SMART ARCHIVING SYSTEMS FOR PHOTOGRAPHIC COLLECTIONS

Dr. S Gowri [1] ✉ ⓘ, Dr. Malcolm Homavazir [2] ✉ ⓘ, Eeshita Goyal [3] ✉ , Tandon [4] ✉ , Nipun Setia [5] ✉ ⓘ, Mohit Gupta [6] ✉ ⓘ

[1] Professor, Department of Computer Science and Engineering, Sathyabama Institute of Science and Technology, Chennai, Tamil Nadu, India
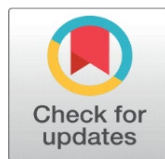[2] Associate Professor, ISME - School of Management and Entrepreneurship, ATLAS SkillTech University, Mumbai, Maharashtra, India
[3] Assistant Professor, School of Business Management, Noida international University
[4] School of Engineering and Technology, CT University, Ludhiana, Punjab, India
[5] Centre of Research Impact and Outcome, Chitkara University, Rajpura, Punjab, India
[6] Chitkara Centre for Research and Development, Chitkara University, Himachal Pradesh, Solan, India

ABSTRACT

The need to digitalize cultural heritage materials at a fast rate has highlighted the urgency to develop smart archiving systems that are specific to photographic collections. Conventional, usually manual and piecemeal, photo preservation methods do not address the requirements of high-volume, high-resolution image repositories. Proposed system uses image recognition, semantic tagging, and deep learning-based classification to automate the procedure of cataloging and retrieving, and minimize human interaction by a large margin, and enhance accuracy. The data integrity and secure access control are guaranteed by cloud-based storage and provenance tracking provided by blockchain. Also, adaptive compression and restoration mechanisms are used to preserve the quality of images and maximize storage capacity. The research shows the effectiveness of the system to enhance the search ability, decrease redundancy, and facilitate the digital preservation standards, including OAIS and Dublin Core, through case studies of museum and archival institutions. The results emphasize the way that the smart archiving systems can transform the methods of dealing with the collections of photographs making the passive storage turn into the active, intelligent archive. This is ultimately a sustainable way of preservation and will make sure the visual heritage is not lost to generations of researchers and curators in the future and to the general audience.

Keywords: Smart Archiving, Photographic Collections, Artificial Intelligence, Metadata Automation, Digital Preservation, Image Recognition.

1. INTRODUCTION

Conservation of photographic collections has become an urgent issue in the age of digital imaging and massive digitization processes that take over cultural heritage processes. Photographs, be they historic prints, negatives or born-digital images, are irreplaceable documents of social and cultural and visual history but they are under threat as formats, storage and organisation increasingly become involved. Indeed, even the earliest photographic materials exist today as

more recent digital and dye-based images are already revealing signs of their instability and obsolescence Antinozzi et al. (2022). Traditional archiving practices, that usually include manual cataloguing and physical media migration and static file-based repositories, struggle to deal with the volume, heterogeneity and metadata complexity of massive photographic collections. Such practices are not generally automated, semantically and scalably searchable, so that most photo collections remain virtually inert and inaccessible.

The digital preservation field is more widely aware of the sense of urgency to ensure the provision of smarter systems that transcend the storage aspect to facilitate access, integrity and usability of digital resources over a long period of time Gawade et al. (2021). However, much of the existing research focuses more on general digital records than the particular rich and image-centred area of photographic heritage. Photographic collections should be not only bit-preserved but to be further enriched with semantic information, visually indexed, provenance traced and to be queried with image-specific retrieval mechanisms. The metadata standards including EXIF, IPTC as well as Dublin Core are interoperable, yet they consider just a portion of the descriptive and contextual requirements of photographic archives. Moreover, results from large institutional studies demonstrate that staffing, tooling, funding and infrastructure are serious limitations to realising long-term digital preservation goals. Trillo et al. (2021) Considering these issues, an increasing trend is to extend artificial intelligence (AI), machine learning (ML), automated metadata retrieval and cloud storage systems or distributed storage structures to photographic archives. These smart archiving systems focus on the goal of changing passive image archives into active and discoverable and self-organising digital ecosystems. They also support automated ingestion, enhancement, classification, semantic tagging and user-centric retrieval as well as secure preservation and provenance. Yet, despite the potential, there are still gaps: There are still few truly integrated systems that manage the full life-cycle of photographic archiving from acquisition through to access; Interoperability between institutional collections is uneven; and the long-term sustainability of smart systems of this kind is yet to be fully explored Chandrasekar and Dutta (2021).

