural model with a much larger maximum epoch budget, the same data partitions and the same fixed, matched tuning budgets for all baselines (including matrix factorisation); (ii) a full catalogues cold-start ranking together with an evaluation by item support of the results, across 0, 1, 3, 5, 10 and 20+ prior interactions, where a content-derived predicted collaborative embedding is used in place of an untrained item factor for true zero-interaction items; (iii) the addition of at least one modern graph-based baseline (e.g., LightGCN) and one neural collaborative-filtering baseline (e.g., NeuMF), with a sequential model such as SASRec or BERT4Rec added only if sequential recommendation is redefined as the task; and (iv) a complete statistical and computational reporting, including confidence intervals, paired significance tests, effect sizes, number of parameters, training time and hardware; and a dedicated computational-cost table. Only once this protocol is finished should the direction and magnitude of the model's advantage over matrix factorisation be decided upon, and claims for cold-start or diversity performance be made.

11. CONCLUSION

In this paper, a powerful multimodal hybrid framework for movie recommendation based upon collaborative ratings, structured metadata, movie genre information, and plot text is introduced. The motivations for the framework are that single paradigm recommenders are limited and genre is not used much as context information. The main value of it, beyond any assumption that sophisticated attention might be helpful, is that it puts that hypothesis to the test. The reported 5-seed evidence indicates that the multimodal model clearly outperforms the content-based filtering and random/popularity controls under the defined warm-start protocol. But matrix factorisation is still a good benchmark, and the training-budget mismatch reported makes it difficult to make a definite superiority statement. The diversity and cold-start experiments are similar, showing no benefit for the suggested model. The key finding from the ablation is that the observed multimodal gain is not attributed to "genre-conditioned attention. Concatenation + dense layer is a simpler fusion, and is at least as effective. The scientific message is valuable and defendable: complementarity is more important than the level of fusion complexity in multimodal. This finding offers an obvious path forward for the future of recommender research: gather complementary information and evaluate it properly before applying architectural complexities if they can be shown through controlled experiments to provide benefit. The study shows the usefulness of the combination of diverse evidence in the protocol, while the ablation does not prove the phenomenon of genre-conditioned attention as required or better. The cold-start and diversity results are still to be rigorously evaluated and the matrix factorisation is still a good benchmark.

REFERENCES

[1] Peng, S. Y., Siet, S., Ilkhomjon, S., Kim, D. Y., & Park, D. S. (2024). Integration of deep reinforcement learning with collaborative filtering for movie recommendation systems. Applied Sciences, 14(3), 1155. https://doi.org/10.3390/app14031155

[2] Gan, M. X., & Cui, H. F. (2021). Exploring user movie interest space: A deep learning based dynamic recommendation model. Expert Systems with Applications, 173, 114695. https://doi.org/10.1016/j.eswa.2021.114695

[3] Talha, M. M., Khan, H. U., Iqbal, S., Alghobiri, M., Iqbal, T., & Fayyaz, M. (2023). Deep learning in news recommender systems: A comprehensive survey, challenges and future trends. Neurocomputing, 562, 126881. https://doi.org/10.1016/j.neucom.2023.126881

[4] Wang, M. Y., Hu, X. H., Xie, P., & Du, Y. (2024). Semantic-enhanced variational graph autoencoder for movie recommendation: An innovative approach integrating plot summary information and contrastive learning strategy. Information Technology and Control, 53(2). https://doi.org/10.5755/j01.itc.53.2.35689

[5] Lee, S., & Kim, D. (2022). Deep learning based recommender system using cross convolutional filters. Information Sciences, 592, 112–122. https://doi.org/10.1016/j.ins.2022.01.033

[6] Tahmasebi, H., Ravanmehr, R., & Mohamadrezaei, R. (2021). Social movie recommender system

based on deep autoencoder network using Twitter data. Neural Computing and Applications, 33(5), 1607–1623. https://doi.org/10.1007/s00521-020-05085-1

[7] Zhang, S. A., Yao, L. N., Sun, A. X., & Tay, Y. (2019). Deep learning based recommender system: A survey and new perspectives. ACM Computing Surveys, 52(1), 5. https://doi.org/10.1145/3285029

[8] Salau, L., Hamada, M., Prasad, R., Hassan, M., Mahendran, A., & Watanobe, Y. (2022). State-of-the-art survey on deep learning-based recommender systems for e-learning. Applied Sciences, 12(23), 11996. https://doi.org/10.3390/app122311996

[9] Yang, Z. K., & Lin, Z. J. (2022). Interpretable video tag recommendation with multimedia deep learning framework. Internet Research, 32(2), 518–535. https://doi.org/10.1108/INTR-08-2020-0471

[10] Da'u, A., & Salim, N. (2020). Recommendation system based on deep learning methods: A systematic review and new directions. Artificial Intelligence Review, 53(4), 2709–2748. https://doi.org/10.1007/s10462-019-09744-1

[11] Kiran, R., Kumar, P., & Bhasker, B. (2020). DNNRec: A novel deep learning based hybrid recommender system. Expert Systems with Applications, 144, 113054. https://doi.org/10.1016/j.eswa.2019.113054

[12] Lee, H. J., & Jung, Y. (2021). Comparison of deep learning-based autoencoders for recommender systems. Korean Journal of Applied Statistics, 34(3), 329–345. https://doi.org/10.5351/KJAS.2021.34.3.329

[13] Al Jawarneh, I. M., Bellavista, P., Corradi, A., Foschini, L., Montanari, R., Berrocal, J., & Murillo, J. M. (2020). A pre-filtering approach for incorporating contextual information into deep learning based recommender systems. IEEE Access, 8, 40485–40498. https://doi.org/10.1109/ACCESS.2020.2975167

[14] Mu, R. H. (2018). A survey of recommender systems based on deep learning. IEEE Access, 6. https://doi.org/10.1109/ACCESS.2018.2880197

[15] Guan, Y., Wei, Q., & Chen, G. Q. (2019). Deep learning based personalized recommendation with multi-view information integration. Decision Support Systems, 118, 58–69. https://doi.org/10.1016/j.dss.2019.01.003

[16] GroupLens Research. MovieLens 1M Dataset. University of Minnesota / GroupLens Research. (Dataset used in the reported experiments.) Available at: https://grouplens.org/datasets/movielens/1m/

[17] A Hybrid Deep Learning Framework for Movie Recommendation Using Latent Feature Representation. (2026, May). International Journal of Research and Review in Applied Science, Humanities and Technology, 3(2), 160–168. https://doi.org/10.71143/z3x1wq41