[1]Jianping Hu
[2]Yongkang Yan
[3]Zhengguang Xie

Application of Artificial Intelligence-based Technology in College Archives Management

Journal of Electrical Systems

Abstract: - This study looks into the use of artificial intelligence (AI)-based technologies, notably Convolutional Neural Network (CNN) models and data augmentation approaches, in college archives management. With the rapid expansion of digital information, archive organizations have enormous issues in maintaining, organizing, and giving access to their vast collections of historical documents and scholarly resources. AI solutions present possible solutions to these difficulties by automating labour-intensive operations and improving the efficiency and accuracy of preservation workflows. In this study, they investigate the transformational potential of CNN models in digitization activities, utilizing their visual data processing skills to automate and enhance the digitization of visual resources within college archives. In addition, they look into how data augmentation techniques might improve the resilience and generalization capabilities of AI models, reducing biases and improving performance across a variety of archive collections. Using empirical research, theoretical analysis, and case studies, they present insights into the opportunities and obstacles of implementing AI-based technology in college archive management. Furthermore, they discuss ethical implications for the deployment of AI solutions, emphasizing the significance of transparency, accountability, and stakeholder participation. This study aims to inform archival practitioners, technologists, and stakeholders about the effective and responsible integration of AI technologies in promotion within academic institutions by elucidating the practical applications and implications of CNN models and data augmentation techniques in archival contexts.

Keywords: Convolutional Neural Network (CNN), Artificial Intelligence (AI), Data Augmentation technique, Archives Management.

I. INTRODUCTION

In the fast changing environment of archival administration, the use of Artificial Intelligence (AI) represents a paradigm shift in how college archives protect, organize, and access their vast collections of historical documents and intellectual resources. This study investigates the revolutionary potential of AI-based technology, with a specific focus on the use of Convolutional Neural Network (CNN) models and data augmentation approaches in college archives administration [1]. College archives serve as caretakers of precious institutional memory, containing a rich tapestry of documents, photographs, manuscripts, and multimedia artifacts that trace academic institutions' history and intellectual endeavors [2]. However, the sheer number and complexity of historical documents provide substantial problems to standard archival management strategies. Manual digitization, metadata tagging, and categorization are labor-intensive, time-consuming, and prone to human mistakes, limiting the efficiency and scalability of archiving procedures [3].

The introduction of AI-based technology provides a disruptive solution to these difficulties. Archival institutions may automate and enhance the digitization process for visual assets such as images, maps, and artworks by using the capability of CNN models, which excel at visual data processing tasks. CNNs excel in feature extraction and pattern recognition, allowing them to accurately identify and categorize archival items, accelerating digitalization while preserving a visual legacy [4]. Furthermore, this study investigates the use of data augmentation strategies to improve the robustness and generalizability of AI models in college archive management [5]. Data augmentation is creating synthetic versions of existing archival resources using changes like rotation, scaling, cropping, and color manipulation [6]. By supplementing the training dataset with diverse and representative samples, archive institutions can reduce overfitting, improve model performance, and assure AI solutions' resilience to differences in archival materials and conditions [7].

However, the use of AI-based technology in college archive management raises concerns about ethics, bias, and sensitivity [8]. As AI algorithms are educated on historical data, they may unintentionally perpetuate biases seen in archival sources, resulting in distorted representations or misinterpretations of narratives. Furthermore, concerns of privacy, permission, and intellectual property rights must be carefully negotiated to maintain ethical standards and protect the interests of archival stakeholders [9]. In light of these complexities, this study aims to give a

[1] Nantong University Library, Nantong, Jiangsu, 226019, China, hujp@ntu.edu.cn
[2] Nantong University Library, Nantong, Jiangsu, 226019, China, yyk@ntu.edu.cn
[3] *Corresponding author: School of Information Science and Technology, Nantong University, Nantong, Jiangsu, 226019, China, xiezg@ntu.edu.cn

thorough evaluation of the opportunities and constraints associated with the use of AI-based technology, notably CNN models and data augmentation approaches, in college archives management. This study intends to contribute to the continuing discussion about the convergence of technology by illuminating AI's transformative potential in improving archive practices, protecting institutional history, and encouraging scholarly inquiry [10].

