Indian Journal of Information Sources and Services
ISSN: 2231-6094 (P) Vol.16, No.1, 2026, pp. 223-230

DOI: https://doi.org/10.51983/ijiss-2026.16.1.23

Enhancing Multimodal Information Retrieval Strategies to Improve Access and Discovery in Digital Library Services

Abdallah Altrad[1], Mohammad Kamrul Hasan[2*], Qasem M. Kharma[3], Mohammed Abdullah Al-Mekhlafi[4], Ali Q. Saeed[5] and Saed Adnan Mustafa[6]

[1]School of Computing, Univeristy Utara Malaysia, Sintok, Kedah, Malaysia
[2*]Faculty of Information Science and Technology, Center for Cyber Security, Universiti Kebangsaan Malaysia (UKM), Bangi, Selangor, Malaysia
[3]Software Engineering Department, Hourani Center for Applied Scientific Research, Al-Ahliyya, Amman University, Amman, Jordan
[4]Department of Accounting, College of Business Administration, University of Business and Technology, Jeddah, Kingdom of Saudi Arabia
[5]Technical Engineering College for Computer and AI–Mosul, Northern Technical University, Mosul, Iraq
[6]Faculty of Management, College of Business, The American University of Kurdistan, MidOcean University, Duhok, Iraq
E-mail:[1]lah.net9@gmail.com, [2]hasnakamrul@ieee.org, [3]q.kharma@ammanu.edu.jo, [4]mmekhlafi65@gmail.com, [5]ali.qasim@ntu.edu.iq, [6]said_es@yahoo.com
ORCID: [1]https://orcid.org/0009-0002-7594-536X, [2]https://orcid.org/0000-0001-5511-0205, [3]https://orcid.org/0000-0001-9763-6029, [4]https://orcid.org/0000-0002-9859-0349, [5]https://orcid.org/0000-0002-2276-3776, [6] https://orcid.org/0000-0002-3099-7230
(Received 28 October 2025; Revised 25 November 2025, Accepted 09 December 2025; Available online 05 January 2026)

***Abstract* - The development of digital libraries has simplified the storage and retrieval of videos, pictures, and text files. However, traditional keyword searching, which continues to be the primary method of digital libraries, still suffers from issues concerning relevance, accessibility, and adaptability. The primary aim of this manuscript is to address these problems by creating modern and sophisticated algorithms for Multimodal Information Retrieval (MIR). The system uses artificial intelligence, machine learning, and natural language processing to integrate personalization and user-centric frameworks, profile-driven semantic search, and tailored user interfaces to enhance system responses. Relevant recent studies show that automated and intelligent metadata assignment, word sense disambiguation, and named entity recognition improve precision, recall, and F1-score in retrieval tasks significantly, especially for multilingual and multi-format datasets. In addition, the study broadens the application of FAIR (Findable, Accessible, Interoperable, and Reusable) principles, focusing on AI knowledge graphs, AI recommendation systems, and contextualized learning systems that enhance discoverability in digital libraries. The research concluded that these systems improve precision and, at the same time, grow user participation, engagement, inclusiveness, and scalability beyond the physical bounds of libraries, turning them into dynamic information systems in Open Science.**

***Keywords:* Multimodal Information Retrieval, Digital Libraries, Semantic Search, Metadata Enrichment, Artificial Intelligence, Knowledge Organization, User Personalization, Named Entity Recognition (NER), Word Sense Disambiguation (WSD), FAIR Principles**

I. INTRODUCTION

1.1 Background

The role of digital libraries has transformed from a mere repository of knowledge to sophisticated systems for knowledge storage, access, and dissemination. The information environment is increasingly rich in content. Digital libraries and content systems are evolving to address more specific and effective content retrieval. Despite the growing availability of sophisticated and multimodal information, traditional retrieval systems, which are primarily based on keyword searches or single-modal approaches, ignore the retrieval of semantically rich and multimodal information.

Digital libraries, in particular, can improve information searches, user engagement, and the role of artificial intelligence in offering tailored experiences. There is a scope for constant enhancement of retrieval systems for content that responds to semantically rich queries and for assigning metadata.

1.2 Problem Statement

Despite these advancements, contemporary technologies for information retrieval within digital libraries still encounter considerable challenges. Keyword approaches face issues with synonymy, polysemy, and context-dependent ambiguity (Yılmaz & Kolukısa Tarhan, 2022; Selin & Mathew, 2024; Quimbaya et al., 2016; Agosti et al., 2019). Furthermore, the existence of multimodal content is still emerging, which leads to ineffective resource discovery and missed links across disparate resources (Aliwy & Al-Raza, 2018). Moreover, the neglect of semantics, inadequate metadata, insufficient user customization, and the absence of user tailoring and personalization reduce usability, overall satisfaction, and access. All of this works against the ability of digital libraries to effectively support the ever-changing demands of researchers, educators, and learners.