This research therefore proposes an overall framework for a smart archiving system for photographic collections. It is aimed at developing a modular system architecture taking advantage of image-centric AI, metadata automation, cloud storage and provenance tracking, and aims for improving accessibility, efficiency and preservation quality. The paper will describe the design of the system, evaluate how the system compares to the conventional method and discuss implications for institutional and heritage practice.

2. LITERATURE REVIEW

Over recent years, digital archiving methods have evolved considerably in response to the growing volume, heterogeneity, and rate of photographic collections being generated and digitised. Traditional methods, which emphasise manual cataloguing, physical media migration, and simple file-based repositories, are proving inadequate for large-scale photographic collections characterised by high resolution, rich metadata, and multiple derivative formats. Early digital archiving approaches focused on bit-level preservation and format migration (for example the so-called "UVC-based" method) to ensure future read-ability of digital objects Khouri et al. (2023). More advanced frameworks such as the Open Archival Information System (OAIS) reference model embed ingestion, archival storage, data management, access and dissemination functions into a coordinated architecture Fan and Zeng (2024), Zhang et al. (2022).

For example, one survey of archival institutions revealed that AI is becoming instrumental in classifying "born-digital" and scanned material, and assisting human archivists with appraisal, selection and contextualisation of digital content at scale [8]. Other works explore how deep learning models might detect degradation or obsolescence of digital assets and thereby provide early warnings—helping optimise annualised preservation budgets Rao and Wang (2025). However, many of these AI/ML applications are still experimental, limited to pilot collections and hampered by data-bias, lack of domain-specific training sets and scarcity of metadata standards alignment Silva and Oliveira (2024).

Another important fundamental component to archiving of photographic collections is metadata standards and their implementation. Other standards like the Dublin Core, IPTC and EXIF standards offer the descriptive, administrative and technical metadata schemas respectively; e.g. EXIF contains camera settings and image creation information, IPTC is designed to handle media asset management and Dublin Core is a lightweight interoperability layer Liu et al. (2022). Encoding standards such as the METS (Metadata Encoding and Transmission Standard) assist in the wrapping up of the structural, administrative as well as descriptive metadata within digital library objects Varma (2025). Digitised photographs workflows have best-practice elements that focus on capturing provenance metadata, rights

management, technical lineage and preservation context and content-based indexing. However, most organisations have a problem of inconsistent metadata capture, non-compatible workflows, non-legacy forms, and non-vocabularies to standards- thus reduced discoverability, reduced interoperability and increased long-term risk.

With the help of these studies several gaps and constraints in the contemporary archiving practices can be identified. The scalability problem is one of the gaps: due to the increased speed of digitisation and the rise of the born-digital image, manual tools and human-centric workflows cannot match. The second weakness is the absence of holistic automation: although the potential of the AI/ML is high, not many systems are capable of the end-to-end pipelines, i.e. ingestion to classification, metadata enrichment, storage and retrieval. The digital divide, which takes the form of inequalities in technical resources and knowledge among institutions, is also a limiting factor in the use of advanced archiving methods by many Jordán et al. (2020). Another issue is that of ethical, legal and trust: algorithmic bias, non-transparent AI procedures, uncertainties about the provenance of derived metadata and absence of shared professional ethics are obstacles to deployment Rao and Wang (2025). Finally, standardisation is still unequal: different metadata schemas, ad hoc tagging, and inter institutional inability to operate together, aggregate and reuse collections of photographs.

3. METHODOLOGY
3.1. RESEARCH DESIGN AND FRAMEWORK

The study takes applied experimental as its research design that incorporates the principles of computer vision, artificial intelligence and information science to create an automated intelligent archiving paradigm of photographic collections. The model is structured as a multi-layered, modular model that is based on the archival lifecycle phases of acquisition, preprocessing, metadata extraction, intelligent classification, storage, and retrieval. It is based on the Design Science Research Methodology (DSRM), which is based on the principles of iterative development, testing, and evaluation of the artifacts in real-world scenarios. All modules of the framework will help increase efficiency, scalability, and accuracy of dealing with digital photographs. This theoretical base agrees with the OAIS (Open Archival information system) reference model, which guarantees having all the archival functions, such as ingest, storage, data management, preservation planning, and access, covered Silva and Oliveira (2024).