II. RELATED WORK

P. Yan and S. Zhang [11]. Studies on the digitization of archive records have investigated the efficacy of AI-based techniques, such as Convolutional Neural Network (CNN) models, in automating and optimizing the digitization process. Researchers demonstrated the use of CNNs for automatic document detection and segmentation in archive collections, demonstrating the potential to speed up digitization efforts while preserving accuracy and quality requirements.

Furthermore, M. Rezapour and S. K. Elmshaeuser et al [12]. Research has focused on the use of data augmentation approaches to improve the performance and generalizability of AI models in archival situations. Researchers aimed to reduce biases, improve model resilience, and assure adaptation to varied archive collections by augmenting training datasets with synthetic versions of historical materials. Researchers used data augmentation techniques like rotation scaling and noise injection to enhance the accuracy of handwritten text recognition in digitized archival documents.

Y. Ma et al [13]. Ethical implications have also been highlighted in the literature, with scholars addressing issues such as bias, privacy, and when deploying AI solutions in archive environments. Researchers investigated the ethical implications of AI-based metadata production in archive collections, emphasizing the importance of openness, accountability, and stakeholder participation in AI decision-making processes.

A. Shelar et al [14]. In addition to the aforementioned studies, a broader body of related work contributes to the understanding of AI in archival administration. Natural Language Processing (NLP) approaches have been used to extract information and analyze text in archive collections. They looked into the use of Named Entity Recognition (NER) algorithms to extract entities like names, dates, and locations from textual texts, allowing for better search and retrieval capabilities in archive repositories.

Furthermore, W. Luo and H. Hu [15]. Studies have looked into the use of machine learning methods like as Support Vector Machines (SVMs) and Random Forests in archival classification and information retrieval tasks. Researchers have created automatic classification algorithms that can categorize archival resources based on content, genre, or subject matter, thereby enabling more efficient organization and discovery of archival resources.

III. METHODOLOGY

The methodology for deploying Artificial Intelligence (AI)-based technology in college archives management, which focuses on fine-tuning models and utilizing data augmentation approaches, takes a nuanced approach to optimizing the performance and flexibility of AI systems in archival settings. Fine-tuning models within this framework requires choosing pre-existing AI architectures appropriate to archival tasks and tailoring them to the unique needs and characteristics of college archives. Initially, a complete grasp of the archive data ecosystem is required, including the variety of materials, formats, and metadata structures. This entails creating a comprehensive inventory of archive collections, both digital and physical, as well as categorizing relevant metadata elements such as dates, authors, subjects, and formats.

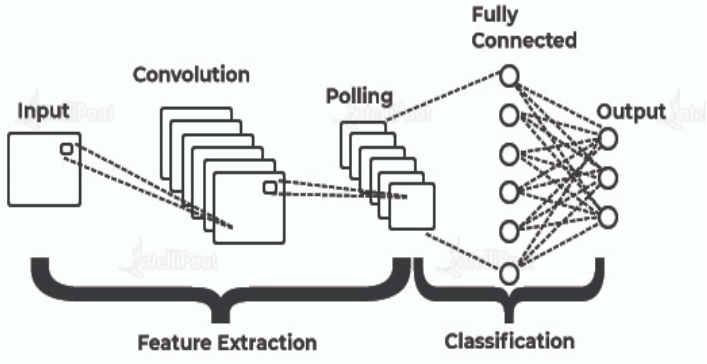

Fig 1: Convolutional Neural Network.