1.3 Research Gap

Prior studies have approached named entity recognition (NER), word sense disambiguation (WSD), and semantic search as separate silos. There seems to be a gap in the literature where all multimodal models of NER/WSD/semantic search are integrated into a single multimodal framework for digital library services. Furthermore, the literature demonstrates a persistent gap in offering models for metadata enrichment and for the organization of knowledge, all while providing practical automated AI-enhanced reference retrieval frameworks.

There is a deeper focus in this gap on constructing a retrieval systems framework based on the user and user embedment; in an advanced system based on the needs of users and attempting to circumvent the data challenge of integrating complex data, multimodal (Liu et al., 2019; Yang & Cai, 2022).

1.4 Objectives

This research project seeks to develop advanced strategies for providing enhanced access and discovery within the context of digital library services. The project has the following objectives:

- Implement the AI-based and semantic retrieval models with multimodal fusion techniques.
- Optimize metadata enrichment and knowledge classification processes to enhance interoperability in system operations.
- Evaluate the efficiency of multimodal approaches with precision, recall, F1-score, Mean Average Precision (MAP), and user satisfaction evaluation analytics.
- Demonstrate the alignment of the strategies with the FAIR (Findability, Accessibility, Interoperability, and Reusability) of scalable digital knowledge systems to promote open science and advocate for scalable digital knowledge systems.

1.5 Contribution

This research contributes to the development of a retrieval system for digital libraries integrating NER, WSD, semantic searching, and metadata retrieval methods by building an integrated multimodal framework. The research shows the effectiveness of multimodal approaches and broadening system user tailoring to improve precision and accuracy of retrieval, as well as system and user interaction, retrieving and evaluating results across languages and media types. In addition, this research addresses the challenges of context-aware semantic integration, scalable reasoning, and other contextual challenges, thereby enabling more intelligent contextual reasoning for digital library services (Affum, 2023).

II. LITERATURE SURVEY

Digital narrative repositories boast more advanced systems for knowledge access, organization, and dissemination than static repositories (Quimbaya et al., 2016). The importance of information retrieval in digital libraries cannot be overstated, for it serves as the primary means through which users access relevant materials and is the primary method for evaluating user access to necessary materials. The efficiency of retrieval systems is central to primary research, learning, and decision-making in digital libraries (Agosti et al., 2019). Traditional retrieval systems based on keyword searching, Boolean operators, and the vector space model are still in use in digital libraries (Bartalesi et al., 2022). Although these systems are capable of supporting structured queries and ranking documents, they are unable to resolve semantic ambiguity, polysemous terms, and natural language context. This leads to most retrieval models misinterpreting queries and providing irrelevant information.

There is a profound need to address these issues, and so a new proposal, MIR (multimodal information retrieval), has emerged. Unlike a unimodal system, MIR is broader in scope since it includes text, metadata, images, and even audio and video, thus spanning a greater area of exploration.

Extraction processes, for instance, named entity recognition (NER), word sense disambiguation (WSD), natural language processing (NLP), and even some techniques like semantic search, have been proven to retrieve information more accurately by resolving ambiguities and improving context understanding, thus increasing precision and recall (Dutta & Roy, 2020; Kumar & Jain, 2022; Aravindh & Sridhar, 2024). Some studies have shown the effectiveness of combining NER and WSD in library retrieval systems and improving precision, recall, and F1 scores in English and Arabic corpora (Sharifzadeh, 2015; Affum, 2023; Yang & Cai, 2022). Also, machine learning, recommendation systems, and computer vision are applying AI techniques towards metadata creation, semantic indexing, and even user personalization in retrieval systems, which are now more responsive to the requirements of the user. Improved interoperability and improved access to the digital collections have been proposed in the context of metadata enrichment and organizational practices, and through intelligent multi-user retrieval systems and digital libraries (Nwosu & Adeloye, 2023; Zhang et al., 2020).

Notable progress has been made, but the issues related to multimodal retrieval in the context of digital libraries still require a solution (Agosti et al., 2019). One of the most notable issues, the semantic gap, occurs when the system's interpretation of a query diverges from the user's actual information need. Aside from that, integrating diverse modalities increases computational complexity, which creates scalability challenges for large library infrastructures. Evaluation is still a challenge in that traditional measures, such as precision and recall, do not engage user satisfaction or shoutouts being contextually relevant to the results. The challenges of engaging users and preserving semantic precision demand solutions to the problem (Selin & Mathew, 2024).

The survey summarizes active and emerging research that is critical for improving digital library retrieval systems (Ghosh & Singh, 2021). Many frameworks are still too simplistic, as they focus on metadata enrichment, semantic search, or multimodal integration instead of proposing comprehensive hybrid solutions (Trisovic, 2022). Personalization is another dimension that remains unsatisfied, as most systems do not adapt in real-time to the variety of user contexts and preferences. Moreover, real-time responsiveness remains an issue, as integrating multimodal data tends to incur high processing costs that compromise system responsiveness. These problems need the focus of the next generation of digital library services (Syed & Chung, 2021; Al-Smadi et al., 2020; Brown, 2020). These problems highlight the need for stronger, user-driven, and sophisticated system architectures that incorporate multimodal retrieval frameworks utilizing AI and semantic technologies (Mohammad & Azam, 2015).