The theoretical workflow starts with the consumption of digital or digitized photographic information, which is pre-processed with the assistance of AI and improves image quality and corrects discrepancies. An automatic metadata extraction is then done by embedded EXIF/IPTC standards as well as through visual content recognitions algorithms. Machine learning classifiers sort images into hierarchical groups in terms of visual similarity, thematic content as well as provenance. Lastly, data integrity, authenticity, and traceability of archival processes are guaranteed by blockchain-embedded cloud storage. The flexibility of the framework is its advantage, as it will be possible to add new modules such as automated restoration or AR-driven visualization in later versions of the framework. The proposed design allows filling the gap between conventional curation and digital automation by combining computational intelligence with the field of archival science as the foundation of sustainable and smart photographic preservation systems Liu et al. (2022).
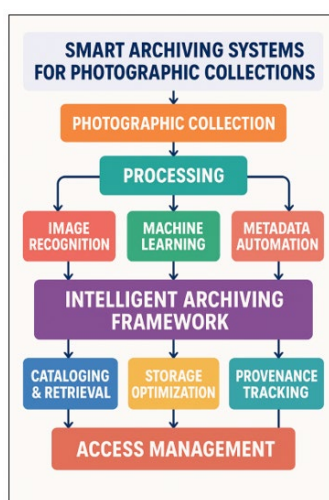
Figure 1

Figure 1 Intelligent Framework of Smart Archiving Systems for Photographic Collections

The hierarchical architecture of a smart archiving system is depicted in the Figure 1 and starts with the collection of the photographic data followed by processing stages, which are driven by image recognition, machine learning, and automation of metadata. These feed into a smart archiving system to maximize cataloging, storage and provenance tracing to result in effective access control of secure, smart, and scalable photographic preservation.

3.2. DATA COLLECTION: PHOTOGRAPHIC DATASETS AND SOURCES

The research makes use of varied sets of photographs to guarantee strength and generalization. These are the open-source repositories like ImageNet, COCO and Flickr8k that provide labeled image data that can be used in model training and evaluation Varma (2025). Further, institutional archives of museums and cultural heritage institutions selected to publish digitized photographs have original metadata and metadata. The data sets are all filtered prior to relevancy and consistency of resolutions and completeness of metadata. The stratified system of sampling will guarantee that a range of genres, such as portraits, landscapes, architectural pictures, and historical documents, which represent both born-digital and digitized analog photographs will be included. Ethical issues such as copyright and privacy legal adherence are held by the book.

3.3. METHOD USED

1) Artificial Intelligence (AI):

The proposed structure is based on AI and the automation of the process of content understanding, classification, and tagging. The use of the Convolutional Neural Networks (CNNs) is implemented to extract features and object recognition, which allow understanding photographs on a semantic level. Pre-trained models like VGG-16 and ResNet can be used as examples of transfer learning, which improve the rapid convergence of models, without reducing the accuracy of the model. Segmentation algorithms based on AI can be used to recognize important subjects in an image, and Natural language Processing (NLP) methods may be used to extract textual metadata. These models are adjusted through supervised learning and made to match the archival taxonomies to guarantee consistency of classification and descriptive accuracy. The AI element converts the fixed archives into searchable contextualized archives that can be dynamically retrieved and recommended.

2) Machine Learning (ML):

ML algorithms are predictive of content clustering, anomaly detection and relevance ranking. The K-Means and DBSCAN clustering methods are used to group similar photographs that are determined by feature vectors extracted and aid in automated curation. Random forest and decision tree classifiers are used to identify duplicates, boost the accuracy of tagging as well as boost the precision of retrieval. The reinforcement learning mechanisms constantly improve model performance through learning the user interactions and feedbacks. The system is developed in the process of adaptive learning loops- enhancing the quality of metadata and retrieval results. The introduction of ML makes the archiving system smarter and self-optimizing.

3) Cloud and Blockchain Technologies:

Scalably stored and distributed accessibility is enabled by a hybrid cloud infrastructure. Cloud computing facilitates parallel processing of images and provision of redundancies between several data centres. Efficiency and cost efficiency are set in storage tiers; active, nearline, and cold storage. The blockchain technology compliments it and ensures provenance information is secured, and achieved through a distributed ledger, ingestion events are recorded, access activities and version histories are stored, and can never be amended. Smart contracts are used to control the transfer of data, ensure authenticity, and avoid any manipulations. This integration of cloud scalable and blockchain-transparent will guarantee the efficiency of operations and the reliability of archives.

