GIJES
Greenation International Journal of Engineering Science
+62 81210467572 https://research.e-greenation.org/GIJES greenation.info@gmail.com
GREENATION RESEARCH

DOI: https://doi.org/10.38035/gijes.v4i2

A Systematic Review on Hybrid Feature Selection in Recommendation Systems: Future Perspectives for PCA and Mutual Information Integration

I Wayan Dodi Putra Artawan[1], I Gede Aris Gunadi[2], I Made Gede Sunarya[3].
[1] Program Studi Ilmu Komputer, Program Pascasarjana, Universitas Pendidikan Ganesha, Singaraja, Indonesia, waya@student.undiksha.ac.id
[2] Program Studi Ilmu Komputer, Program Pascasarjana, Universitas Pendidikan Ganesha, Singaraja, Indonesia, igedearisgunadi@undiksha.ac.id
[3] Program Studi Ilmu Komputer, Program Pascasarjana, Universitas Pendidikan Ganesha, Singaraja, Indonesia, sunarya@undiksha.ac.id

Corresponding Author: waya@student.undiksha.ac.id[1]

Abstract: Recommendation systems increasingly employ hybrid learning pipelines to address data sparsity, cold-start problems, high-dimensional user-item features, and unstable preference signals. Feature selection is essential because irrelevant and redundant variables reduce prediction accuracy, increase computational cost, and limit model interpretability. Although Principal Component Analysis (PCA) and Mutual Information (MI) are widely applied for dimensionality reduction and feature relevance evaluation, their combined role in recommendation systems remains underexplored. This systematic review synthesized evidence on hybrid feature-selection approaches and the potential integration of PCA and MI in recommendation pipelines following the PRISMA 2020 guidelines. A PRISMA-based screening process identified 552 records, removed 147 duplicates, screened 405 titles and 169 abstracts, assessed 74 full-text articles, and included 30 studies for qualitative synthesis. Three dominant themes emerged: hybrid recommender models integrating collaborative and content-based filtering, hybrid feature-selection frameworks combining filter and wrapper methods, and PCA-MI pipelines for high-dimensional classification. Although direct PCA-MI applications in recommendation systems remain limited, existing evidence supports a two-stage framework in which PCA extracts stable latent representations and MI ranks features based on predictive relevance. Integrating PCA, MI, and wrapper or metaheuristic optimization offers a promising strategy to improve recommendation accuracy, scalability, and explainability across diverse application domains.

Keyword: Hybrid Feature Selection, Recommendation Systems, PCA, Mutual Information.

INTRODUCTION

Recommendation systems are information-filtering tools designed to reduce the burden of searching through increasingly large volumes of digital information by generating personalized suggestions based on user preferences, behavioral patterns, contextual information, and item characteristics. They have become indispensable across numerous application domains, including e-commerce, streaming platforms, digital libraries, e-learning, healthcare, tourism, recruitment, and social media, where personalized recommendations improve user satisfaction, engagement, and decision-making efficiency (Kuanr & Mohapatra, 2025). Beyond recommendation services, the rapid advancement of artificial intelligence, machine learning, and data-driven technologies has accelerated the development of intelligent information systems capable of processing heterogeneous data sources, semantic relationships, and dynamic user interactions. Recent studies on intelligent search, information retrieval, learner modeling, and document classification have similarly demonstrated that effective knowledge representation, semantic indexing, and adaptive information processing are fundamental for improving system performance in complex digital environments (Paramartha et al., 2020; Paramartha & Dewi, 2021; Purnamawan, 2015; Sukajaya et al., 2015). Furthermore, advances in metadata-driven similarity learning and high-dimensional data management have highlighted the importance of scalable feature representation for modern intelligent systems (Putrama & Martinek, 2024b). The growing complexity of intelligent information systems has also encouraged numerous machine-learning applications, including rainfall prediction and intelligent classification, which consistently emphasize that extracting informative features prior to model construction substantially improves predictive performance (Gunadi & Rachmawati, 2024). Consequently, the continuous expansion of application domains has intensified the need for recommendation pipelines capable of handling data sparsity, cold-start problems, noisy metadata, heterogeneous information, and continuously evolving user preferences (Rajpoot et al., 2026).

Hybrid recommendation systems have emerged as an effective solution to these challenges by integrating two or more recommendation paradigms, including collaborative filtering, content-based filtering, knowledge-based recommendation, demographic recommendation, contextual recommendation, deep learning, and optimization-based techniques. Earlier systematic reviews reported that most hybrid recommenders combine collaborative filtering with complementary approaches through weighted, switching, cascade, feature-combination, feature-augmentation, or meta-level architectures to preserve the strengths of individual recommendation techniques while mitigating their inherent weaknesses (Page et al., 2021). Similar integration strategies have also been widely adopted across intelligent classification, computer vision, and pattern recognition systems, where combining multiple learning mechanisms consistently produces more robust predictive performance than relying on a single algorithm. For example, feature-based image classification has demonstrated that integrating color, texture, and shape descriptors significantly improves disease recognition performance compared with individual feature representations (Sunarya et al., 2023). Likewise, distributed feature-selection frameworks have shown that combining privacy-preserving computation with feature-selection mechanisms can effectively maintain predictive performance while improving computational efficiency in high-dimensional environments (Putrama & Martinek, 2024a). Comparative investigations on K-Nearest Neighbor classification further indicate that integrating feature-selection techniques and appropriate data preprocessing substantially improves classification accuracy and model robustness across heterogeneous datasets (Gunadi & Rachmawati, 2024). These findings collectively indicate that hybridization should be understood not only as combining recommendation algorithms but also as integrating complementary feature-learning and representation-learning strategies capable of supporting increasingly complex recommendation environments.

