IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 16 January 2025, accepted 25 February 2025, date of publication 27 February 2025, date of current version 13 March 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3546499

RESEARCH ARTICLE

Intelligent Archive Management Based on Deep Learning Technology Driven by Artificial Intelligence

JIAHANG LI[1] AND JIASHU WANG[2]

[1]Party Committee Student Work Department, Changchun Finance College, Changchun 130000, China
[2]School of Management, Changchun University of Chinese Medicine, Changchun 130000, China

Corresponding author: Jiashu Wang (249211660@qq.com)

This work was supported by the Research on Cultivating the Historical Outlook of Young Marxists in Higher Vocational Colleges and Universities, Jilin Province Educational Science Planning Program, under Grant SZ22027.

ABSTRACT To improve the efficiency and intelligence level of an archive management system (AMS) in multimodal data processing, this study proposes and designs an intelligent AMS based on deep learning (DL). However, the traditional AMS faces problems, such as inaccurate data classification, low query efficiency, long response delay, and insufficient system expansibility, when dealing with massive multimodal data. To address these issues, mainstream DL models such as Decoding-Enhanced Bidirectional Encoder Representation from Transformers with Disentangled Attention (DeBERTa), Contrastive Language-Image Pretraining (CLIP), and Swin Transformer are compared with the proposed optimized models. The optimized model's performance improvement in multimodal archive management tasks has been validated through multidimensional experimental assessments. In comprehensive performance comparison experiments, the optimized model demonstrates excellent performance across several key metrics, including resource consumption, response time, data processing throughput, query efficiency, and fault recovery capability. For instance, the optimized model's response time in text processing tasks is 98.367 milliseconds (ms), significantly lower than the Swin Transformer's 156.234 ms. Regarding audio processing tasks, the optimized model's resource consumption is only 4.387 GigaByte (GB), markedly lower than DeBERTa's 6.823 GB. Furthermore, in terms of user satisfaction, the proposed model scores as high as 9.238 in text processing, indicating an enhancement in the user experience. Through effectiveness evaluation experiments, this study further confirms the superiority of the optimized model in terms of accuracy, processing delay, self-learning ability, error rate, security assessment, and system scalability. Moreover, the optimized model achieves an accuracy of 94.23% in text processing, nearly 4% higher than DeBERTa, and reduces the error rate in audio processing to 3.78%, showing greater stability and reliability. Therefore, this study provides new solutions for AMS in the fields of multimodal data processing and intelligent management, especially in enhancing system performance, optimizing user experience, and strengthening system security and scalability.

INDEX TERMS Intelligent archive management, deep learning, multimodal data processing, system optimization, security and scalability.

I. INTRODUCTION

A. RESEARCH BACKGROUND AND MOTIVATIONS

In the wave of the digital era, archive management is gradually transitioning from traditional paper archives to electronic archives. How to efficiently and securely manage and utilize massive amounts of archival information has become a significant challenge for various organizations and institutions. The traditional archive management system (AMS) relies heavily on manual operations, which struggle to cope with the rapid growth of data volume and the complexity of information retrieval. This often leads to issues

The associate editor coordinating the review of this manuscript and approving it for publication was Vlad Diaconita.

such as information redundancy, low retrieval efficiency, and both false positives and negatives in detection, thereby affecting the efficiency and accuracy of archive management [1]. Furthermore, with the increasing demands for privacy protection and data security, AMS is also facing pressures related to data security, privacy breaches, and access control. In this context, advancing Artificial Intelligence (AI) and Deep Learning (DL) technologies provide innovative solutions for intelligent archive management. With their powerful data processing and pattern recognition capabilities, DL technologies can extract useful patterns from vast amounts of unstructured data, thereby enabling automated classification, precise retrieval, and information recommendation, significantly enhancing the intelligence level of archive management [2], [3], [4]. Additionally, these technologies possess robust self-learning and optimization abilities, allowing for dynamic adjustments and upgrades in response to changes in archive data, which provides AMS with greater flexibility and adaptability.

Based on this, this study proposes an intelligent archive management model based on DL technology. It aims to address various bottlenecks and issues in traditional AMS and promote the intelligence and automation of archive management. Simultaneously, the study offers new theoretical foundations and technical support for improving management efficiency, accuracy, and security.

B. RESEARCH OBJECTIVES

(1) By applying models from DL such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), developing an archive classification system capable of efficiently processing multimodal data is achieved. This facilitates the intelligent categorization and organization of archives, enhancing the efficiency and accuracy of classification tasks.

(2) Different DL algorithms and their performance in archive management tasks are compared, and the optimal model and architecture are selected. This ensures that the system can efficiently process massive data while maintaining low resource consumption and response time, thereby improving the system's overall operational efficiency.

(3) By designing reasonable experiments, the performance of a DL-based intelligent AMS on multiple indicators (such as response time, resource consumption, data processing throughput, etc.) is evaluated. Moreover, the system's accuracy, security, and user satisfaction in practical applications are analyzed.

II. LITERATURE REVIEW

Archive management technology has evolved from traditional paper-based systems to electronic and, more recently, intelligent management. The swift progress of information technology in recent years has further transformed archive management models. Jaillant and Rees pointed out that traditional AMS faced the challenge of a rapidly growing number of archives, and the parallel management of both paper and electronic archives increased the complexity of classification and retrieval, posing challenges for meeting modern management needs [5]. Kunduru and Kandepu argued that with the advancement of digitization, electronic archive management gradually became mainstream. However, many existing systems depended heavily on database technology and lacked intelligent functions, resulting in inefficiency when dealing with large-scale data and making it difficult to achieve fast and efficient archive retrieval [6]. As one of the core technologies of AI, DL has been increasingly applied in the archive management field, especially in classification and retrieval. Abdulwahid et al. applied CNNs for the automatic classification of archive images. They found that DL-based classification models demonstrated high accuracy and efficiency when processing large volumes of archive images, markedly enhancing the intelligence of archive classification [7]. Zhao et al. studied the application of DL in Natural Language Processing (NLP). They proposed the use of semantic analysis to automatically classify and retrieve archive texts, addressing the issues of inaccuracy and inefficiency present in traditional keyword-matching methods [8]. As the digital transformation of archive management continues, the challenges surrounding archive data security and privacy have become increasingly prominent. Ensuring data security in intelligent management systems has become a critical research topic in academic and practical fields. Hawkins noted that the main data security risks in archive management stemmed from data breaches and insufficient access control. Moreover, traditional encryption and access control measures were limited in intelligent systems, necessitating the integration of AI technologies to enhance archive security management [9].