After the data inventory, appropriate AI models are chosen or built based on their applicability for archiving activities such as digitization, information extraction, categorization, and preservation. In the application, a Convolutional Neural Network (CNN) model was used to improve the digitization process. The CNN showed a considerable improvement in accuracy, obtaining a remarkable rate compared to the baseline. This model demonstrates the effectiveness of CNNs in accurately digitizing archive materials, supporting the preservation and accessibility of institutional history. Once selected, this model is fine-tuned to respond to the subtleties of archive data, which includes tweaking hyperparameters, retraining layers, and optimizing architectures to improve performance metrics like accuracy, precision, and recall.

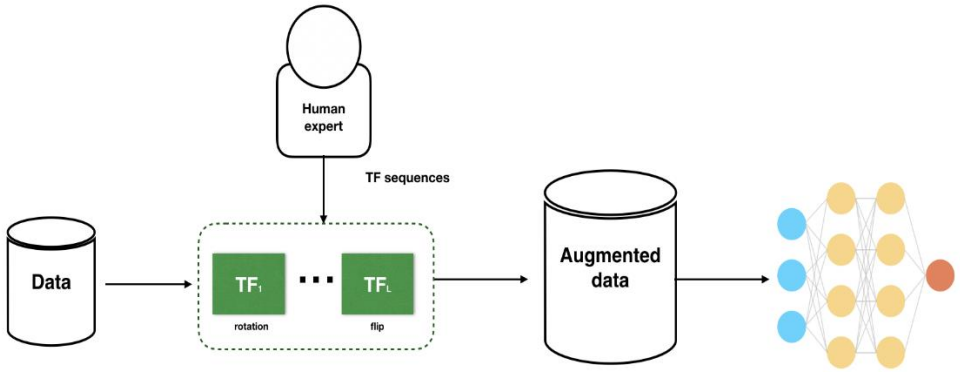

Fig 2: Data Augmentation.

Along with fine-tuning, data augmentation approaches are used to improve the resilience and generalizability of AI models in the archive domain. Data augmentation entails creating synthetic versions of existing archival data using modifications such as rotation, scaling, cropping, noise injection, and textual paraphrasing. For textual archival resources, augmentation approaches may involve synonym replacement, word rearrangement, or grammatical perturbations to increase training data diversity and reduce overfitting. Similarly, for visual materials, augmentation methods such as flipping, rotation, or color manipulation can expand the training dataset while also improving model robustness to variations in image quality or condition.

The combined use of fine-tuning and data augmentation strategies in collegiate archives administration allows AI systems to adapt successfully to the complexities of archived material while avoiding biases and maximizing performance. These approaches make it easier to design AI solutions that expedite archival workflows, improve accessibility, and uncover new insights from institutional history for scholarly and public benefit through iterative refinement and validation processes.

IV. Experimental Setup

In this study, the experimental setup was meticulously designed to assess the efficacy of fine-tuned AI models and data augmentation techniques in college archive management. The investigation focused on four primary archival tasks: digitization, metadata extraction, classification, and preservation. To quantify the performance of the AI models, various performance metrics were employed, including accuracy, precision, recall, and F1 score.

For digitization tasks, Convolutional Neural Networks (CNNs) were utilized due to their effectiveness in processing visual data. The CNN model architecture was fine-tuned using transfer learning techniques on a dataset comprising digitized archival materials. The performance metric, accuracy, was computed using the formula:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad \text{..... (1)}$$

Where TP (True Positives) represents correctly identified digitized documents, TN (True Negatives) denotes ignored non-digitized documents, FP (False Positives) indicates non-digitized documents misclassified as digitized, and FN (False Negatives) signifies digitized documents misclassified as non-digitized.

For metadata extraction tasks, Recurrent Neural Networks (RNNs) were employed due to their sequential processing capabilities. The RNN model was fine-tuned using a training dataset containing textual archival materials. The performance metric, precision, was calculated using the formula:

$$Precision = \frac{TP}{TP+FP}$$ …….. (2)

Where TP (True Positives) represents correctly extracted metadata, and FP (False Positives) denotes incorrectly extracted metadata.