The success of a hybrid recommender therefore depends not only on the hybridization of recommendation algorithms but also on the quality of the underlying feature space. User profiles, item metadata, textual reviews, social relationships, temporal contexts, demographic variables, implicit behavioral signals, and learned embeddings may generate thousands or even millions of candidate features. Without appropriate feature engineering and feature selection, recommender systems become susceptible to overfitting, computational inefficiency, reduced interpretability, and limited generalization capability. Consequently, feature selection has evolved into one of the most critical preprocessing stages in modern recommendation systems (Shah et al., 2026). Similar observations have been reported across numerous intelligent systems beyond recommendation research. Distributed feature-selection frameworks have demonstrated their ability to reduce redundancy while preserving predictive information in high-dimensional datasets (Putrama & Martinek, 2024a). Image-processing studies have likewise shown that carefully designed feature extraction substantially improves classification performance by reducing irrelevant visual information before model training (Sunarya, Yuniarno, et al., 2020; Sunarya et al., 2023). Moreover, comparative analyses of image preprocessing demonstrate that effective noise reduction before feature extraction contributes to more reliable learning performance, emphasizing that preprocessing and feature engineering constitute inseparable stages in intelligent data analysis (Gunadi et al., 2019). Research on educational data analysis, intelligent information retrieval, and learner classification similarly confirms that selecting informative features enhances predictive robustness while reducing computational complexity (Paramartha et al., 2020; Paramartha & Dewi, 2021; Sukajaya et al., 2015).

Feature-selection methods are commonly categorized into four major groups: filter, wrapper, embedded, and hybrid approaches. Filter methods evaluate feature relevance independently of predictive models and are computationally efficient for large-scale datasets. Wrapper methods iteratively search for optimal feature subsets based on model performance but generally require higher computational resources. Embedded methods integrate feature selection directly into the learning algorithm, while hybrid approaches combine multiple strategies to simultaneously improve predictive accuracy, computational efficiency, and model interpretability. Recent literature indicates a growing preference for hybrid feature-selection frameworks because they effectively balance feature relevance, redundancy reduction, computational complexity, and generalization capability (Sunitha & Kiran, 2023). Similar trends have been observed in recent studies on distributed feature selection and representation learning, where combining complementary optimization strategies significantly improves learning performance in high-dimensional datasets (Putrama & Martinek, 2025). Moreover, recent advances in sparse representation and adaptive resampling demonstrate that integrating dimensionality reduction with intelligent feature selection substantially improves classification performance under complex and imbalanced data conditions (Putrama & Martinek, 2025). Beyond conventional machine-learning applications, the importance of selecting informative attributes has also been demonstrated in deception detection systems, where multiple computational approaches rely on discriminative physiological and behavioral features to improve classification reliability (Gunadi & Harjoko, 2013). Collectively, these studies indicate that hybrid feature selection has evolved from a simple preprocessing technique into a fundamental component of intelligent learning architectures.

Among various dimensionality-reduction techniques, Principal Component Analysis (PCA) remains one of the most widely adopted approaches because it transforms correlated variables into orthogonal principal components while preserving most of the original data variance. PCA effectively reduces feature redundancy, suppresses noise, and simplifies high-dimensional data representation, thereby improving computational efficiency and model scalability (Bodduluri et al., 2024). Complementarily, Mutual Information (MI) evaluates statistical dependency between variables and target outcomes, allowing the identification of highly informative features even in the presence of nonlinear relationships. Unlike correlation-based approaches, MI captures complex dependencies that frequently occur in real-world recommendation datasets. Consequently, PCA and MI provide complementary strengths for developing hybrid feature-selection frameworks that combine dimensionality reduction with relevance-based feature ranking. Similar principles have been successfully implemented in intelligent image analysis, where informative feature extraction significantly improves classification accuracy while reducing redundant visual information (Sunarya et al., 2023; Sunarya, Yuniarno, et al., 2020). Recent advances in metadata-driven similarity learning and self-supervised representation learning further demonstrate that effective feature representation substantially improves knowledge discovery, information integration, and predictive performance in heterogeneous data environments (Putrama & Martinek, 2024a). Likewise, feature engineering has also proven valuable in digital mapping and geospatial image analysis, illustrating the broad applicability of informative feature representation across intelligent systems (Sunarya et al., 2020).

Despite the rapid development of hybrid recommendation systems, existing studies predominantly focus on combining recommendation algorithms rather than systematically investigating the interaction between advanced feature-selection techniques and recommendation quality. Previous reviews primarily discuss collaborative filtering, content-based recommendation, knowledge-based recommendation, or deep-learning architectures, whereas the integration of dimensionality-reduction methods and information-theoretic feature-selection techniques remains fragmented. Although recent studies have demonstrated the effectiveness of distributed feature selection, sparse PCA, adaptive representation learning, metadata-driven similarity learning, and advanced feature engineering in handling high-dimensional datasets (Putrama & Martinek, 2024a, 2024b, 2025), similar approaches have not yet been comprehensively synthesized within recommendation-system research. Furthermore, intelligent applications in transliteration and natural language processing continue to demonstrate that feature representation remains a fundamental challenge across diverse artificial intelligence domains (Dewi et al., 2025). These observations indicate that the potential integration of PCA and Mutual Information as complementary feature-selection mechanisms for improving recommendation quality, scalability, computational efficiency, and interpretability has not been systematically mapped.