Existing research indicates that archive management technology is at a critical stage of intelligent development, and DL has demonstrated significant effectiveness in archive classification, retrieval, and recommendation. At the same time, the widespread use of digital archives has made security and privacy protection key challenges, requiring the exploration of more intelligent solutions. Consequently, this study builds on these foundations to further explore how DL technology drives the smart development of archive management, providing theoretical and technical support for constructing intelligent AMS.

III. RESEARCH METHODOLOGY

A. DL-BASED AUTOMATIC CLASSIFICATION AND RETRIEVAL OF ARCHIVES

Feature extraction is a foundational step in the automatic classification and retrieval system for archives, determining whether the system can effectively understand and process archive data. Archive data often exhibits diversity, including structured data (such as dates and identifiers) and a large amount of unstructured data (such as text, images, and videos). Consequently, different types of data require distinct feature extraction methods. The core task of the archive

TABLE 1. Application of common analysis models.

Model	Analysis
The CNN model	By employing multiple convolutional layers to extract features at different levels of the image, combined with fully connected layers, the model ultimately accomplishes classification.
The bidirectional transformer model	Through pre-training on a large-scale corpus, the model gains the ability to comprehend the contextual semantics of text, making it particularly suitable for classifying unstructured textual archives.
The multimodal classification model	By integrating different types of data features and utilizing a multimodal DL framework, the model fuses data features from various modalities, thereby enhancing the accuracy of archive classification.

TABLE 2. Data security risks in archive management.

Risk	Analysis
Data leakage	Archives store a vast amount of sensitive information, such as personal identity data, corporate financial information, and trade secrets. Unauthorized access to this information could lead to privacy violations, economic losses, or even legal disputes [18-20].
Data tampering	Data tampering refers to the malicious modification or deletion of archive data without authorization, resulting in distortion or unavailability of archive content. This risk can not only undermine the integrity and reliability of the archives but may also negatively impact the decisions and behaviors of archive users [21].
Data loss and disaster recovery	The storage of digital archives relies on physical devices or cloud storage, which may result in data loss due to natural disasters, hardware failures, or network interruptions [22].
System vulnerabilities and malware attacks	AMS may be susceptible to malicious attacks due to software vulnerabilities or misconfigurations. Hackers can exploit these vulnerabilities to launch attacks, implant malware, and encrypt, corrupt, or steal system data.

classification model is to automatically categorize archives into predefined classes based on the extracted features, facilitating subsequent storage, retrieval, and management [10], [11], [12]. DL is widely applied in archive classification, primarily relying on several common models, as detailed in Table 1.

In AMS, text-based archives hold significant importance. Thus, conducting an in-depth analysis of text content is a crucial means to enhance archive management efficiency [13]. Furthermore, before performing content analysis, it is essential to preprocess the text archives, encompassing steps such as tokenization, part-of-speech tagging, stop-word removal, and stemming. These operations help eliminate noise and improve the accuracy of subsequent analyses. In addition, DL-based semantic analysis techniques can transcend traditional keyword matching to achieve a deeper understanding of text archive content. Meanwhile, by applying pre-trained language models, text can be transformed into context-sensitive vector representations, capturing the complex semantic relationships within the text. Thus, this enables the system to identify themes, sentiments, and relevancy in archive content [14], [15], [16].

In summary, the DL-based automatic classification and retrieval system for archives significantly enhances the intelligence and automation of archive management. This is achieved by integrating deep feature extraction, model design, NLP, and intelligent retrieval and recommendation technologies. Hence, it can provide users with more precise and efficient archive management and utilization services.

B. INTELLIGENT SECURITY AND PRIVACY PROTECTION IN ARCHIVE MANAGEMENT

AMS faces various data security risks during its operation, including data leakage, tampering, loss, and unauthorized access. Understanding these risks is essential for designing effective security protection strategies [17]. The specific risks are exhibited in Table 2.

With the development of DL technology, many advanced privacy protection methods have been applied to AMS to enhance the system's security and privacy capabilities [5], [23], [24]. For example, differential privacy is a crucial protection technique that safeguards individual privacy by adding noise to data queries or analysis results. Meanwhile, DL models can incorporate differential privacy techniques to prevent reverse inference of sensitive archival information while maintaining model accuracy. Particularly during the statistical analysis and model training of archive data, differential privacy can effectively prevent external attackers from inferring the original archive content from the model [25].

To sum up, intelligent security and privacy protection in archive management are vital components of the archives' digital and intelligent development [26], [27], [28]. The introduction of DL technology has markedly improved AMS in areas such as data security, privacy protection, identity authentication, access control, and violation detection. These enable it to better address the security challenges of the digital age.

C. DESIGN OF THE INTELLIGENT AMS BASED ON DL

System architecture is the core part of intelligent AMS design, which determines how the system coordinates each module to realize the function of intelligent archive management. The designed AMS architecture based on DL includes an application, data, processing, and interface layers, which cooperate to realize intelligent archive classification, retrieval, storage, and security management. As the number of archives increases and the diversity of data types expands, traditional local storage methods are increasingly inadequate to meet the demands of intelligent management systems. Consequently, integrating archive data storage with cloud