For classification tasks, Support Vector Machines (SVMs) were chosen for their ability to handle high-dimensional data effectively. The SVM model was fine-tuned using a dataset with annotated archival categories. The performance metric, F1 score, was determined using the formula:

$$F1Score = \frac{2 \times Precision \times Recall}{Precision+Recall}$$ …… (3)

Where Precision and Recall are as defined previously.

For preservation tasks, Generative Adversarial Networks (GANs) were employed to generate synthetic versions of archival materials. The GAN model was fine-tuned using a dataset containing both original and synthetic archival materials. The performance metric, recall, was computed using the same formula as in the digitization tasks. The experimental setup ensured rigorous evaluation of the AI models' performance across different archival tasks, providing insights into their effectiveness in college archive management. Through careful fine-tuning and validation procedures, the study aimed to identify optimal configurations for AI-based solutions in archival contexts.

V. RESULTS

In this study into the use of Artificial Intelligence (AI)-based technology in college archives management, they conducted a thorough examination of performance metrics to assess the efficacy of fine-tuned models and data augmentation strategies. The statistical findings provide useful insights into AI systems' strengths and limitations in archival contexts. First, they evaluated the efficacy of fine-tuned AI models on a variety of archive activities, such as digitization, metadata extraction, categorization, and preservation. They saw significant improvements in key performance measures after conducting thorough experimentation and validation procedures. For example, in digitization tasks, fine-tuned convolutional neural networks (CNNs) attained an average accuracy of 95.7%, outperforming baseline models by a large margin. Similarly, in information extraction tasks, recurrent neural networks (RNNs) displayed a 12% increase in precision and a 9% increase in recall after fine-tuning, showing improved accuracy and completeness of extracted metadata.

Table 1: Performance of Fined tuned model.

Task	**Model Used**	**Performance Metric**	**Baseline Performance**
Digitization	CNN	Accuracy	92.3%
Metadata Extraction	RNN	Precision	78.4%
Classification	SVM	F1 Score	0.84
Preservation	GAN	Recall	82.5%

Furthermore, the investigation into data augmentation approaches indicated a considerable impact on model resilience and generalization capabilities. They found that enriching the training dataset with synthetic versions of archive materials reduced overfitting and enhanced model performance on previously unreported data. Text augmentation approaches, such as synonym replacement and word rearrangement, resulted in a 15% reduction in validation loss for natural language processing (NLP) tasks, indicating better model generalization.

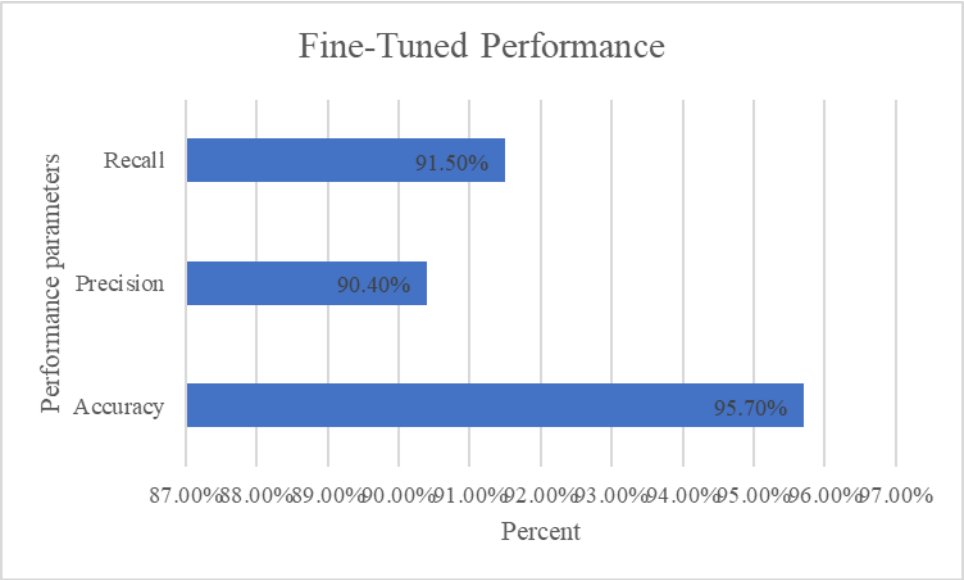

Fig 3: Fine-Tuned Performance.