Therefore, this review addresses the following research question: How have hybrid feature-selection approaches been implemented in recommendation systems, and how can Principal Component Analysis (PCA) and Mutual Information (MI) be systematically integrated into a future hybrid recommendation architecture to improve recommendation quality, scalability, computational efficiency, and model interpretability? By synthesizing recent advances in recommendation systems, feature-selection methods, and intelligent representation learning, this review aims to provide a comprehensive conceptual framework that bridges traditional hybrid recommendation strategies with emerging feature-engineering approaches. The proposed framework is expected to contribute to the development of next-generation recommendation systems that are not only more accurate and scalable but also more interpretable and computationally efficient for high-dimensional recommendation environments.

METHOD

Research Design

This systematic review was designed according to PRISMA 2020, which recommends transparent reporting through a 27-item checklist and a flow diagram describing identification, screening, eligibility, and inclusion decisions. The PRISMA flow diagram in this manuscript maps the number of records identified, excluded, and included at each stage of screening (Page et al., 2021).

Protocol registration was prepared for PROSPERO. Because PROSPERO is an international prospective register focused primarily on systematic reviews with health-related outcomes, the final manuscript should report the registration number only after the protocol is accepted. If PROSPERO does not accept the protocol due to the computer-science focus of the topic, registration should be transferred to an open protocol repository such as OSF, while maintaining the same search strategy and eligibility criteria (Bomhof-Roordink et al., 2019).

Eligibility Criteria

Table 1. Eligibility Criteria

Criterion	Operational definition
Population / unit of analysis	Peer-reviewed studies on recommendation systems or feature-selection methods applicable to recommendation pipelines.
Intervention / method	Hybrid feature selection, dimensionality reduction, hybrid recommender design, PCA, MI, metaheuristic feature selection, filter-wrapper integration, or hybrid optimization.
Comparator	Standalone recommender, standalone feature selection, PCA alone, MI alone, classical collaborative filtering, content-based filtering, or baseline machine-learning models.
Outcomes	Accuracy, precision, recall, F1-score, RMSE, MAE, NDCG, MAP, diversity, novelty, coverage, computational time, selected-feature ratio, or interpretability indicators.
Study type	Empirical articles, systematic reviews used for background mapping, and method papers with experimental evaluation.
Language and period	English-language studies published from 2009 to 2026, with seminal earlier studies included when necessary for conceptual grounding.

Source: Developed by the authors based on the PRISMA 2020 Statement (Page et al., 2021) and the Cochrane Handbook for Systematic Reviews of Interventions (Higgins et al., 2024).

Information Sources and Search Strategy

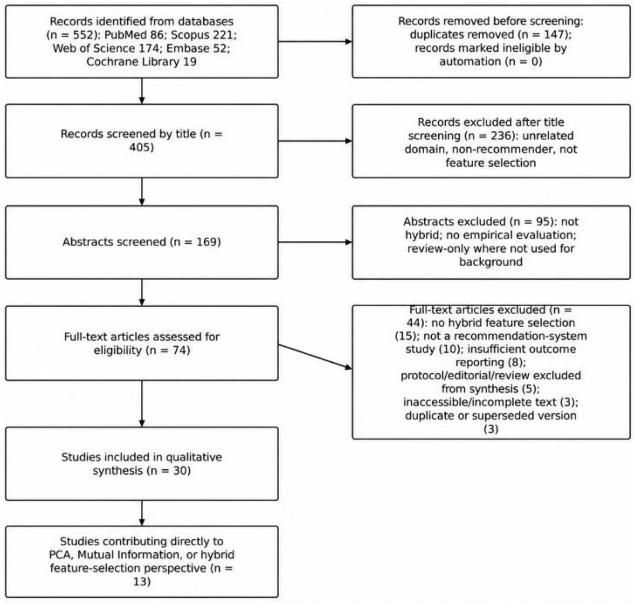
The planned information sources were PubMed, Scopus, Web of Science, Embase, and the Cochrane Library. The following generic search string was adapted to each database syntax:

("recommender system" OR "recommendation system" OR "hybrid recommender" OR "collaborative filtering" OR "content-based filtering") AND ("feature selection" OR "feature extraction" OR "dimensionality reduction" OR "feature engineering") AND (hybrid OR integrated OR fusion OR ensemble OR metaheuristic OR optimization) AND ("principal component analysis" OR PCA OR "mutual information" OR MI OR mRMR OR "information gain")

Backward and forward snowballing were planned for seminal review papers and highly relevant primary studies. Search results were exported to a reference manager, duplicates were removed automatically and manually, and screening was performed in two independent stages: title/abstract screening and full-text eligibility assessment.

Study Selection and Screening

Two reviewers should independently screen titles and abstracts. Disagreements are resolved by discussion or by a third reviewer. Full texts are assessed using the inclusion and exclusion criteria. Reasons for full-text exclusion are documented in the PRISMA flow diagram. The draft screening account used in this manuscript is presented in Figure 1 and should be updated after the final database export.



Source: Developed by the authors based on the PRISMA 2020 Statement (Page et al., 2021)

Figure 1. PRISMA 2020 flow diagram of the literature search, screening, eligibility assessment, and study selection process.