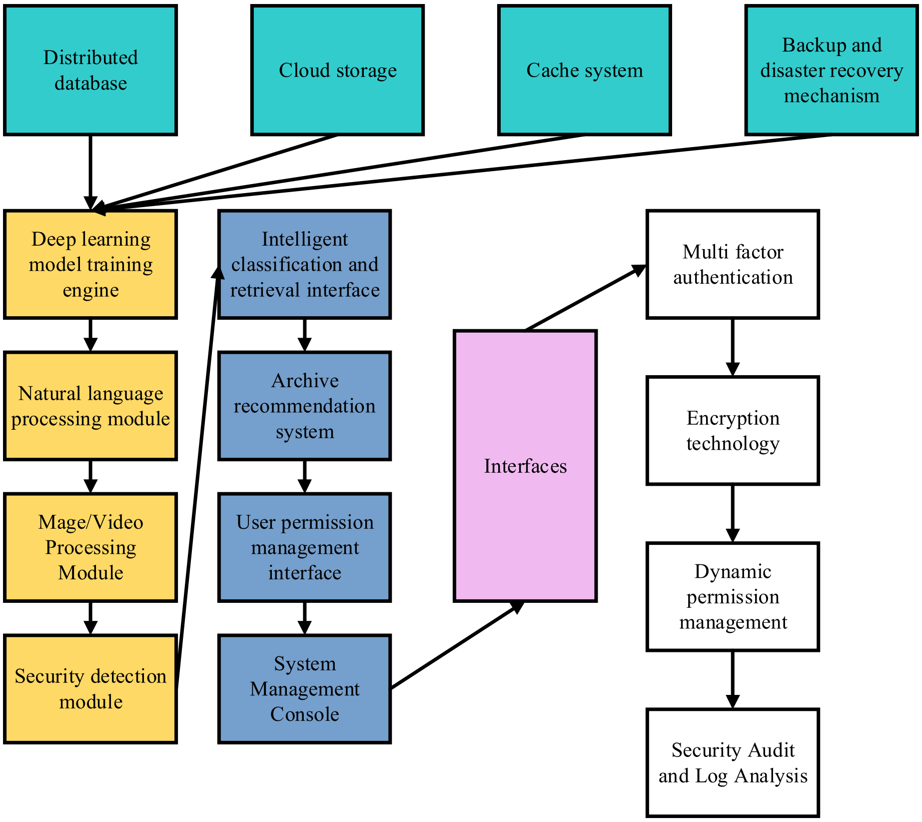

FIGURE 1. **The architecture of the intelligent AMS.**

services has become a trend in smart archive management. Moreover, cloud storage offers high scalability, flexibility, and cost-effectiveness, providing a stable storage environment for massive archive data. In this system, a hybrid cloud architecture is employed for archive data storage, combining the advantages of private and public clouds to achieve flexibility and security in data storage. Sensitive archive data can be stored in the private cloud to ensure data security and privacy protection. In contrast, non-sensitive data is stored in the public cloud to reduce costs and enhance data access efficiency. The system adopts a data tiering strategy to optimize the utilization of storage resources. Archive data is categorized for tiered storage based on attributes such as usage frequency and importance. For instance, frequently accessed archives (hot data) are stored on high-speed storage devices to ensure rapid retrieval, while less frequently accessed archives (cold data) are stored on more cost-effective storage devices to minimize system overhead. The system utilizes distributed storage technology, replicating and distributing archive data across multiple nodes to achieve high availability and fault tolerance. Additionally, a disaster recovery backup mechanism is integrated into the system, which regularly backs up archive data to off-site storage centers to ensure rapid data recovery in the event of a failure or disaster. Thus, it can safeguard the continuous availability and integrity of the archives. The intelligent AMS is also integrated with multiple cloud service platforms, such as Amazon Web Services and Microsoft Azure. Overall, the data analysis, machine learning (ML), and content management services provided by these cloud platforms can further enhance the system's level of intelligence. For example, the system can automatically analyze, classify, and recommend archive data by utilizing the cloud platform's ML application programming interface.

The specific architecture is displayed in Figure 1:

This study optimizes the model architecture, which is divided into five key layers: the data layer, processing layer, application layer, interface layer, and security layer. Each layer focuses on the storage, intelligent functionality, and security design of archival data, integrating advanced computational technologies to support the processing of large-scale multimodal data. The data layer, as the system's foundation, is responsible for storing and managing archival data. To accommodate various types and scales of archives, the data layer integrates distributed databases and cloud storage technologies to handle structured and unstructured data (such as documents, images, and audio). A caching mechanism stores frequently accessed data, reducing database query latency. Additionally, backup and disaster recovery systems are introduced to ensure data security and system stability. The processing layer is the system's core, focusing on the development of intelligent functionalities based on DL technologies, including archival classification, retrieval, recommendation, and security detection. DL models extract features from text and images, and multimodal data classification is achieved through weighted fusion. The semantic retrieval module implements accurate search via semantic matching algorithms, while the recommendation feature combines collaborative filtering and DL algorithms to offer personalized archival recommendations. The security detection module utilizes DL to analyze user access logs in real-time, automatically identifying and blocking potential threats. The application layer is directly user-facing, providing the main functionalities required for archive management. Users can quickly locate target archives through classification and retrieval functions, and manage user roles and access control via the permission management module. Furthermore, the system's administrative console enables administrators to monitor system status, operation logs, and data management, significantly improving operational efficiency. The interface layer facilitates integration with external systems and services, offering standardized interfaces via RESTful API and GraphQL API. It supports data exchange with other business systems and cross-platform, multi-device access. The interface layer also provides integration with third-party cloud services, enhancing the system's compatibility and flexibility. The security layer is embedded throughout the entire system architecture, providing comprehensive security protection for each layer. Multi-factor authentication ensures the uniqueness of user identities, and advanced encryption algorithms are used to protect sensitive information during data transmission and storage. Additionally, the intelligent audit module, combined with DL-based log analysis tools, detects anomalous operations and potential threats in real-time, thus enhancing the system's security and reliability.

The overall architecture adopts a distributed design with cloud storage technology, supporting parallel processing and dynamic scaling of large-scale data. It ensures fast response capabilities under high concurrency scenarios through efficient caching and indexing mechanisms. By incorporating DL-optimized retrieval algorithms, the system significantly improves retrieval efficiency and accuracy in large-scale data scenarios. The multi-layered security protection mechanisms ensure data privacy and system security. The design of this optimized model not only achieves efficient multimodal data processing but also reaches new heights in performance, intelligence, and security, offering strong technical support for intelligent archive management.

IV. EXPERIMENTAL DESIGN AND PERFORMANCE EVALUATION

A. DATASETS COLLECTION, EXPERIMENTAL ENVIRONMENT, AND PARAMETERS SETTING

The experimental dataset, ArchivalQA, is specifically designed for open-domain question answering and focuses on archive management tasks related to historical news. It comprises approximately 1.8 million news articles spanning from 1987 to 2007, with a significantly larger dataset size compared to other similar datasets. In addition, the dataset primarily consists of textual news articles, especially for document understanding and retrieval tasks in archive management. The dataset can be downloaded from the official website (https://www.openml.org/). The ArchiveQA dataset's selection is based on its scale, content diversity, task design, and high adaptability to archive management tasks. This dataset provides a real and complex testing environment for the intelligent AMS in this study. It validates the model's multimodal processing capabilities and tests its performance with historical data spanning a long period. Compared to other datasets, ArchiveQA more comprehensively meets the research requirements, offering higher practical value and academic significance.