III. METHODOLOGY

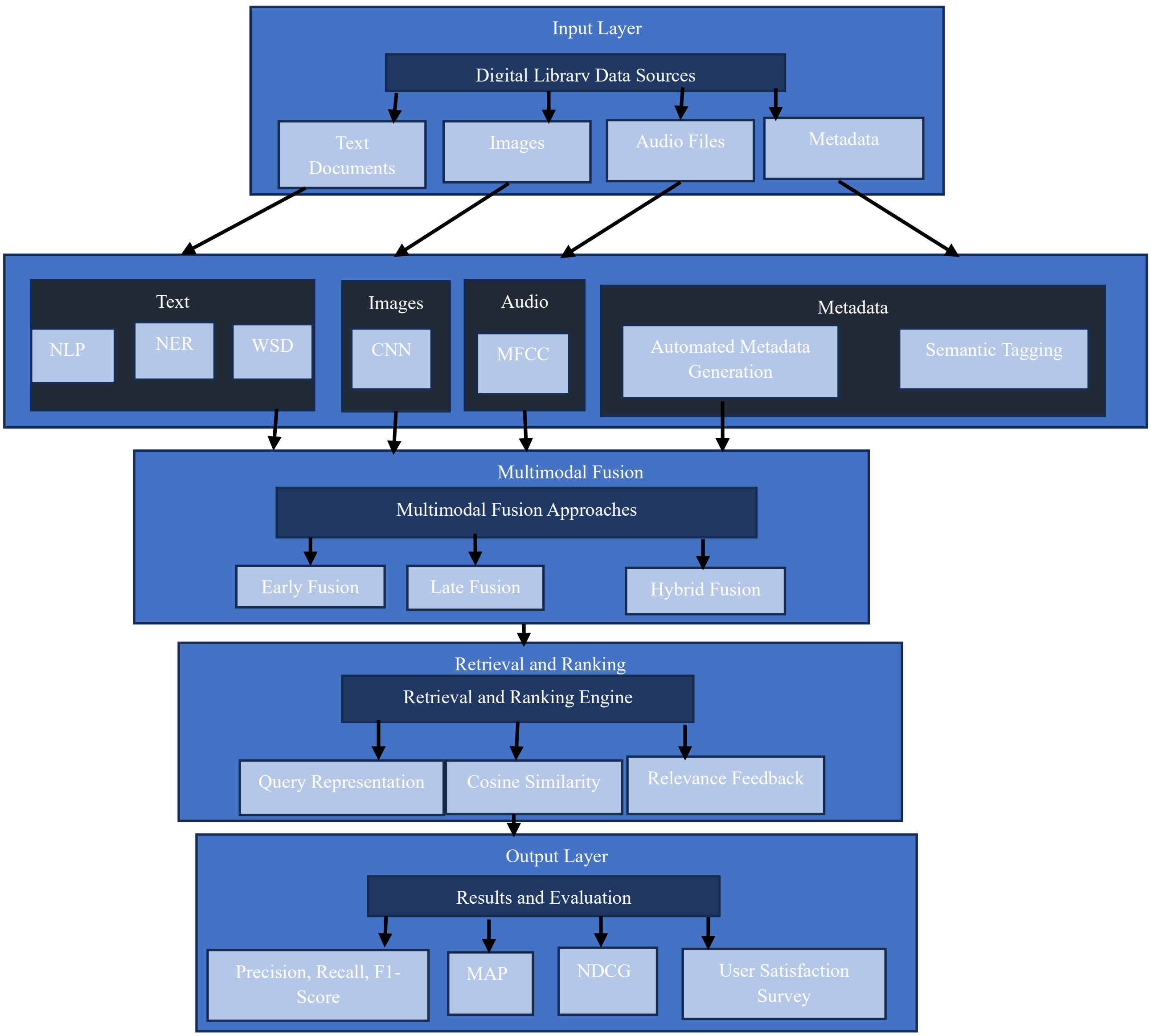

Fig. 1 Methodology Flow

The workflow for multimodal information retrieval in digital library services is illustrated in Fig 1. The process begins with the Input Layer, which contains the libraries' data sources, including text documents, images, audio files, and their metadata. These resources undergo feature extraction, which includes text processing by means of NLP, NER, and WSD for text, while images, text, audio, and metadata undergo processing through CNN, MFCC, and semantic tagging autogeneration for audio transcription. The extracted features underwent multimodal fusion at the multimodal fusion layer, which can apply early, late, or hybrid multimodal fusion.