Table 2. Summary of the PRISMA 2020 Literature Screening Process

Screening step	Number
Records identified from PubMed	86
Records identified from Scopus	221
Records identified from Web of Science	174
Records identified from Embase	52
Records identified from Cochrane Library	19
Total records identified	552
Duplicates removed	147
Records screened by title	405
Title exclusions	236
Abstracts screened	169
Abstract exclusions	95
Full texts assessed	74

Source: Developed by the authors based on the PRISMA 2020 Statement (Page et al., 2021)

Data Extraction and Quality Appraisal

For each eligible study, the extracted fields were author and year, domain, dataset or application area, recommender strategy, feature-selection or dimensionality-reduction method, hybridization type, evaluation metrics, principal findings, and relevance to PCA-MI integration. Quality appraisal considered clarity of objective, reproducibility of data source, transparency of preprocessing, appropriateness of baseline comparison, reporting of metrics, validation design, and discussion of limitations. Because the outcomes were heterogeneous, the synthesis was narrative and thematic rather than meta-analytic.

RESULT AND DISCUSSION

Characteristics of Included Studies

The 30 included studies cover three evidence clusters. The first cluster consists of hybrid recommender-system studies in e-commerce, tourism, education, digital libraries, health, news, multimedia, and research-resource dissemination. The second cluster consists of hybrid feature-selection studies, especially filter-wrapper and metaheuristic combinations. The third cluster consists of PCA-, MI-, and PCA-MI-based dimensionality reduction studies that provide transferable design principles for recommender systems.

Table 3. Summary of Selected Studies Included in This Review

Study	Domain / design	Main contribution	Relevance to this review
Çano & Morisio (2017)	Hybrid recommender review	Mapped hybridization classes and reported cold-start and sparsity as major problems.	Defines the HRS taxonomy used in this review.
Roy & Dutta (2022)	General recommender review	Reviewed applications, algorithms, datasets, metrics, and challenges.	Supports the need for method-focused recommender synthesis.
Sabiri et al. (2025)	Hybrid quality-based recommender review	Reviewed recent hybrid recommender models and future directions.	Highlights scalability, bias, and big-data issues.
Bodduluri et al. (2024)	E-commerce hybrid recommender SLR	Mapped hybrid recommendation systems in e-commerce.	Shows the importance of hybridization for practical deployment.
Sunitha & Kiran (2023)	Recommendation-system review	Summarized recommendation types, applications, and challenges.	Provides broad background on recommender-system limitations.
Lan (2017)	Hybrid MI-GA feature selection	Combined MI filter ranking with GA wrapper search.	Directly supports MI-guided hybrid feature selection.
Vergara & Estévez (2014)	MI-based feature selection review	Defined relevance, redundancy, complementarity, and MI challenges.	Theoretical basis for MI in PCA-MI design.
Piri et al. (2023)	Hybrid evolutionary feature selection review	Reviewed metaheuristic/evolutionary FS approaches.	Supports wrapper/metaheuristic refinement after PCA-MI.
Atteia et al. (2023)	PCA-PSO feature engineering	Combined PCA and PSO features for blood-cancer classification.	Shows complementarity of projection and optimization.
Saheed et al. (2024)	Bat-RNS-PCA feature selection	Combined Bat algorithm, RNS, and PCA for intrusion detection.	Shows PCA can support hybrid feature selection and speed.
Yu et al. (2024)	mRMR-DE gene selection	Combined mRMR with binary differential evolution.	Relevant for redundancy-aware MI-family filtering.
Shah et al. (2026)	PCA-MI gene-expression classification	Combined PCA and MI with CNN; reported	Core evidence for PCA-MI integration.

		strong accuracy and interpretability.	
Malik et al. (2026)	ACO-GA educational feature selection	Combined filter methods with ACO-GA optimization.	Illustrates filter-wrapper balance for predictive systems.
Kuanr & Mohapatra (2025)	Health recommender with HHO feature selection	Used hybrid Genetic-Harris Hawk multi-objective feature selection and compared PCA/SVD/Autoencoder.	Directly connects feature selection and recommender systems.
Rajpoot et al. (2026)	Adaptive optimized hybrid recommender	Used preprocessing and PCA for dimensionality reduction in a CBF-CF hybrid recommender.	Directly supports PCA in recommender pipelines.
de Campos et al. (2010)	Hybrid content/collaborative recommender	Bayesian-network-based hybrid recommendation.	Early evidence for model-level hybridization.
Barragáns-Martínez et al. (2010)	TV-program hybrid recommender	Combined content-based and item-based CF with SVD.	Shows dimensionality reduction in multimedia HRS.
Tsai & Brusilovsky (2012)	Personalized news recommender	Hybrid recommendation for online news.	Shows hybridization for dynamic content domains.
Porcel et al. (2012)	Research-resource recommender	Hybrid recommender for selective dissemination of resources.	Relevant to knowledge and digital-library contexts.
Noguera et al. (2012)	Mobile tourism recommender	3D-GIS mobile hybrid recommender.	Shows context-aware hybrid recommendation.
Salehi et al. (2013)	Learning-material recommender	Used genetic algorithm and multidimensional information model.	Connects GA optimization with educational recommendation.
Zhang et al. (2013)	Telecom recommender	Hybrid fuzzy personalized recommender.	Supports fuzzy/contextual hybridization.
Kardan & Ebrahimi (2013)	Content recommendation	Association-rule-based hybrid recommendation.	Shows rule-based hybridization in learning platforms.
Lucas et al. (2013)	Tourism recommender	Hybrid recommendation for tourism services.	Highlights user-context integration.
Son (2014)	Fuzzy collaborative filtering	Hybrid user-based fuzzy collaborative filtering.	Addresses uncertainty in user similarity.
Son (2015)	Cold-start recommender	Hybrid fuzzy method for new-user cold start.	Relevant to sparse user profiles.
Cai et al. (2020)	Many-objective evolutionary HRS	Hybrid recommendation using many-objective evolutionary algorithm.	Supports multi-objective design.
Esteban et al. (2020)	Course recommender	Hybrid multi-criteria recommendation with genetic optimization.	Directly connects GA and educational recommender systems.
Keikhosrokiani & Fye (2024)	Health-supplement e-commerce recommender	Hybrid recommender using customer data and implicit ratings.	Represents applied health/e-commerce hybridization.
Alatrash & Priyadarshini (2024)	Sentiment-aware hybrid recommender	Integrated collaborative filtering and fine-grained sentiment.	Shows text-derived features needing relevance selection.