The experimental environment is as follows. The graphics processor model is NVIDIA Tesla V100, the memory capacity is 256GB DDR4, the storage device model is Samsung PM1733 NVMe SSD, and the motherboard model is Supermicro X11 series server motherboard. Meanwhile, the network device model is Mellanox ConnectX-5 100GbE network card, the operating system version is Ubuntu 20.04 LTS 64-bit, and the Python version is Python 3.8.10. The comparative models selected for the experiment are Contrastive Language-Image Pretraining (CLIP), Decoding-Enhanced Bidirectional Encoder Representation from Transformers with Disentangled Attention (DeBERTa), and Swin Transformer. CLIP is a cross-modal contrastive learning model that maps text and images into the same vector space, enabling joint analysis of multimodal data. This study involves multimodal data such as text, images, and audio, and CLIP's cross-modal processing ability provides an important reference for comparison. CLIP performs excellently in multimodal tasks such as image classification and text-image retrieval, demonstrating advanced processing capabilities and broad practical applications. A comparison with CLIP can validate the improvements made by the optimized model in multimodal data processing. DeBERTa is an optimized model based on Bidirectional Encoder Representations from Transformers (BERT) that excels in NLP tasks such as text classification and question answering. This study involves a significant amount of textual archive data, and DeBERTa's high semantic understanding capacity serves as a vital baseline for comparison. DeBERTa's decoupled attention mechanism more effectively captures contextual information in text, providing a powerful baseline for text classification and semantic retrieval tasks. Swin Transformer is a vision processing model based on a windowing mechanism that has achieved outstanding results in image classification and object detection tasks. The image archive data in this study requires strong visual feature extraction capabilities, and Swin Transformer provides a reliable comparative model. Compared to traditional Transformer models, Swin Transformer's shifted window mechanism markedly reduces computational complexity. A comparison with Swin Transformer can validate the efficiency improvements made by the optimized model. The selection of CLIP, DeBERTa, and Swin Transformer as comparative models is based on their outstanding performance in different data types (text, images, and multimodal data). They represent the highest level of mainstream DL models in text processing, image processing, and multimodal processing fields. By comparing with these models, this study can comprehensively verify the advantages of the optimized model in multimodal AMS, while enhancing the academic value and practical significance of the research findings.

The model parameters in this study are carefully chosen to meet the requirements of the experimental task and the limitations of the hardware environment, ensuring the experimental results' validity and reliability. Firstly, the learning rate is set to 0.01, a commonly used starting value in DL models, which ensures the model converges quickly while avoiding issues like gradient explosion or oscillation. Combined with the optimizer and learning rate scheduling strategy, this setting balances convergence efficiency with training stability. The batch size is 16, increasing the training speed within the limits of GPU memory, while effectively balancing computational efficiency and gradient update stability, which is suitable for handling multimodal data. The maximum sequence length for text input is set to 128, which is sufficient to cover most archive texts while avoiding the increased computational cost associated with longer sequences. The number of training epochs is 3, providing sufficient learning for most small- and medium-scale datasets while avoiding overfitting or lengthy training processes. For the image modality, the input size is chosen to be 224×224, which is the standard size for many mainstream visual models such as CLIP and Swin Transformer. This setting strikes a balance between detail preservation and computational efficiency, suitable for the handling of image data in archive management. In the contrastive learning task, the temperature coefficient is set to 0.07, a value that controls the distinction between positive and negative samples. It is one of the empirically proven optimal values in contrastive learning, effectively enhancing the feature separation ability of the samples. Finally, the window size in the Swin Transformer model is 7, which is its recommended default value. It can achieve a good balance between local feature extraction and global feature fusion, especially suited for the multimodal archive management tasks in this study.

In summary, these parameter settings are not only based on practical experience in the DL field but also tailored to the characteristics of the research task. While ensuring model performance, these parameters effectively control the usage of computational resources, enhancing the scientific rigor and reproducibility of the experiments.

B. PERFORMANCE EVALUATION

1) PERFORMANCE COMPARISON TEST

The study first conducts a performance comparison experiment, evaluating metrics such as response time, resource consumption, query efficiency, data processing throughput, fault recovery capability, and user satisfaction. Furthermore, the dataset is divided into three dimensions: text, image, and audio data, with 10,000 entries for each dimension. The results of the performance comparison experiment are suggested in Figure 2:

Figure 2a reveals that regarding response time, the model proposed here performs best, with response times of 98.367 milliseconds (ms) for text, 112.945 ms for images, and 145.589 ms for audio, significantly lower than other models. In contrast, the Swin Transformer has the highest response time when processing audio, reaching 265.732 ms, while the CLIP model also exhibits a longer response time for image processing. In Figure 2b, the optimized model demonstrates significant advantages across all data types, consuming 2.674 GigaByte (GB) (text), 3.789 GB (image), and 4.387 GB (audio). This is more resource-efficient compared to the high resource usage of DeBERTa and CLIP. In Figure 2c, the optimized model maintains its lead, with query times of 98.654, 113.345, and 145.984 ms for text, images, and audio, showing a clear time advantage over DeBERTa and Swin Transformer. In Figure 2d, the optimized model processes text, image, and audio at speeds of 2.238 GB/s, 1.889 GB/s, and 1.356 GB/s, respectively, surpassing the other models, especially in image and audio processing. In Figure 2e, the optimized model has the shortest recovery times, with 5.768, 3.678, and 4.783 seconds for audio, text, and images, remarkably outperforming Swin Transformer and DeBERTa. In Figure 2f, the proposed model receives the highest scores across all data types, with user satisfaction ratings of 8.764, 9.238, and 8.467 for images, text, and audio, indicating a notable improvement in user experience compared to other models.

2) EFFECTIVENESS EVALUATION EXPERIMENT

An effectiveness evaluation experiment is conducted to verify the effectiveness of the proposed optimized model. It compares metrics such as accuracy, processing delay, self-learning ability, error rate, security assessment, and system scalability. The results are depicted in Figure 3:

First, Figure 3a shows that in terms of accuracy, the proposed optimized model reaches accuracies of 94.23% for text, 91.89% for images, and 89.54% for audio, outperforming the other models. CLIP also performs well in image classification tasks, with an accuracy of 88.56%, but it lags behind the optimized model in text and audio processing. Second, in Figure 3b, the proposed model exhibits the lowest delay, with 62.34, 53.89, and 49.78 ms for audio, text, and image processing, apparently lower than that of the Swin Transformer, especially in audio processing. Third, in Figure 3c, the optimized model exhibits higher abilities across all data types, with scores of 8.56, 8.34, and 7.89 for text, image, and audio processing, respectively, exceeding DeBERTa and Swin Transformer. Then, Figure 3d presents that the optimized model has lower error rates of 3.12%,1.76%, and 3.78% for images, text, and audio processing, illustrating higher accuracy and reliability. Next, in Figure 3e, the proposed model achieves higher security scores, with 8.98, 9.01,