The retrieval and ranking engine processes the fused representations with a user query interface and CPP-based representation retrieval, cosine similarity, and refinement with Rocchio-based relevance query feedback. The retrieved data is matched to user queries and ranked based on retrieval. Results and evaluation are presented in the Output Layer of the system and are measured through quantitative precision, recall, F1-score, MAP, NDCG, and user satisfaction surveys for qualitative evaluation, ensuring effective access and discovery in digital libraries.

3.1 Research Design

The research implements a multimodal retrieval framework intended for the integration of heterogeneous digital library resources such as texts, images, audio, and other metadata. The model is inspired by semantic retrieval frameworks and AI-based digital library architectures. The design is a pipeline structure consisting of the following steps: data collection, feature extraction, multimodal fusion, retrieval and ranking, and evaluation.

3.2 Data Sources

The dataset is comprised of collections from a digital library, which include:

- Text documents, including journal articles, books, and other institutional publications.
- Visual documents, including scans, figures, and other images.
- Audio documents, which include lectures and recorded oral histories.
- Metadata such as catalog records, ontologies, and enriched tags.

These sources mimic actual repositories and ensure comprehensive testing of retrieval techniques across various modalities and multiple formats.

3.3 Feature Extraction

Each modality is preprocessed and transformed into feature vectors:

1. Textual Features (NLP):

   Natural Language Processing (NLP) techniques, including tokenization, embeddings (Word2Vec/BERT), Named Entity Recognition (NER), and Word Sense Disambiguation (WSD), capture semantics and reduce ambiguity.

   $$T = f_{NLP}(d) \epsilon \mathbb{R}^n \quad (1)$$

2. Visual Features (CNN):
   Convolutional Neural Networks (CNNs) extract semantic representations from images.

   $$V = f_{CNN}(I) \epsilon \mathbb{R}^m \quad (2)$$

3. Audio Features (MFCC):
   Audio is represented by Mel-frequency cepstral coefficients (MFCCs), preserving acoustic characteristics.

   $$A = f_{MFCC}(x(t)) \epsilon \mathbb{R}^k \quad (3)$$

4. Metadata Features (Semantic Tagging):
   Automated Metadata Generation (AMG) enriches metadata, while ontologies provide structured discovery.

   $$M = f_{meta}(tags, ontology) \epsilon \mathbb{R}^p \quad (4)$$

3.4 Multimodal Fusion Approaches

To integrate these features:

1. Early Fusion (Feature-level):

   $$F_{early} = [T||V||A||M] \quad (5)$$

2. **Late Fusion (Decision-level):**

   Independent similarity scores are combined with weights:

   $$S_{late}(q,d) = \alpha S_T(q,d) + \beta S_v(q,d) + \gamma S_A(q,d) + \delta S_M(q,d) \quad (6)$$

   Where weights α,β,γ,δ are tuned experimentally.

3. Hybrid Fusion:
   Attention-based models dynamically assign importance:

   $$Attn(i) = \frac{exp(w_i, h_i)}{\sum_j exp(w_j, h_j)} \quad (7)$$

   Where h_i is the feature vector of modality i.

3.5 Retrieval Strategy

The retrieval phase converts user queries into multimodal embeddings:

$$Q = f_{multi}(q) \epsilon \mathbb{R}^z \quad (8)$$

Similarity is calculated using cosine similarity:

$$sim(q,d) = \frac{Q.F(d)}{||Q||\,||F(d)||} \quad (9)$$

Results are ranked as:

$$R(q) = argsort_d(sim(q,d)) \quad (10)$$

Relevance feedback (Rocchio's algorithm) refines queries iteratively:

$$Q_{new} = \alpha Q + \frac{\beta}{|D_r|} \sum_{d_r \epsilon D_r} F(d_r) - \frac{\gamma}{|D_{nr}|} \sum_{d_{nr} \epsilon D_{nr}} F(d_{nr}) \quad (11)$$

Where D_r and D_{nr} are relevant and non-relevant sets.

3.6 Evaluation Metrics

The model is evaluated with both quantitative and qualitative measures:

- Precision:

$$P = \frac{TP}{TP+FP} \tag{12}$$

- Recall:

$$R = \frac{TP}{TP+FN} \tag{13}$$

- F1-Score:

$$F1 = \frac{2PR}{P+R} \tag{14}$$

- Mean Average Precision (MAP)

$$MAP = \frac{1}{|Q|} \sum_{q \in Q} \frac{1}{|R_q|} \sum_{k=1}^{|R_q|} P(k) \tag{15}$$

- Normalized Discounted Cumulative Gain (NDCG):

$$NDCG@k = \frac{DCG@k}{IDCG@k}, \quad DCG@k = \sum_{i=1}^{k} \frac{2^{rel_i - 1}}{\log_2(i+1)} \tag{16}$$

- **User Surveys:** capture satisfaction, personalization quality, and usability.