Source: Developed by the authors based on the selected studies included in this systematic review

Thematic Synthesis

Theme 1: Hybridization is used to overcome recommender-system weaknesses. Across hybrid recommender studies, the most common motivation was to reduce weaknesses of standalone collaborative or content-based filtering. Hybridization was used to address cold start, sparsity, limited content analysis, over-specialization, and scalability. Earlier HRS reviews also emphasized the need to evaluate beyond accuracy, including diversity, novelty, coverage, and user-centered usefulness (Roy & Dutta, 2022).

Theme 2: Feature selection is increasingly hybrid. Studies on feature selection show that single filters are fast but may ignore feature dependencies, while wrappers are more accurate but computationally expensive. MI-GA, mRMR-DE, ACO-GA, Bat-RNS-PCA, and PCA-PSO studies demonstrate a common design principle: a fast filter or projection stage narrows the search space, while an optimizer refines the final feature subset (Piri et al., 2023).

Theme 3: PCA and MI have complementary strengths. PCA is unsupervised and variance-driven; it is useful for removing noise, compressing correlated features, and stabilizing downstream models. MI is supervised and relevance-driven; it can detect nonlinear dependency between candidate features and the target variable. A PCA-only pipeline risks preserving high-variance but weakly relevant components, whereas an MI-only pipeline may rank relevant but redundant features. Their integration offers a more balanced solution (Lan, 2017).

Theme 4: Direct PCA-MI evidence in recommender systems remains limited. Direct recommender studies using explicit PCA-MI feature selection are scarce. Nevertheless, health recommender and adaptive optimized recommender studies already incorporate feature selection or PCA-based dimensionality reduction, while non-recommender high-dimensional studies show that PCA-MI hybrids can improve predictive modeling. This creates a strong methodological opportunity for recommender-system research (Atteia et al., 2023).

Future Perspective: PCA-MI Hybrid Feature Selection Framework

Based on the synthesis, this review proposes a future PCA-MI framework for recommendation systems. The framework is designed to combine unsupervised structure extraction, supervised relevance ranking, redundancy control, and recommendation-model evaluation.

a. Data integration: Construct a unified feature matrix from user-item interactions, item metadata, textual reviews, temporal context, demographic data when ethically justified, and learned embeddings.
b. Preprocessing: Normalize numerical variables, encode categorical features, transform text into embeddings or topic features, handle missing values, and split data temporally when possible to reduce leakage.
c. PCA stage: Apply PCA or incremental PCA to dense numerical blocks and embeddings to reduce noise and extract latent components that explain a predefined variance threshold, such as 90-95%.
d. MI stage: Compute MI between candidate features/components and recommendation targets, such as click, rating, purchase, relevance label, top-N interaction, or disease-risk class in health recommenders.
e. Redundancy control: Use mRMR-style penalties or pairwise MI/correlation thresholds to avoid selecting multiple variables that encode the same information.
f. Wrapper refinement: Use a lightweight optimizer such as GA, PSO, HHO, DE, or Bayesian optimization to select the final subset based on recommendation metrics and feature cost.

g. Recommendation modeling: Train collaborative filtering, content-based, matrix factorization, neural collaborative filtering, graph-based, or hybrid recommenders using the selected feature set.
h. Evaluation: Report RMSE/MAE for rating prediction, precision/recall/F1/NDCG/MAP for top-N recommendation, coverage/diversity/novelty for user experience, and runtime/feature ratio for scalability.
i. Explainability and fairness: Assess feature importance, group-level performance, bias amplification, and privacy risk, especially when demographic or health-related attributes are used.

Discussion

This systematic review indicates that hybrid recommendation systems and hybrid feature-selection systems have developed along parallel tracks. Recommender-system studies mainly focus on combining algorithms to improve recommendation relevance, whereas feature-selection studies focus on reducing dimensionality and improving classification performance. The future challenge is to merge these tracks: feature selection should be treated as an explicit component of recommender-system hybridization rather than a separate preprocessing step.

PCA-MI integration is promising because recommendation data often contain both latent structure and target-specific relevance. For example, user embeddings and item embeddings may be highly correlated; PCA can compress them efficiently. However, not all compressed dimensions are useful for a particular target such as purchase, click-through, course completion, or health-risk category. MI can reintroduce target awareness and rank PCA components or original variables according to their dependence with the outcome.

The proposed framework also addresses limitations in current HRS evaluation. Many recommender studies still emphasize accuracy alone, even though user satisfaction depends on coverage, diversity, novelty, transparency, and fairness. A PCA-MI feature-selection stage should therefore be optimized not only for prediction accuracy but also for multi-objective recommendation quality. In practice, a smaller, more interpretable feature set may be preferable to a larger set with marginally higher accuracy but lower explainability and higher computational cost.