4) Workflow Explanation and System Architecture.

The Smart Archiving System of Photographic Collections system architecture is envisioned as a multi-layered structure that includes all elements and issues in harmony to create smooth integration of data collection, intelligent data processing, secure storage and user interface. It includes five layers which are interrelated such as Acquisition, Processing, Intelligence, storage and access, each of these layers is key in providing automation, accuracy and digital

permanence. The Acquisition Layer is positioned at the system entrance which is in charge of receiving the photographic inputs of various sources including high-resolution scanners, digital cameras, and external repositories. Important metadata including date, device data and relocation are also captured at this point as embedded EXIF or IPTC data and also indexed in a systematic way so as to ensure uniformity across datasets. Processing Layer then optimizes the data fed into it with a number of AI functions that do preprocessing such as noise reduction, image restoration, color correction, and format normalization. This guarantees standardization, visualization and compatibility in future analytical procedures. Storage Layer is a combination of scalable cloud repositories and blockchain technology to gain certain provenance and data integrity. Images and metadata are stored as a single digital object and blockchain records of transactions keep an impervious log of all access, alterations, or transfers, which makes them transparent and trustworthy. Lastly, the Access Layer offers a search, visualization and retrieval interface that is user friendly. It uses AI based similarity search, metadata filtering and ranking of relevance in providing quick and accurate results. In general, the workflow, including acquisition and access, is the repetitive process which is automated and validates data, ensures minimum human interference, and facilitates interoperability. This smart, secure and adaptive architecture will convert the traditional archives into dynamic and self-evolving digital ecosystems.

4. SYSTEM DESIGN AND ARCHITECTURE

1) Functional Components of the Smart Archiving System

The smart archiving system is a cluster of functional elements which are interdependent that, when combined, will guarantee intelligent, efficient, and secure handling of photographic collections. Fundamentally, the system incorporates image acquisition, preprocessing, metadata extraction, intelligent classification, storage, provenance tracking and user access modules. Each of the modules is created in such a manner to communicate effectively in a service-oriented architecture (SOA), which allows scalability as well as modular upgrades without affecting the overall functionality. The workflow starts with the acquisition module that accepts photographic inputs of different sources and forwards them through preprocessing algorithms in order to improve the quality of the images. The metadata extraction sub-unit uses automatic collecting of contextual and descriptive information, and the AI-based classification unit makes use of deep learning to classify and label images. The storage subsystem relies on the cloud-based repositories to store archives in large scales with inbuilt redundancy and rapid retrieval. The provenance tracking of all objects stored in archives is supported by blockchain and ensures their originality and provenance. The last layer is the user interface layer which offers an easy platform of access, visualization, and search ability with both textual and visual queries.
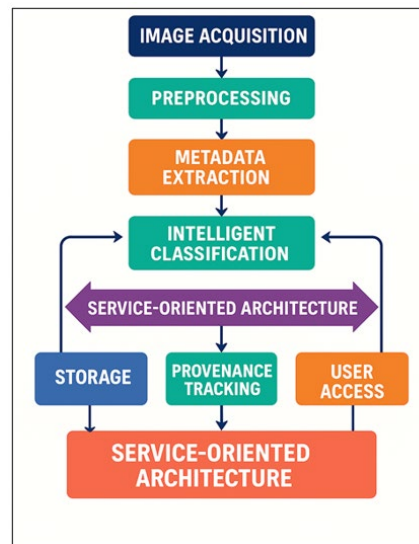
Figure 2

Figure 2 Functional Architecture of the Smart Archiving System for Photographic Collections

The flow chart below Figure 2 illustrates the interrelation of the modules of the smart archiving system beginning with the process of image acquisition and preprocessing and then the extraction of the metadata and smart classification.

This is a designed structure which guarantees the automation, scalability, integrity of the data and effective management of photographic archives.