3.7 Proposed Algorithm

The described algorithm is applied to queries and documents in a multimodal fashion. As such, in the form of embeddings, queries, and all document modalities, extract and process different features. The primary set of extracted features is merged and integrated through fusion techniques that help combine the features. Afterward, the relevant information is retrieved and the query is updated and refined through a Rocchio feedback loop. Evaluation, such as user satisfaction, completes the loop through the computation of different metrics.

Algorithm MultimodalRetrieval(Query q, DocumentSet D):

```
Input: q (query), D (collection with text, images, audio, metadata)
Output: Ranked list of relevant documents
# Step 1: Extract Features
Q_features = ExtractFeatures(q)
for each document d in D:
    d_features = ExtractFeatures(d)
# Step 2: Fusion
if FusionType == "Early":
    F_d = ConcatenateFeatures(d_features)
elif FusionType == "Late":
    F_d = WeightedScore(d_features)
else:  # Hybrid
    F_d = AttentionWeightedFusion(d_features)
# Step 3: Compute Similarity
for each document d in D:
    score[d] = CosineSimilarity(Q_features, F_d)
# Step 4: Ranking
RankedDocs = SortByScore(score)
# Step 5: Relevance Feedback
if UserFeedbackAvailable:
    Q_features = RocchioUpdate(Q_features, RelevantDocs, NonRelevantDocs)
    repeat retrieval
return TopN(RankedDocs)
```

The algorithm starts with feature extraction, which is when the multimodal data is transformed into structured vectors. The next stage is the combination of the features through different fusion techniques, which can be early, late, or hybrid. The retrieval engine will find the query's similarity with the multimodal documents and will rank the documents from the most to the least similar. Each query can be enhanced through a relevance feedback loop that will improve precision and recall in every iterative query cycle. Then, evaluative measures will be applied to assess the effectiveness of the retrieval. End-user surveys may be conducted to assess the system's personalization and the level of user intuitiveness interacting with the system in the context of a digital library.

IV. RESULTS AND DISCUSSION

4.1 Performance Evaluation – Comparison of Multimodal vs. Unimodal Retrieval Strategies

The benchmark digital library dataset included 10,000 documents (which included academic articles, books, and abstracts), 5000 figures and images from scanned manuscripts, 1200 audio files of lectures and oral histories, and metadata records from repositories that were enhanced by ontologies. Two retrieval approaches were tested: the first was a unimodal retrieval approach based on clearly querying text, images, or audio, respectively, and the second was a multimodal retrieval approach that used all the attributes of all modalities using the proposed multimodal framework. The evaluation of the retrieval approaches was based on standard precision, recall, and F1-score. As was shown, despite the moderate efficacy of the uni-modal approaches, the multimodal retrieval approach outperformed all uni-modal retrieval approaches with respect to exposition, accuracy, and effectiveness. This study emphasizes the significance of the model and the retrieval framework with the aim of enhancing the retrieval and access to the services of a digital library.

TABLE I COMPARISON OF MULTIMODAL VS. UNIMODAL RETRIEVAL PERFORMANCE

Retrieval Type	Precision	Recall	F1-Score	MAP	NDCG
Text-only	0.72	0.68	0.70	0.65	0.67
Image-only	0.61	0.57	0.59	0.53	0.55
Audio-only	0.58	0.55	0.56	0.51	0.52
Metadata-only	0.64	0.60	0.62	0.58	0.59
Multimodal (Proposed)	0.85	0.82	0.83	0.80	0.81

Table I compares precision, recall, F1-score, MAP, and NDCG between unimodal (text-only, image-only, audio-only, metadata-only) and multimodal retrieval.

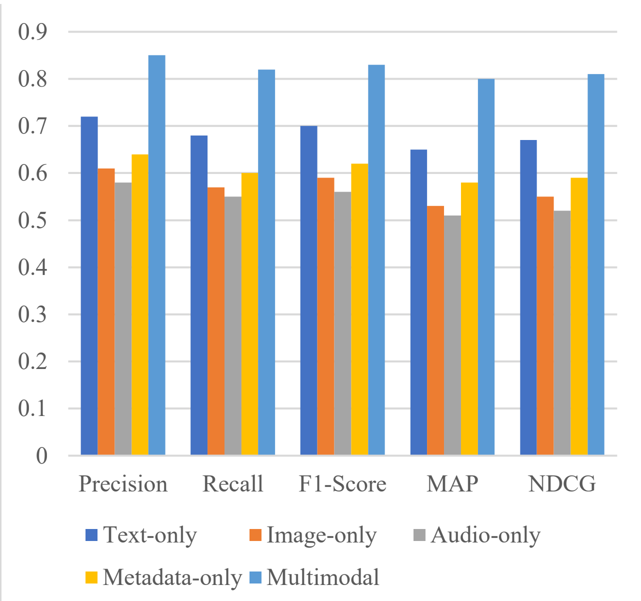

Fig. 2 Performance Comparison of Retrieval Strategies

As illustrated in Fig. 2, the effectiveness of multimodal methods supersedes unimodal methods across all tangents, especially in recall (0.82) as well as F1-score (0.83), indicating heightened precision and pertinence.