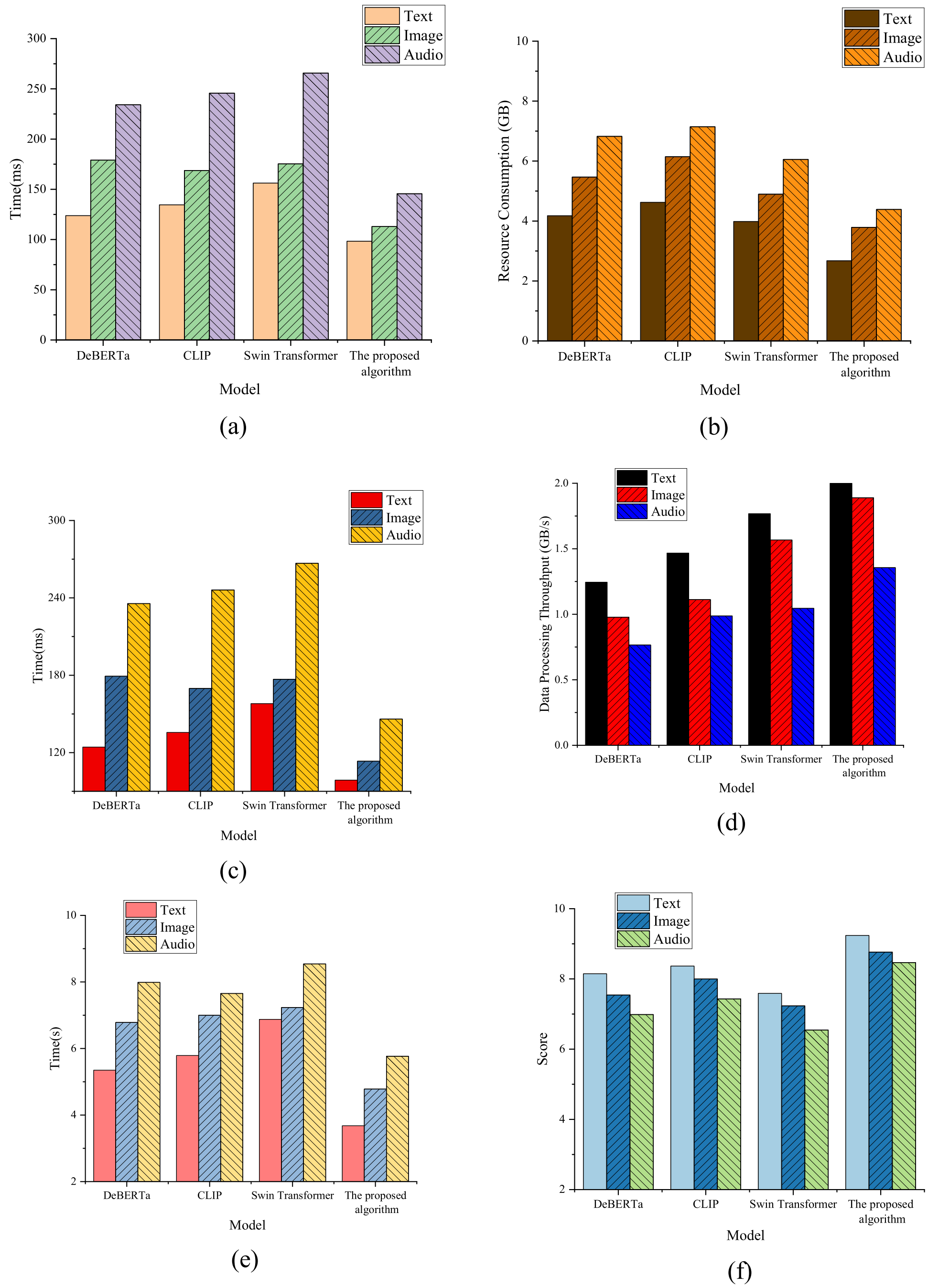

FIGURE 2. Performance comparison results ((a): Response Time; (b): Resource Consumption; (c): Query Efficiency; (d): Data Processing Throughput; (e): Fault Recovery Time; (f): User Satisfaction).

and 9.12, for audio, images, and text, significantly outperforming the Swin Transformer, especially in audio processing tasks. Lastly, in Figure 3f, the optimized model attains higher scalability scores across all data types, with scores of 8.34, 8.78, and 8.12 for images, text, and audio, demonstrating superior scalability compared to CLIP and DeBERTa.

C. DISCUSSION

The performance comparison experiment demonstrates that the proposed optimized model shows significant advantages across all key metrics. Firstly, the optimized model achieves a remarkable reduction in response time, particularly when handling large-scale image and audio data, highlighting its great potential for efficiently processing multimodal data. Secondly, in terms of resource consumption, the optimized model markedly reduces memory and computational resource usage through more efficient algorithms and hardware resource management. This means that under the same hardware conditions, the proposed model can process more data, thereby enhancing the system's processing capacity.