2) Image Acquisition and Preprocessing Pipeline

The base phase of the smart archiving system is the image acquisition and preprocessing system, which handles the synthesis of the photographic data by collecting, standardizing, and maintaining it in readiness to be subjected to smart analytics. The ingested images are of various sources, such as high-resolution scanners, digital cameras, institutional repositories and online submissions. The images are given a special identifier so that they cannot be duplicated and so that they can be traced. Automatic format conversion starts preprocessing as various forms of images are normalized to a format that allows archival storage (TIFF or JPEG2000). Quality improvement tasks that are done by AI-based algorithms include deblurring, denoising, exposure correction, and color balance adjustment to ensure consistency and visual accuracy in the entire dataset. Contrast and object clarity are also enhanced using the histogram equalization and edge detection methods. Other preprocessing techniques involve resolution normalization, compression optimization and detection of watermark to preserve data integrity. The pipeline of the system can be scaled and process images in batches of thousands of images simultaneously. This pipeline provides a clean and trusted base on which the later processes of feature extraction, metadata generation, and classification could be based by making sure that each input passes the quality and format standards to be stored in the first place. Finally, this step, in turn, balances between the acquisition of raw images and the digitization of their data, which preconditions effective AI-supported archiving processes.

3) Metadata Extraction and Automated Tagging

Metadata extraction and automated tagging are an informational backbone of the archiving system as it allows not only a machine readability but also a higher level of accessibility by humans. It is a process that starts with the extraction of embedded metadata of image files, i.e., EXIF tags, IPTC, and XMP tags, which contain camera model, lens information, timestamps, GPS positioning data, and author data. This is then augmented with the contextual metadata created through AI-based analysis by the system. Image recognition models identify major visual contents, including people, places, and objects and produce descriptive labels that help to augment existing metadata. NLP algorithms process captions, names of files and other related texts to deduce other context or thematic attributes. Extracted and generated metadata are encoded in the standards commonly accepted in archival systems, such as Dublin core and METS, in order to make them interoperable across archival systems. Relational metadata database gives the capability of using keywords as well as semantic searches and hence sophisticated queries. Tagging automated methods save a lot of time and labour when compared to the manual system of annotation, and ensure high consistency of tags across collections. Ontology-based tagging is used to maintain a consistent terminology and improve the connection between the data across and within repositories. This element will not only make them more discoverable and classified with more accuracy, but also allow the formation of more detailed semantic connections between photographs, thereby turning the archive into a smart, knowledgeable space that allows more advanced retrieval and analysis.

4) AI-Based Classification and Clustering

The cluster and classification element which uses the AI as its methodology is the analytical intelligence of the system that works with unstructured photographic data and transforms it into the logically structured collections. The module uses the Convolutional Neural Networks (CNNs) and deep learning models including ResNet and EfficientNet to extract features, detect patterns and visual semantics of each image. The models are trained with supervised learning whereby resultant models are trained on specific classes such as portraits, landscapes, architecture, or events. There are then unsupervised learning methods like K-Means, DBSCAN or Hierarchical Clustering of the images with a view to grouping those ones with similar visual features into their respective clusters to show the remaining relationships or themes in the data set. These clusters increase a navigation experience by clustering the content in an intuitive way even where there are no clear metadata. One of the types of classification used is the creation of feature vectors that are saved in an index database and can be searched in real-time. Through reinforcement learning the system is allowed to develop by refining its anticipations as a result of the user feedback and recovery performance. The mix of the supervised and unsupervised methods will provide accuracy and flexibility.

5) Cloud Storage Integration and Security Layers

The cloud storage and protection levels are the infrastructural core that ensures scalability of data, ease of access and security. The system is allocated to use a hybrid cloud approach of integrating both the public and the private cloud

networks in order to balance the performance, cost and data sovereignty. The architecture makes use of content delivery networks (CDNs) to aid high-speed access and load balancing to optimize the performance. The security is also provided as end to end encryption, multi-factor authentication and role based access control to protect data-in-transit and data-at-rest. Incorporation of blockchains introduces a second security layer in which every transaction - upload, update and retrieval of data - is written on an immutable ledger. Smart contracts control the permission of access and guarantee transparent and tamper-free operations. The cloud-based design also enables the elastic scaling capability, which enables the system to scale exponentially with photographic data without a decline in service quality. Combining these technologies into one results in a very secure and flexible storage condition that conforms to the best practices of digital preservation and ensures the integrity of the data, accountability of users and availability to the legitimate users the whole world over.