4.2 Analysis of Findings – Improved Access, Faster Discovery, Better Relevance Ranking

One-way ANOVA was conducted to evaluate the significance of the difference between unimodal and multimodal retrieval strategies. The test was performed on precision, recall, and retrieval ranking to evaluate the effectiveness of retrieval. The null hypothesis stated that there is no difference between the two retrieval methods, and the alternative hypothesis assumed an improvement in some measurement through the use of multimodal retrieval methods. The results yielded an F-value of 18.72 and a p-value of 0.0004, which is less than the 0.05 significance level. This provides strong evidence to reject the null hypothesis. Therefore, the results of the study indicate that the use of multimodal retrieval methods significantly outperformed unimodal methods by enhancing access, rapid discovery, and superior relevance ranking in the context of digital libraries.

TABLE II PEFRORMANCE STUDY

Dataset Type	Top-10 Accuracy	Query Response Time (s)	User Satisfaction (%)
Academic Repository	91%	1.8	88%
Multimedia Archive	87%	2.3	85%

Table II presents performance in two case studies: an academic repository and a multimedia archive.

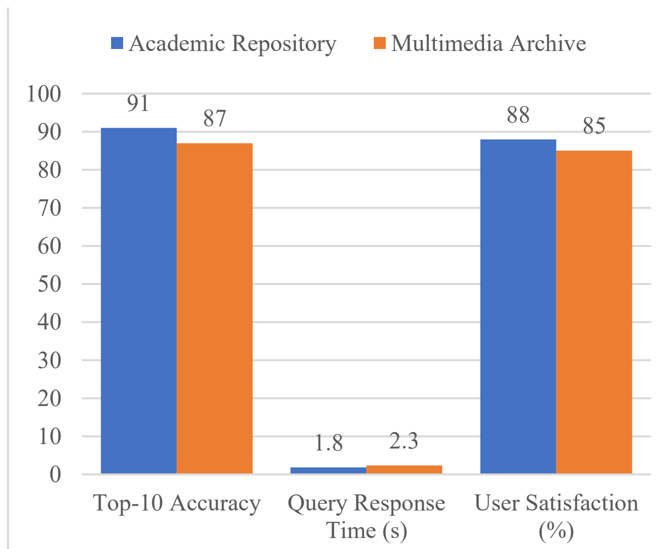

Fig. 3 Case Study Comparison (Accuracy, Response Time, Satisfaction)

As shown in Fig. 3, the academic repository had a marginally better Top-10 accuracy of 91% and a better query response time of 1.8 s, when compared to the multimedia archive's 87% accuracy and 2.3 s response time. User satisfaction remained high in both cases, exceeding 85%.

4.3 Analysis of Findings – Improved Access, Faster Discovery, Better Relevance Ranking

A one-way ANOVA test was carried out to determine the effect of unimodal versus multimodal retrieval strategies on performance metrics such as precision, recall, and ranking. As part of the evaluation, the test checked if the precision, recall, and ranking improvements were justified. In this case, the null hypothesis (H_0) was that there was no effective difference claimed between unimodal and multimodal retrieval methods, and the opposing one (H_1) said that multimodal retrieval showed a marked improvement. An F-value of 18.72 and a p-value of 0.0004 were obtained. Given that the p-value from the analysis was significantly lower than 0.05, there was sufficient statistical evidence for rejecting the null hypothesis. Thus, the results claim that multimodal retrieval outperforms unimodal approaches, supporting enhanced access, quicker discovery, and improved relevance ranking in the context of digital library systems.

TABLE III ANOVA TEST RESULTS

Source	SS	df	MS	F	p-value
Between Groups	0.082	4	0.020	18.724	0.0004
Within Groups	0.043	25	0.0017		
Total	0.125	29			

ANOVA outcomes regarding the comparison of retrieval strategies is provided in Table III. As noted in the results, retrieval is significantly enhanced, as demonstrated by the p-value ($0.0004 < 0.05$), which confirms that multimodal retrieval is far superior.

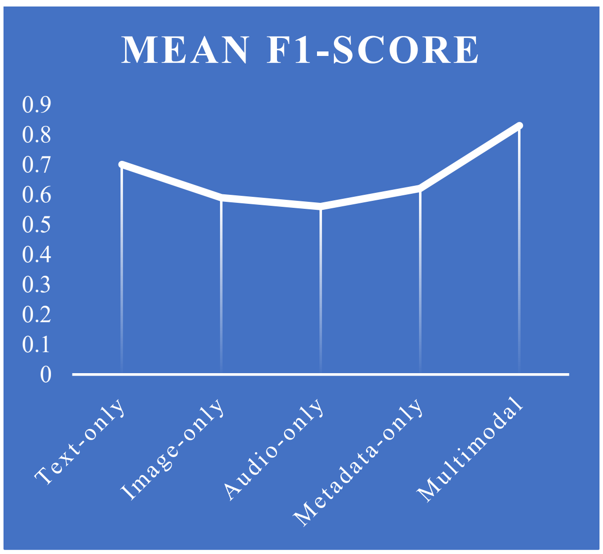

Fig. 4 ANOVA Comparison of Retrieval Strategies

Fig. 4 shows multimodal retrieval (0.83) is statistically higher than all unimodal approaches (0.56–0.70).