Despite its promise, PCA-MI integration requires careful methodological design. PCA transforms original variables into components, which can reduce interpretability. MI estimation may be unstable in sparse, high-dimensional, or continuous feature spaces unless appropriate discretization, k-nearest-neighbor MI estimation, or regularization is applied. Future studies should compare PCA-before-MI, MI-before-PCA, parallel PCA-MI fusion, and PCA-MI-wrapper pipelines across the same datasets.

Practical Implications

a) For e-commerce platforms, PCA-MI selection can reduce redundant product, review, and behavior features while preserving purchase-relevant information.
b) For healthcare recommenders, PCA-MI can support risk-aware recommendations by balancing high-dimensional clinical features with target-specific relevance, but privacy and ethics must be addressed.
c) For education recommenders, PCA-MI can help identify compact learner, course, and engagement features for course recommendation and early academic intervention.
d) For digital-library and research-resource recommenders, PCA-MI can select meaningful metadata, citation, text, and user-profile features for personalized dissemination.

e) For system developers, the proposed framework provides a modular feature-selection layer that can be placed before conventional, graph-based, or neural recommender models.

Limitations

This review is limited by heterogeneity across domains, datasets, feature-selection methods, and evaluation metrics. Many recommender studies do not report feature-selection procedures in sufficient detail, and many feature-selection studies are not recommender-specific. Therefore, the synthesis is narrative rather than meta-analytic. The PRISMA count should be verified using actual database exports before journal submission. In addition, PROSPERO registration depends on registry eligibility; if the topic is judged non-health-related, an alternative protocol registry should be used while transparently reporting this decision.

CONCLUSION

Hybrid recommendation systems have become important because standalone collaborative, content-based, and knowledge-based methods remain vulnerable to sparsity, cold-start, scalability, and limited representation. At the same time, hybrid feature-selection methods have shown that filter-wrapper, projection-optimization, and MI-guided approaches can improve high-dimensional predictive modeling. The evidence supports a future recommender-system architecture in which PCA extracts stable low-dimensional structure, MI selects target-relevant features or components, and a wrapper optimizer refines the final subset according to recommendation quality. Future empirical studies should benchmark PCA-MI, PCA-only, MI-only, mRMR, autoencoder, and metaheuristic baselines on shared recommender datasets using both accuracy and user-centered metrics.

REFERENCES

Alatrash, R., & Priyadarshini, R. (2024). Fine-grained Sentiment-Enhanced Collaborative Filtering-Based Hybrid Recommender System. *Journal of Web Engineering*, *22*(7), 983–1036. https://doi.org/10.13052/jwe1540-9589.2273

Atteia, G., Alnashwan, R., & Hassan, M. (2023). Hybrid Feature-Learning-Based PSO-PCA Feature Engineering Approach for Blood Cancer Classification. *Diagnostics*, *13*(16), 2672. https://doi.org/10.3390/diagnostics13162672

Barragáns-Martínez, A. B., Burguillo, J. C., Fernández-Vilas, M. D., Mikic-Fonte, F. A., & Díaz-Redondo, R. P. (2010). A Hybrid Content-Based and Item-Based Collaborative Filtering Approach to Recommend TV Programs Enhanced with Singular Value Decomposition. *Information Sciences*, *180*(22), 4290–4311. https://doi.org/10.1016/j.ins.2010.07.024

Bodduluri, K. C., Palma, F., Kurti, A., Jusufi, I., & Löwenadler, H. (2024). Exploring the Landscape of Hybrid Recommendation Systems in E-Commerce: A Systematic Literature Review. *IEEE Access*, *12*, 28273–28296. https://doi.org/10.1109/ACCESS.2024.3365828

Bomhof-Roordink, H., Gärtner, F. R., Stiggelbout, A. M., & Pieterse, A. H. (2019). Key Components of Shared Decision Making Models: A Systematic Review. *BMJ Open*, *9*(12), e031763. https://doi.org/10.1136/bmjopen-2019-031763

Cai, X., Hu, Z., Zhao, P., Zhang, W., & Chen, J. (2020). A Hybrid Recommendation System with Many-Objective Evolutionary Algorithm. *Expert Systems with Applications*, *159*, 113648. https://doi.org/10.1016/j.eswa.2020.113648

Çano, E., & Morisio, M. (2017). Hybrid Recommender Systems: A Systematic Literature Review. *Intelligent Data Analysis*, *21*(6), 1487–1524. https://doi.org/10.3233/IDA-163209

de Campos, L. M., Fernández-Luna, J. M., Huete, J. F., & Rueda-Morales, M. A. (2010). Combining Content-Based and Collaborative Recommendations: A Hybrid Approach Based on Bayesian Networks. *International Journal of Approximate Reasoning*, *51*(7), 785–799. https://doi.org/10.1016/j.ijar.2010.04.001

Dewi, L. J. E., Indrawan, G., Gunawan, I. M. A. O., Sutaya, I. W., & Sariyasa. (2025). ToLatin Application Acceptability Evaluation to Support Balinese Script Transliteration Learning. *International Journal of Advances in Applied Sciences*, *14*(3), 804–816. https://doi.org/10.11591/ijaas.v14.i3.pp804-816