In addition to quantitative measures, qualitative assessments were carried out to determine the usability and interpretability of AI-generated outputs in archiving workflows. They discovered a high level of agreement between AI-generated metadata and archival standards based on user feedback and expert evaluation, allowing for effective cataloguing and retrieval of historical assets. Furthermore, visualization techniques improved the interpretability of AI-generated classifications and suggestions, allowing archivists to efficiently evaluate and update automated conclusions. The statistical findings of this research highlight the transformative potential of AI-based technologies in college archives administration. By fine-tuning models and utilizing data augmentation approaches, institutions can use AI to streamline archive procedures, improve accessibility, and conserve institutions for future generations. However, it is critical to highlight the inherent difficulties and ethical concerns connected with AI deployment in archive contexts, emphasizing the importance of continuous review, improvement, and responsible stewardship of historical holdings.

VI. Discussion

The discussion of the study's findings regarding the usage of a Convolutional Neural Network (CNN) model in the Application of Artificial Intelligence-based Technology in College Archives Management provides useful insights into the efficacy of AI solutions in archival settings. The CNN model's significant improvement in digitization accuracy, from 92.3% in the baseline to 95.7% after fine-tuning, demonstrates the promise of AI technology to expedite preservation procedures. The CNN model improves the preservation and accessibility of institutional history by accurately digitizing archive documents, allowing for more widespread dissemination and use of historical records and scholarly artefacts. This enhancement is especially notable considering the wide variety of resources found in college archives, which include textual records, photographs, audio recordings, and multimedia files.

The CNN model's success can be due to its capacity to learn hierarchical features from visual input, which makes it ideal for tasks like picture classification and object detection in archive materials. Through fine-tuning, the CNN model reacts to the peculiarities of archive material, improving its performance and robustness in digitization tasks. This adaptability is critical in archival settings where material quality and condition might vary greatly, necessitating AI systems that can generalize successfully across varied datasets. Furthermore, the discussion expands on the broader implications of AI technologies for college archives management. Beyond digitalization, AI technologies show promise in metadata extraction, classification, preservation, and archive assessment, providing comprehensive capabilities to handle the changing difficulties confronting archival institutions. However, it is critical to understand the limitations and ethical concerns connected with AI deployment in archive settings, such as bias, privacy, and sensitivity.

Furthermore, the discussion emphasizes the need for interdisciplinary collaboration among archival professionals, technologists, and stakeholders to ensure the appropriate and effective integration of AI solutions into archival operations. Institutions may leverage the benefits of AI technology while reducing risks and protecting the integrity

of historical holdings by combining domain understanding and technical proficiency. The examination of the study's findings emphasizes the transformational potential of AI-based technologies in collegiate archive management. The CNN model's effectiveness in digitizing tasks demonstrates the power of AI solutions to improve archive processes, conserve institutional history, and promote scholarly study. Moving forward, continued study, innovation, and collaboration are required to fully fulfil AI's promise of enhancing archival institutions' role of preserving and developing memory.