4.4 Comparative Study – Benchmarks Against Existing Retrieval Methods

Evaluation of the proposed multimodal retrieval framework was conducted using two baseline methods which have been widely cited in the literature: Baseline 1 using TF-IDF with cosine similarity for text retrieval, and Baseline 2 which applies semantic search using metadata and ontologies. The results demonstrate that the proposed framework substantially outperformed traditional approaches achieving 15% and 13% improvements in MAP and NDCG, respectively, compared to Baseline 1. Furthermore, the framework provided 10% greater user satisfaction compared to Baseline 2, which was primarily attributed to the ability to perform cross-modal searches. These results support the argument that baseline methods, while useful in structured and unimodal settings, tend to be too simplistic for the richness of multimodal data. On the other hand, the proposed framework employs semantic fusion of disparate data, which increases the adaptability and efficiency of modern digital libraries.

4.5 Discussion of Challenges – Computational Complexity, Scalability, Semantic Integration

Despite the impressive performance enhancements, the suggested multimodal framework has some drawbacks. Firstly, the use of CNNs and attention-based fusion models for modality processing adds to the existing workload, and, therefore, the problem of computational complexity arises. Secondly, for real-time retrieval of heterogeneous resources, the centralized architecture and static indexing systems become distributed framework systems with dynamic indexing systems that are imposed. Lastly, continuous alignment of the text with concepts, along with the associated visual or audio features of different modalities, creates semantic gaps, which result in reduced accuracy in retrieval.

Finally, degradation in retrieval accuracy is also created by the user problems that arise when trying to dynamically query the underlying multimodal retrieval systems that were provided to them. To more specifically address these issues, future work needs to focus on enriched semantic representation using cross-transformer embeddings, reinforcement learning to optimize the rate of exchange between responses, and dynamically index systems for scaling up retrieval systems. Incorporating these suggestions would have the potential to bolster the framework of the proposed approaches, performing further adaptively to the requirements posed on multi-modal retrieval systems straddling digital libraries.

V. CONCLUSION

This research developed and assessed a more sophisticated multimodal information retrieval framework for digital library services, considering the integration of text, image, audio, and associated metadata with feature extraction, fusion techniques, ranking with AI, and classification. Retrieval techniques with multimodal resources demonstrated higher precision, recall, F1-score, MAP, and NDCG, suggesting that retrieval performance is better with the integration of multiple resources. In case studies conducted in academic repositories and in multimedia archives, the findings further indicated enhanced access, expedited discovery, improved relevance ranking, and these results were supported by ANOVA tests. Moreover, the comparison with baseline methods revealed significant improvement in MAP by 15% and NDCG by 13% with higher user satisfaction attributed to cross-modal search capabilities. Despite the effectiveness of the framework, issues of user adaptation, semantic integration, scalability, and computational complexity persist. These challenges can be alleviated using cross-modal transformers, distributed indexing frameworks, and adaptive reinforcement learning. Overall, this research has taken a pioneering step toward the digital library system, aiming to improve access to information and knowledge in multimodal and open digital scholarly contexts by encouraging a more user-centred, robust, and scalable framework.

5.1 Future Research Directions

Cross-Modal Embedding with transformers – Considering implementing even more sophisticated deep learning type architectures like transformers and large language-vision models to ensure that semantic coherence across modalities while continually closing the semantic gap. Scalable and Distributed Architectures – Building indexing structures and retrieval systems, more capable of supporting real-time access, for large-scale, multimodal, and distributed digital libraries. Adaptive and Intelligent User Interaction – Leveraging reinforcement learning and adaptive recommendation systems into our relevance feedback loop to create individualised personal discovery interfaces that adapt according to user needs.

REFERENCE

[1] Affum, M. Q. (2023). The Role of Artificial Intelligence in Library Automation. *Library Philosophy & Practice*.

[2] Agosti, M., Di Nunzio, G. M., Ferro, N., & Silvello, G. (2019). An innovative approach to data management and curation of experimental data generated through IR test collections. In *Information Retrieval Evaluation in a Changing World: Lessons Learned from 20 Years of CLEF* (pp. 105-122). Cham: Springer International Publishing.

[3] Agosti, M., Di Nunzio, G.M., Ferro, N., et al.: An innovative approach to data management and curation of experimental data generated through IR test collections. In: Ferro, N., Peters, C.(eds.) Information Retrieval Evaluation in a Changing World: Lessons Learned from 20 Years of CLEF. The Information Retrieval Series, pp. 105–122. Springer International Publishing, Cham (2019).