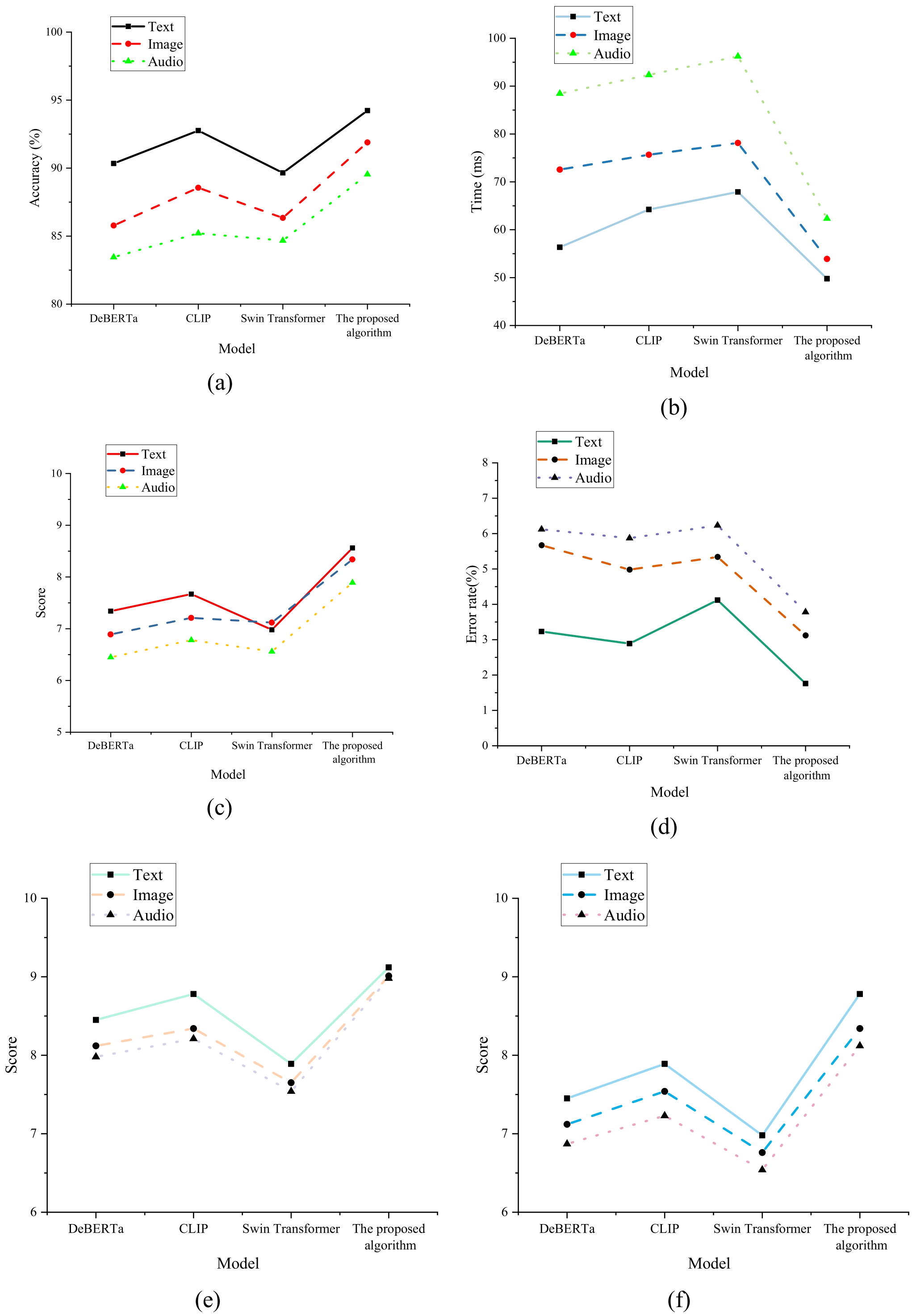

FIGURE 3. Effectiveness evaluation experiment ((a): Accuracy; (b): Processing Delay; (c): Self-Learning Ability; (d): Error Rate; (e): Security Assessment; (f): System Scalability).

The results for data processing throughput also highlight the optimized model's superior performance, consistently outperforming other models across various data types. Combined with its shorter fault recovery time, the optimized model ensures system stability and reliability even under high-load conditions. Finally, the user satisfaction survey further validates the optimized model's advantages in practical applications, particularly the improvements in audio and image processing, which greatly enhance the user experience. These findings demonstrate that the optimized model excels in performance and provides a better user experience.

The effectiveness evaluation experiment reveals that the proposed optimized model performs exceptionally well across multiple evaluation metrics. In terms of accuracy, the optimized model markedly outperforms the comparison models in processing text, images, and audio, particularly excelling in handling complex data. Furthermore, regarding processing latency, the optimized model's efficient algorithms notably reduce data processing delays, improving system real-time performance and response speed. Simultaneously, in evaluating self-learning capability, the optimized model exhibits stronger self-optimization abilities due to its improved adaptive algorithms, which is crucial for dynamic AMS. Additionally, the reduction in error rates further validates the proposed model's reliability, decreasing the probability of misjudgment in multimodal data processing. The security and scalability evaluation results suggest that the optimized model is more reliable in data protection. Also, it exhibits better scalability when dealing with large-scale data and high-concurrency access, meeting the growing demands of the system. Overall, the performance and security of the proposed optimized model provide robust support for the application of intelligent AMS.

This study adopts a DL approach for multimodal data processing, comparing multiple models and proposing an optimized model. The choice of this methodology is based on the practical needs of the intelligent AMS, especially the requirements for efficiently handling text, image, and audio data. Through experimental design and performance evaluation, it has been demonstrated that the selected approach adapts well to the complexity of multimodal data while meeting the multiple objectives of system performance, resource management, and security. Therefore, the proposed method is highly applicable in the current field of archive management. Furthermore, this study's main innovation lies in the optimization and integration of existing DL models. By introducing adaptive learning mechanisms and efficient model architecture design, the optimized model significantly improves the system's accuracy, throughput, and scalability. Meanwhile, this study innovatively applies DL privacy protection technologies (such as encryption and anomaly detection) to archive management, greatly enhancing the system's security and reliability. Furthermore, through exploring the distributed processing of multimodal data and dynamic optimization mechanisms, the proposed methods offer new solutions for intelligent AMS.

Existing literature, such as CLIP, focuses on cross-modal processing of text and images, but its support for other modalities, such as audio, is limited. DeBERTa highlights text processing and lacks the ability to handle multimodal data. Swin Transformer is primarily used for visual tasks and struggles to scale to non-visual modalities. The optimized model proposed here integrates multiple DL technologies to support the processing of multimodal data, such as images, text, and audio, notably enhancing the system's capabilities in multimodal scenarios. At the same time, the model achieves feature fusion across different modalities, enabling it to capture the potential correlations in multimodal data more comprehensively. Mainstream models, such as CLIP and Swin Transformer, typically require significant computational resources, especially when handling high-dimensional images or multimodal data, making them heavily dependent on hardware. While DeBERTa performs excellently in text processing, its computational complexity may lead to excessive resource consumption in large-scale semantic tasks. This study reduces resource consumption and response latency by introducing adaptive learning mechanisms and optimized model architectures. Experiments show that, under the same hardware conditions, the optimized model uses fewer resources and achieves efficient data processing.

Raciti et al. made groundbreaking advances in multimodal learning tasks, especially cross-modal text-image processing, by proposing the CLIP model [29]. CLIP can map text and images to the same vector space, significantly improving the efficiency of multimodal retrieval and classification. However, their study primarily focuses on the fusion of text and image modalities, with limited support for other modalities, such as audio. Moreover, CLIP requires substantial computational resources for high-precision multimodal tasks, and its system scalability and adaptability are relatively insufficient. The proposed optimized model supports text and image processing and markedly enhances support for audio and other non-visual modal data. Through cross-modal feature extraction and fusion technologies, it achieves more comprehensive archival data management. Rasmussen et al. introduced the Swin Transformer model, which excelled in computer vision tasks. Its shift-window mechanism efficiently captured both local and global features of images, making it widely used for image classification and object detection tasks [30]. However, Swin Transformer primarily concentrates on image modalities and struggles to handle multimodal data like text and audio. Additionally, the model's response time and system scalability in large-scale data scenarios still have room for improvement. While the Swin Transformer has clear advantages in image modality processing, it lacks support for text and audio modalities. The optimized model in this study offers a comprehensive optimization for text, image, and audio modalities, enabling efficient handling of various modal data and enhancing the system's adaptability in complex archive management scenarios.