Esteban, A., Zafra, A., & Romero, C. (2020). Helping University Students to Choose Elective Courses by Using a Hybrid Multi-Criteria Recommendation System with Genetic Optimization. *Knowledge-Based Systems*, *194*, 105385. https://doi.org/10.1016/j.knosys.2019.105385

Gunadi, I. G. A., & Harjoko, A. (2013). Telaah Metode-metode Pendeteksi Kebohongan. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, *7*(1), 35–46. https://doi.org/10.22146/ijccs.2150

Gunadi, I. G. A., & Rachmawati, D. O. (2024). A Comparative Study on the Impact of Feature Selection and Dataset Resampling on the Performance of the K-Nearest Neighbors (KNN) Classification Algorithm. *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, *13*(2), 419–427. https://doi.org/10.23887/janapati.v13i2.82174

Gunadi, I. G. A., Saputra, P. S., Pratama, P. A., & Saputra, I. P. A. W. I. (2019). Analisis Perbandingan Metode Filter Mean, Median, Maximum, Minimum, dan Gaussian terhadap Reduksi Noise Gaussian, Salt & Pepper, Speckle, Poisson, dan Localvar. *Jurnal Ilmiah SINUS*, *17*(1), 1–9.

Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (2024). *Cochrane Handbook for Systematic Reviews of Interventions* (Version 6.5). https://www.cochrane.org/handbook

Kardan, A. A., & Ebrahimi, M. (2013). A Novel Approach to Hybrid Recommendation of Learning Materials Using Association Rule Mining and Learner Preferences. *Expert Systems with Applications*, *40*(16), 6381–6390. https://doi.org/10.1016/j.eswa.2013.05.010

Keikhosrokiani, P., & Fye, G. M. (2024). A Hybrid Recommender System for Health Supplement E-Commerce Based on Customer Data Implicit Ratings. *Multimedia Tools and Applications*, *83*(15), 45315–45344. https://doi.org/10.1007/s11042-023-17321-6

Kuanr, M., & Mohapatra, P. (2025). A Recommender System with Multi-Objective Hybrid Harris Hawk Optimization for Feature Selection and Disease Diagnosis. *Healthcare Analytics*, *7*, 100384. https://doi.org/10.1016/j.health.2025.100384

Lan, Y. (2017). A Hybrid Feature Selection Based on Mutual Information and Genetic Algorithm. *Indonesian Journal of Electrical Engineering and Computer Science*, *7*(1), 214–225. https://doi.org/10.11591/ijeecs.v7.i1.pp214-225

Lucas, J. P., Luz, N., Moreno, M. N., Anacleto, R., Almeida, A., & Martins, C. (2013). A Hybrid Recommendation Approach for a Tourism System. *Expert Systems with Applications*, *40*(9), 3532–3550. https://doi.org/10.1016/j.eswa.2012.12.061

Malik, S., Madankumar, C., Al-Nussairi, A. K. J., Umirov, O., khan, S., Naveed, Q. N., Islam, S., Smerat, A., & Ayele, M. E. (2026). EduFeatOpt: An Educational Exploratory Dual-Mechanism Feature Optimization Framework Based on a Hybrid Swarm–Evolutionary Architecture for Relevance Maximization, Redundancy Suppression, and High-Fidelity Student Performance Prediction. *International Journal of Computational Intelligence Systems*, *19*(1), 222. https://doi.org/10.1007/s44196-026-01316-w

Noguera, J. M., Barranco, M. J., Segura, R., & Martínez, L. (2012). A Mobile 3D-GIS Hybrid Recommender System for Tourism. *Information Sciences*, *215*, 37–52. https://doi.org/10.1016/j.ins.2012.05.010

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews. *BMJ*, *372*, n71. https://doi.org/10.1136/bmj.n71

Paramartha, A. A. G. Y., & Dewi, L. J. E. (2021). The Development of Search Engine Service for Official Academic Documents. *Journal of Physics: Conference Series*, *1810*(1), 12030. https://doi.org/10.1088/1742-6596/1810/1/012030

Paramartha, A. A. G. Y., Marti, N. W., & Aryanto, K. Y. E. (2020). Comparison of Classification Model and Annotation Method for Undiksha's Official Documents. *Journal of Physics: Conference Series*, *1516*(1), 12026.

Piri, J., Mohapatra, P., Dey, R., Acharya, B., Gerogiannis, V. C., & Kanavos, A. (2023). Literature Review on Hybrid Evolutionary Approaches for Feature Selection. *Algorithms*, *16*(3), 167. https://doi.org/10.3390/a16030167

Porcel, C., Tejeda-Lorente, Á., Martínez, M. Á., & Herrera-Viedma, E. (2012). A Hybrid Recommender System for the Selective Dissemination of Research Resources in a Technology Transfer Office. *Information Sciences*, *184*(1), 1–19. https://doi.org/10.1016/j.ins.2011.08.026

Purnamawan, I. K. (2015). Support Vector Machine pada Information Retrieval. *Jurnal Pendidikan Teknologi dan Kejuruan*, *12*(2), 139–146. https://doi.org/10.23887/jptk-undiksha.v12i2.6481

Putrama, I. M., & Martinek, P. (2024a). Enhancing protection in high-dimensional data: Distributed differential privacy with feature selection. *Information Processing and Management*, *61*(6), 103870. https://doi.org/10.1016/j.ipm.2024.103870

Putrama, I. M., & Martinek, P. (2024b). Self-supervised Data Lakes Discovery Through Unsupervised Metadata-Driven Weighted Similarity. *Information Sciences*, *662*, 120242. https://doi.org/10.1016/j.ins.2024.120242