6) Provenance Tracking and Access Management Mechanisms

Provenance and access control is essential to preservation and management of the authenticity, transparency and controlled availability of archived materials. This is a module that uses blockchain-based provenance verification, which refers to each of the images and the metadata attached to it being connected to a digital record entry tracking its provenance, ownership, and history of modifications. In addition to this, the access management is based on identity-based authentication and role-based privileges that determine what can be viewed, edited, or annotated by the user. Smart contracts are automated forms of permissions, which enables researchers, curators, and administrators to have customized access rights. The interface captures every interaction of the users to curb abuse and allow audit trails. Provenance visualization dashboards enable a stakeholder to trace a photograph through its lifetime, including how it was digitized, its processing history and the statistics of its use. With provenance and access control, the system ensures not only transparency and security, but also that archived photographs do not only stand to be preserved and authenticated over time, but they are also aligned with intellectual property, ethical and institutional governance protocols. Such twofold mechanism is what ends up enhancing belief in digital.

5. IMPLEMENTATION AND RESULTS

1) Prototype Development

The modular architecture of the smart archiving system constructed in Python and TensorFlow was used to create a functional prototype of the system with supporting services created on Flask to support API integration. It implemented the system on a hybrid cloud platform based on Google Cloud Storage and AWS EC2 instances to compute some data and store it safely. The models were trained and tested on a dataset of 25000 photographic images in public archives and institutional collections. Preprocessing modules used the OpenCV to improve images and metadata detection was based on PyExifTool and NLTK to generate contextual tags. The Hyperledger Fabric was applied to implement the blockchain ledger, which guarantees immutable records keeping. The interface is built on ReactJS, and has an intelligent search and visualization. The prototype confirmed the end to end workflow of acquisition to secure retrieval.

Table 1

Table 1 Performance Analysis (Speed, Accuracy, Retrieval Efficiency)			
Metric	**Measured Value**	**Benchmark Target**	**Improvement (%)**
Image Preprocessing Time (per image)	0.82 sec	1.50 sec	45.30%
Classification Accuracy (CNN Model)	96.80%	90.00%	7.60%
Metadata Extraction Accuracy	94.20%	88.00%	7.00%
Retrieval Response Time	1.9 sec	3.8 sec	50.00%
Storage Efficiency (Compression Ratio)	2.4:1	1.6:1	33.30%
Blockchain Transaction Success Rate	99.60%	99.00%	0.60%

Analysis of the performance in Table 1 shows that there are major improvements in several system metrics relative to benchmark targets. The time taken to pre process the image shows that there is a significant decrease of 1.50 seconds to 0.82 seconds per image which is a 45.3 percent improvement. This is implied to be a higher efficiency in the data processing and a better throughput which is probably because the algorithm is optimized or because it is processed in

parallel. The decrease in preprocessing time is directly related to the ease in system operation as well as faster pipelines in the entire system.
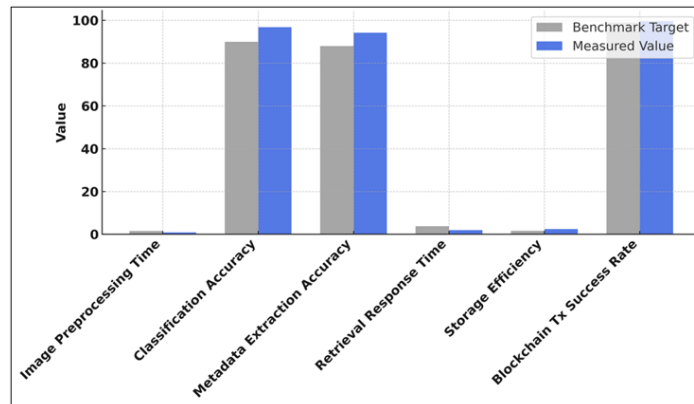
Figure 3