V. CONCLUSION

A. RESEARCH CONTRIBUTION

The contributions of this study are mainly reflected in the following aspects. First, addressing the issues of low data classification, retrieval, and processing efficiency in traditional AMS, this study proposes and implements an intelligent AMS based on DL technologies. This system, through the integration of multiple DL models, efficiently handles multimodal archival data (text, images, audio, etc.), remarkably enhancing the automation and intelligence of archive management. Second, this study proposes an optimized model based on existing DL models, achieving significant improvements in resource consumption, response time, and query efficiency. By introducing adaptive learning mechanisms and efficient model architecture optimization, the accuracy of archival data processing has significantly improved. Moreover, the error rate has been greatly reduced, providing strong support for enhancing the system's efficiency and reliability.

Furthermore, this study addresses security issues in AMS by designing a DL-based intelligent security monitoring and privacy protection mechanism. By improving data encryption, access control, and anomaly detection techniques, the system shows greater adaptability in security and privacy protection, effectively ensuring the safety of archival data during the management process. Finally, this study expands the application of DL in the archive management field. Also, it provides new solutions for intelligent management, model optimization, system security, and scalability, with significant theoretical and practical implications.

In conclusion, these contributions significantly enrich the application scenarios of DL technologies in archive management, providing innovative support for intelligent management, performance optimization, and security assurance. This offers important references for research and practice in the field.

B. FUTURE WORKS AND RESEARCH LIMITATIONS

Although the proposed optimized model performs exceptionally well across multiple experiments, there is still room for improving computational efficiency as archive data grows. Future research can be conducted from the following aspects. 1. Future work focuses on developing more lightweight DL models to reduce computational resource requirements and explore more efficient distributed training methods for large-scale datasets. Additionally, introducing multimodal pre-trained models could further enhance multimodal data processing. While the current system can process text, image, and audio data, handling more complex archive data (e.g., video, and sensor data) remains a challenge. 2. Future work concentrates on improving multimodal data fusion techniques, incorporating methods from NLP, computer vision, and temporal analysis to enhance system performance on complex data. Concurrently it can explore automatically extracting deep semantic information from multimodal data. Moreover, although this study designs an intelligent security and privacy protection mechanism based on DL, AMS may face increasingly complicated security threats due to the diversification of cyberattacks. 3. Future work prioritizes researching more advanced privacy protection technologies, such as federated learning and differential privacy, to address more severe data security challenges. Additionally, dynamic access control and adaptive security strategies within archival systems can be explored to ensure continued system security in ever-changing environments.

REFERENCES

[1] M. Ngoepe, L. Jacobs, and M. Mojapelo, "Inclusion of digital records in the archives and records management curricula in a comprehensive open distance e-learning environment," *Inf. Develop.*, vol. 40, no. 2, pp. 190–201, Jun. 2024.

[2] H. Benmakhlouf and A. Chouaou, "Electronic document, information, and archive management systems in economic institutions: A descriptive study of the onbase system," *Int. J. Prof. Bus. Rev.*, vol. 9, no. 6, p. 11, Jul. 2024.

[3] L. Jaillant, "How can we make born-digital and digitised archives more accessible? Identifying obstacles and solutions," *Archival Sci.*, vol. 22, no. 3, pp. 417–436, Sep. 2022.

[4] L. Jaillant and A. Caputo, "Unlocking digital archives: Cross-disciplinary perspectives on AI and born-digital data," *AI Soc.*, vol. 37, no. 3, pp. 823–835, Sep. 2022.

[5] L. Jaillant and A. Rees, "Applying AI to digital archives: Trust, collaboration and shared professional ethics," *Digit. Scholarship Humanities*, vol. 38, no. 2, pp. 571–585, May 2023.

[6] A. R. Kunduru and R. Kandepu, "Data archival methodology in enterprise resource planning applications (Oracle ERP, peoplesoft)," *J. Adv. Math. Comput. Sci.*, vol. 38, no. 9, pp. 115–127, Aug. 2023.

[7] A. H. Abdulwahid, M. Pattnaik, M. R. Palav, S. B. G. T. Babu, G. Manoharan, and G. P. Selvi, "Library management system using artificial intelligence," *IEEE*, vol. 5, no. 3, pp. 1–7, Jan. 2023.

[8] H. Zhao, S. Li, H. Xu, L. Ye, and M. Chen, "The influence of educational psychology on modern art design entrepreneurship education in colleges," *Frontiers Psychol.*, vol. 13, Jun. 2022, Art. no. 843484, doi: 10.3389/fpsyg.2022.843484.

[9] A. Hawkins, "Archives, linked data and the digital humanities: Increasing access to digitised and born-digital archives via the semantic Web," *Archival Sci.*, vol. 22, no. 3, pp. 319–344, Sep. 2022.

[10] M. Chen and W. Du, "Dynamic relationship network and international management of enterprise supply chain by particle swarm optimization algorithm under deep learning," *Expert Syst.*, vol. 41, no. 5, May 2024, Art. no. e13081, doi: 10.1111/exsy.13081.

[11] E. Daga, L. Asprino, R. Damiano, M. Daquino, B. D. Agudo, A. Gangemi, T. Kuflik, A. Lieto, M. Maguire, A. M. Marras, D. M. Pandiani, P. Mulholland, S. Peroni, S. Pescarin, and A. Wecker, "Integrating citizen experiences in cultural heritage archives: Requirements, state of the art, and challenges," *J. Comput. Cultural Heritage*, vol. 15, no. 1, pp. 1–35, Mar. 2022.

[12] A. Näslund, "Image metadata. From information management to interpretative practice," *Museum Manage. Curatorship*, vol. 39, no. 4, pp. 398–418, Jul. 2024.

[13] R. Sacks, Z. Wang, B. Ouyang, D. Utkucu, and S. Chen, "Toward artificially intelligent cloud-based building information modelling for collaborative multidisciplinary design," *Adv. Eng. Informat.*, vol. 53, Aug. 2022, Art. no. 101711.