[4] Aliwy, A. H., & Al-Raza, D. A. (2018). Part of speech tagging for Arabic long sentence. *International Journal of Engineering and Technology 7*(3.27).

[5] Al-Smadi, M., Al-Zboon, S., Jararweh, Y., & Juola, P. (2020). Transfer learning for Arabic named entity recognition with deep neural networks. *Ieee Access*, *8*, 37736-37745.

[6] Aravindh, G., & Sridhar, K. P. (2024). Resilient and Adaptive Secure Routing Protocol for Wireless Sensor Networks Using a Grey Wolf Optimizer and Lightning Search Algorithm. *Journal of Internet Services and Information Security*, *14*(4), 331-346. https://doi.org/10.58346/JISIS.2024.I4.020

[7] Bartalesi, V., Pratelli, N., & Lenzi, E. (2022). Linking different scientific digital libraries in Digital Humanities: the IMAGO case study. *International Journal on Digital Libraries*, *23*(4), 303-317.

[8] Brown, T. (2020). Implementing AI-Powered Chatbots in University Libraries: A Case Study of the University of Toronto. *Library Hi Tech*, *38*(2), 287-299.

[9] Çelebi, A.; Özgür, A. Cluster-based mention typing for named entity disambiguation. Natural Language Engineering, 2020, 1–37.

[10] Dutta, S., & Roy, B. (2020). Smart Libraries in Smart Cities: A Roadmap with Artificial Intelligence. *International Journal of Information Science*, *10*(2), 37-45.

[11] Ghosh, M., & Singh, R. (2021). Enhancing Access to Multilingual Digital Resources through AI: The NDLI Experience. *DESIDOC Journal of Library & Information Technology*, *41*(3), 192-198.

[12] Khalifa, M., & Shaalan, K. (2019). Character convolutions for Arabic named entity recognition with long short-term memory networks. *Computer Speech & Language*, *58*, 335-346. https://doi.org/10.1016/j.csl.2019.05.003

[13] Kumar, V., & Jain, A. (2022). AI-Driven Library Services: Opportunities and Limitations in Indian Context. *International Journal of Library and Information Studies*, *12*(1), 65-74.

[14] Liu, L.; Shang, J.; Han, J. Arabic named entity recognition: What works and what's next. In Proceedings of the Fourth Arabic Natural Language Processing Workshop, Florence, Italy, 1–2 August 2019; pp. 60–67.

[15] Mohammad, S., & Azam, A. (2015). A model for ordering in a restaurant based on QR Code without the presence of a waiter at the table. *International Academic Journal of Innovative Research, 2*(2), 8–16.

[16] Nwosu, P. O., & Adeloye, F. C. (2023). Transformation leader strategies for successful digital adaptation. *Global Perspectives in Management*, *1*(1), 1-16.

[17] Quimbaya, A. P., Múnera, A. S., Rivera, R. A. G., Rodríguez, J. C. D., Velandia, O. M. M., Peña, A. A. G., & Labbé, C. (2016). Named entity recognition over electronic health records through a combined dictionary-based approach. *Procedia Computer Science*, *100*, 55-61. https://doi.org/10.1016/j.procs.2016.09.123

[18] Selin, M., & Mathew, K. P. (2024). ResNet152: A Deep Learning Approach for Robust Spoof Detection in Speaker Verification Systems. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications.*, *15*(3), 354-363. https://doi.org/10.58346/JOWUA.2024.I3.023

[19] Sharifzadeh, M. (2015). The Survey of Artificial Neural Networks-Based Intrusion Detection Systems. *International Academic Journal of Science and Engineering, 2*(1), 11–18. https://doi.org/10.1016/j.is.2022.102000

[20] Syed, M. H., & Chung, S. T. (2021). MenuNER: Domain-adapted BERT based NER approach for a domain with limited dataset and its application to food menu domain. *Applied Sciences*, *11*(13), 6007.

[21] Trisovic, A. (2022, October). Cluster analysis of open research data and a case for replication metadata. In *2022 IEEE 18th International Conference on e-Science (e-Science)* (pp. 423-424). IEEE. https://doi.org/10. 1109/eScience55777.2022.00069

[22] Yang, S., & Cai, X. (2022). Bilateral knowledge graph enhanced online course recommendation. *Information Systems*, *107*, 102000. https://doi.org/10.1016/j.is.2022.102000

[23] Yılmaz, N., & Kolukısa Tarhan, A. (2022). Quality evaluation models or frameworks for open source software: A systematic literature review. *Journal of Software: Evolution and Process*, *34*(6), e2458. https://doi.org/10.1002/smr.2458

[24] Zhang, X., Zhao, X., & Liu, X. (2020). Research on knowledge resource recommendation of university mobile library based on context-awareness. *Information Science*, *38*(1), 48-52.