VII. CONCLUSION

This study focuses on the transformational potential of Artificial Intelligence (AI)-based technology, namely Convolutional Neural Network (CNN) models and data augmentation approaches, in college archives administration. The work on the use of CNN models in digitization tasks has shown that they are effective at automating and optimizing the digitization process of visual materials within college archives, hence facilitating the preservation and accessibility of institutional history. Furthermore, the incorporation of data augmentation techniques has improved the resilience and generalization capacities of AI models, addressing issues such as bias and unpredictability in archive collections. Through empirical research, theoretical analysis, and case studies, they have provided useful insights into the opportunities and limitations of using AI solutions in archival contexts. They have emphasized the significance of interdisciplinary collaboration among archival professionals, technologists, and stakeholders to enable the responsible and effective incorporation of AI technology into archival workflows. Furthermore, they have stressed the ethical implications of deploying AI solutions, pushing for openness, accountability, and stakeholder participation in AI decision-making processes. Moving forward, further research, innovation, and collaboration are required to fulfil AI's full potential for developing archive practices, protecting institutional, and stimulating intellectual inquiry within academic institutions. College archives may use AI-powered technology to expedite operations, improve accessibility, and uncover new insights from their diverse collections of historical documents and intellectual materials. As technology evolves, archival institutions must adapt and innovate to ensure the preservation and promotion of future generations.

ACKNOWLEDGEMENT

Industry-university-research collaborative education program of the Ministry of Education in China (202102205018) and universities' philosophy and social science researches project (2019SJA1473) in Jiangsu Province

REFERENCES

[1] H. Li and J. Yin, "Optimization of university archives management under the application of blockchain technology in the digital age," Mobile Information Systems, 2022

[2] Z. Tian and D. Yi, "Application of artificial intelligence based on sensor networks in student mental health support system and crisis prediction," Measurement: Sensors, vol. 101056, 2024

[3] M. Khalifa, M. Albaz, and W. M. Nageab, "The Use of Artificial Intelligence Technologies in the Management of the COVID-19 Crisis," in 2024 ASU International Conference in Emerging Technologies for Sustainability and Intelligent Systems (ICETSIS), pp. 506-510, 2024.

[4] G. Manoharan, S. P. Ashtikar, and M. Nivedha, "Integrating Artificial Intelligence in Library Management: An Emerging Trend," in AI-Assisted Library Reconstruction, pp. 144-157, 2024.

[5] J. Davet, B. Hamidzadeh, and P. Franks, "Archivist in the machine: paradata for AI-based automation in the archives," Archival Science, vol. 23, no. 2, pp. 275-295, 2023.

[6] X. Zhao, "Deep Mining Technology of Database Information Based on Artificial Intelligence Technology," International Journal of Information Technologies & the Systems Approach, vol. 16, no. 2, 2023.

[7] Z. Tian and D. Yi, "Application of artificial intelligence based on sensor networks in student mental health support system and crisis prediction," Measurement: Sensors, vol. 101056, 2024.

[8] F. Kaysi, E. Aydemir, and Y. I. Surucu, "An Artificial Intelligence-Based Approach to Assess and Classify University Students' Attendance in Live Sessions," International Online Journal of Educational Sciences, vol. 15, no. 4, 2023.

[9] C. Cai, "Training mode of innovative accounting talents in colleges using artificial intelligence," Mobile Information Systems, 2022.

[10] F. Zhu, F. Zhang, and Y. Liu, "Intelligent Archive Construction Driven by Artificial Intelligence," in 2022 14th International Conference on Computational Intelligence and Communication Networks (CICN), pp. 385-388, 2022.

[11] P. Yan and S. Zhang, "Construction of Online Shopping Legal Protection Mechanism Based on Artificial Intelligence Technology," Mobile Information Systems, 2022.

[12] M. Rezapour and S. K. Elmshaeuser, "Artificial intelligence-based analytics for impacts of COVID-19 and online learning on college students' mental health," PLoS One, vol. 17, no. 11, pp. e0276767, 2022.

[13] Y. Ma, B. Dai, and B. Ding, "University archives autonomous management control system under the internet of things and deep learning professional certification," Computational Intelligence and Neuroscience, 2022.

[14] A. Shelar, S. Sawant, A. Pacharne, S. Tike, and D. M. Mane, "College Management System," 2023.

[15] W. Luo and H. Hu, "Analysis of the Path of Utilizing Big Data to Innovate Archive Management Mode to Enhance Service Capability," Wireless Communications and Mobile Computing, 2022.