[14] Y. Liu and M. Chen, "The knowledge structure and development trend in artificial intelligence based on latent feature topic model," *IEEE Trans. Eng. Manag.*, vol. 71, pp. 12593–12604, 2023, doi: 10.1109/TEM.2022.3232178.

[15] B. Wachnik, "Analysis of the use of artificial intelligence in the management of industry 4.0 projects. The perspective of Polish industry," *Prod. Eng. Arch.*, vol. 28, no. 1, pp. 56–63, Mar. 2022.

[16] G. Ramesh, J. Logeshwaran, and V. Aravindarajan, "The performance evolution of antivirus security systems in ultra dense cloud server using intelligent deep learning," *BOHR Int. J. Comput. Intell. Commun. Netw.*, vol. 1, no. 1, pp. 15–19, May 2022.

[17] J. Yang, F. Xiang, R. Li, L. Zhang, X. Yang, S. Jiang, H. Zhang, D. Wang, and X. Liu, "Intelligent bridge management via big data knowledge engineering," *Autom. Construction*, vol. 135, Mar. 2022, Art. no. 104118.

[18] Ö. Albayrak Ünal, B. Erkayman, and B. Usanmaz, "Applications of artificial intelligence in inventory management: A systematic review of the literature," *Arch. Comput. Methods Eng.*, vol. 30, no. 4, pp. 2605–2625, Apr. 2023.

[19] M. Aboian, K. Bousabarah, E. Kazarian, T. Zeevi, W. Holler, S. Merkaj, G. Cassinelli Petersen, R. Bahar, H. Subramanian, P. Sunku, E. Schrickel, J. Bhawnani, M. Zawalich, A. Mahajan, A. Malhotra, S. Payabvash, I. Tocino, M. Lin, and M. Westerhoff, "Clinical implementation of artificial intelligence in neuroradiology with development of a novel workflow-efficient picture archiving and communication system-based automated brain tumor segmentation and radiomic feature extraction," *Frontiers Neurosci.*, vol. 16, Oct. 2022, Art. no. 860208.

[20] J. R. Lechien, A. Maniaci, I. Gengler, S. Hans, C. M. Chiesa-Estomba, and L. A. Vaira, "Validity and reliability of an instrument evaluating the performance of intelligent chatbot: The artificial intelligence performance instrument (AIPI)," *Eur. Arch. Oto-Rhino-Laryngol.*, vol. 281, no. 4, pp. 2063–2079, Apr. 2024.

[21] Y. Huang, H. Peng, M. Sofi, Z. Zhou, T. Xing, G. Ma, and A. Zhong, "The city management based on smart information system using digital technologies in China," *IET Smart Cities*, vol. 4, no. 3, pp. 160–174, Sep. 2022.

[22] S. Rapinel, L. Panhelleux, G. Gayet, R. Vanacker, B. Lemercier, B. Laroche, F. Chambaud, A. Guelmami, and L. Hubert-Moy, "National wetland mapping using remote-sensing-derived environmental variables, archive field data, and artificial intelligence," *Heliyon*, vol. 9, no. 2, Feb. 2023, Art. no. e13482.

[23] A. L. Cushing and G. Osti, "'So how do we balance all of these needs?': How the concept of AI technology impacts digital archival expertise," *J. Document.*, vol. 79, no. 7, pp. 12–29, Apr. 2023.

[24] T. Ye, H. Wang, T. Zeng, M. G. H. Omran, F. Wang, Z. Cui, and J. Zhao, "An improved two-archive artificial bee colony algorithm for many-objective optimization," *Expert Syst. Appl.*, vol. 236, Feb. 2024, Art. no. 121281.

[25] V. Croce, G. Caroti, A. Piemonte, L. De Luca, and P. Véron, "H-BIM and artificial intelligence: Classification of architectural heritage for semi-automatic scan-to-BIM reconstruction," *Sensors*, vol. 23, no. 5, p. 2497, Feb. 2023.

[26] A. V. Singh, M. Varma, P. Laux, S. Choudhary, A. K. Datusalia, N. Gupta, A. Luch, A. Gandhi, P. Kulkarni, and B. Nath, "Artificial intelligence and machine learning disciplines with the potential to improve the nanotoxicology and nanomedicine fields: A comprehensive review," *Arch. Toxicol.*, vol. 97, no. 4, pp. 963–979, Apr. 2023.

[27] A. Iudin, P. K. Korir, S. Somasundharam, S. Weyand, C. Cattavitello, N. Fonseca, O. Salih, G. J. Kleywegt, and A. Patwardhan, "EMPIAR: The electron microscopy public image archive," *Nucleic Acids Res.*, vol. 51, no. D1, pp. D1503–D1511, Jan. 2023.

[28] E. Clough, T. Barrett, S. E. Wilhite, P. Ledoux, C. Evangelista, I. F. Kim, M. Tomashevsky, K. A. Marshall, K. H. Phillippy, P. M. Sherman, H. Lee, N. Zhang, N. Serova, L. Wagner, V. Zalunin, A. Kochergin, and A. Soboleva, "NCBI GEO: Archive for gene expression and epigenomics data sets: 23-year update," *Nucleic Acids Res.*, vol. 52, no. 1, pp. 138–144, Jan. 2024.

[29] P. Raciti, J. Sue, J. A. Retamero, R. Ceballos, R. Godrich, J. D. Kunz, A. Casson, D. Thiagarajan, Z. Ebrahimzadeh, J. Viret, D. Lee, P. J. Schüffler, G. DeMuth, E. Gulturk, C. Kanan, B. Rothrock, J. Reis-Filho, D. S. Klimstra, V. Reuter, and T. J. Fuchs, "Clinical validation of artificial intelligence–augmented pathology diagnosis demonstrates significant gains in diagnostic accuracy in prostate cancer detection," *Arch. Pathol. Lab. Med.*, vol. 147, no. 10, pp. 1178–1185, Oct. 2023.

[30] M. L. R. Rasmussen, A.-C. Larsen, Y. Subhi, and I. Potapenko, "Artificial intelligence-based ChatGPT chatbot responses for patient and parent questions on vernal keratoconjunctivitis," *Graefe's Arch. Clin. Experim. Ophthalmol.*, vol. 261, no. 10, pp. 3041–3043, Oct. 2023.

JIAHANG LI was born in Changchun. He received the LL.M. degree in ideological and political education from Changchun University of Technology. He is currently a Librarian. He is engaged in archives management, education, and historical documentation research.

JIASHU WANG was born in Changchun. She received the LL.M. degree in sociology from Changchun University of Technology. She is currently a Lecturer in the ideological and political education of college students.