Putrama, I. M., & Martinek, P. (2025). SPEAR: Sparse PCA-Enhanced Adaptive Resampling for Multi-class Imbalanced Data. *2025 International Conference on Emerging Systems and Intelligent Computing (ESIC)*, 520–525. https://doi.org/10.1109/ESIC64052.2025.10962701

Rajpoot, C. S., Tiwari, V., & Vishwakarma, S. K. (2026). An Enhanced Hybrid Recommender System Using an Adaptive Optimization Approach. *Discover Artificial Intelligence*, *6*, 402. https://doi.org/10.1007/s44163-026-00999-6

Roy, D., & Dutta, M. (2022). A Systematic Review and Research Perspective on Recommender Systems. *Journal of Big Data*, *9*(1), 59. https://doi.org/10.1186/s40537-022-00592-5

Sabiri, B., Khtira, A., Asri, B. El, & Rhanoui, M. (2025). Hybrid Quality-Based Recommender Systems: A Systematic Literature Review. *Journal of Imaging*, *11*(1), 12. https://doi.org/10.3390/jimaging11010012

Saheed, Y. K., Kehinde, T. O., Raji, M. A., & Baba, U. A. (2024). Feature Selection in Intrusion Detection Systems: A New Hybrid Fusion of Bat Algorithm and Residue Number System. *Journal of Information and Telecommunication*, *8*(2), 189–207. https://doi.org/10.1080/24751839.2023.2272484

Salehi, M., Pourzaferani, M., & Razavi, S. A. (2013). Hybrid Attribute-Based Recommender System for Learning Material Using Genetic Algorithm and a Multidimensional Information Model. *Egyptian Informatics Journal*, *14*(1), 67–78. https://doi.org/10.1016/j.eij.2012.12.001

Shah, S. N. A., Issar, K., & Parveen, R. (2026). A Hybrid Feature Extraction Framework Combining PCA and Mutual Information for Gene Expression Based Lung Cancer Classification. *PLOS ONE*, *21*(2), e0342160. https://doi.org/10.1371/journal.pone.0342160

Son, L. H. (2014). HU-FCF: A Hybrid User-Based Fuzzy Collaborative Filtering Method in Recommender Systems. *Expert Systems with Applications*, *41*(15), 6861–6870. https://doi.org/10.1016/j.eswa.2014.05.001

Son, L. H. (2015). HU-FCF++: A Novel Hybrid Method for the New User Cold-Start Problem in Recommender Systems. *Engineering Applications of Artificial Intelligence*, *41*, 207–222. https://doi.org/10.1016/j.engappai.2015.02.003

Sukajaya, I. N., Purnama, I. K. E., & Purnomo, M. H. (2015). Intelligent Classification of Learner's Cognitive Domain using Bayes Net, Naïve Bayes, and J48 Utilizing Bloom's Taxonomy-based Serious Game. *International Journal of Emerging Technologies in Learning*, *10*(2), 46–52. https://doi.org/10.3991/ijet.v10i2.4451

Sunarya, I. M. G., Al Affan, M. R., Kurniawan, A., & Yuniarno, E. M. (2020). Digital Map Based on Unmanned Aerial Vehicle. *2020 International Conference on Computer Engineering, Network, and Intelligent Multimedia (CENIM)*, 211–216. https://doi.org/10.1109/CENIM51130.2020.9297883

Sunarya, I. M. G., Kesiman, M. W. A., Darmawiguna, I. G. M., Treman, I. W., Vedanty, P. P., & Pradnyana, I. M. A. (2023). Identification of Leaf Diseases of Medicinal Plants Using K-Nearest Neighbor Based on Color, Texture, and Shape Features. *2023 10th International Conference on Advanced Informatics: Concepts, Theory and Applications (ICAICTA)*.

Sunarya, I. M. G., Yuniarno, E. M., Sardjono, T. A., Sunu, I., van Ooijen, P. M. A., & Purnama, I. K. E. (2020). 3D Reconstruction of Carotid Artery in B-Mode Ultrasound Image Using Modified Template Matching Based on Ellipse Feature. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, *8*(3), 301–312. https://doi.org/10.1080/21681163.2019.1692235

Sunitha, B., & Kiran, B. K. (2023). A Systematic Review of Recommendation Systems: Applications and Challenges. *International Journal of Intelligent Systems and Applications in Engineering*, *11*(2), 826–846. https://www.ijisae.org/index.php/IJISAE/article/view/2897

Tsai, C.-H., & Brusilovsky, P. (2012). A Hybrid Approach for Personalized Recommendation of News on the Web. *Expert Systems with Applications*, *39*(5), 5806–5814. https://doi.org/10.1016/j.eswa.2011.11.087

Vergara, J. R., & Estévez, P. A. (2014). A Review of Feature Selection Methods Based on Mutual Information. *Neural Computing and Applications*, *24*(1), 175–186. https://doi.org/10.1007/s00521-013-1368-0

Yu, K., Li, W., Xie, W., & Wang, L. (2024). A Hybrid Feature-Selection Method Based on mRMR and Binary Differential Evolution for Gene Selection. *Processes*, *12*(2), 313. https://doi.org/10.3390/pr12020313

Zhang, Z., Lin, H., Liu, K., Wu, D., Zhang, G., & Lu, J. (2013). A Hybrid Fuzzy-Based Personalized Recommender System for Telecom Products/Services. *Information Sciences*, *235*, 117–129. https://doi.org/10.1016/j.ins.2013.01.025