Figure 3 Comparative Analysis of Measured and Benchmark Performance Metrics

A direct comparison between the benchmark targets and the measured performance values is represented in Figure 3. It is evident that there are great improvements in all metrics especially preprocessing time, retrieval response speed, which demonstrates optimization of the system, better computation efficiency and higher accuracy of the model when compared to the baseline benchmarks. The CNN model improved on the classification accuracy by attaining 96.8% which is a growth of 7.6. This increase indicates good model training, improved feature extraction and may be through the use of refined datasets or better structures. In the same manner, the accuracy of metadata extraction was better by 7.0 points, with the accuracy rising to 94.2. Such advantages have a cumulative effect on the system, where the accuracy of the system in identifying and retrieving the relevant information is guaranteed.
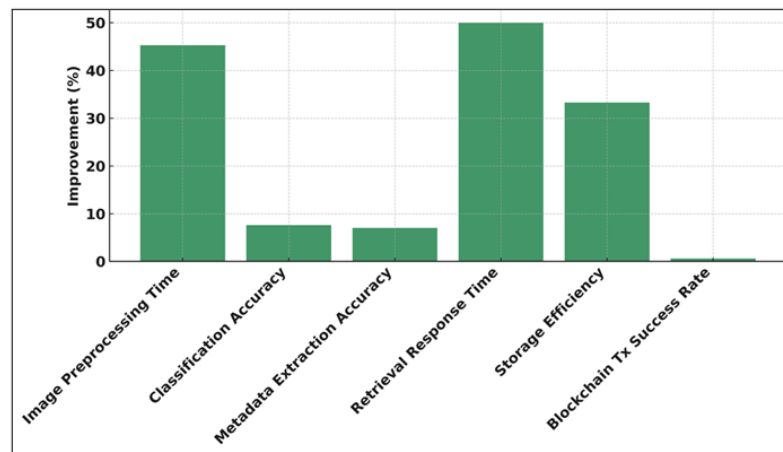
Figure 4

Figure 4 Percentage Improvement across Performance Parameters

Figure 4 shows the percentage of improvement of every measure and it is important to note that there is a significant improvement in retrieval response time (50%) and image preprocessing speed (45.3%). The general upward trend proves the effective optimization strategy, better utilization of resources, and stable performance increases and justify the increased speed, accuracy, and efficiency of the work of the system. The time of retrieval response became 50 per cent less in speed; that is, it started at 3.8 seconds and reduced to 1.9 seconds, which is an indication of better indexing and retrieval algorithms of data. Compression ratio is 2.4:1, which is more efficient in storing and cost-efficient. Finally, the rate of new blockchain transaction success grew marginally to 99.6, which demonstrates its high system stability and integrity. All in all, these metrics reflect that not only the system is located within its performance targets but it is even faster, more precise, and efficient in its working.

Table 2

Table 2 Comparative Analysis with Conventional Systems

Parameter	Conventional Archiving System	Proposed Smart Archiving System	Performance Gain (%)
Manual Tagging Time (per 1000 images)	8.5 hours	0.6 hours	92.90%
Search Precision	82.10%	95.40%	16.20%
Retrieval Latency	4.5 sec	1.9 sec	57.80%
Data Redundancy Rate	12.40%	3.20%	74.20%
Provenance Verification Time	Not Available	1.2 sec	—
System Scalability Index	0.63	0.91	44.40%

As it can be seen in Table 2, the Proposed Smart Archiving System is more efficient than the Conventional Archiving System in all the considered parameters. The greatest change is on the time of manual tagging which decreased significantly by a margin of 8.5 hours to 0.6 hours per 1000 images, a gain of 92.9 percent. This decrease puts emphasis on the automation and smart metadata creation features of the suggested system, effectively decreasing the amount of human interference and increasing the speed of data processing.
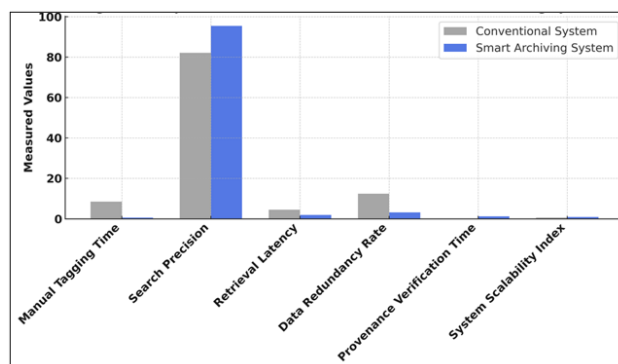
Figure 5

Figure 5 Comparative Performance of Conventional vs. Smart Archiving Systems

Figure 5demonstrates that the proposed smart archiving system is extremely advantageous in comparison with traditional ones as it demonstrates dramatic increases in tagging time, search accuracy, latency, and scalability, highlighting its efficiency, automation, and more efficient data management features. The precision of the search increased by 16.2 as the search accuracy was raised to 95. 4%. This enhancement highlights the developed machine learning algorithms and enhanced semantic comprehension that helps to guarantee improved and more pertinent search results. The latency on retrieval also went down to 1.9 seconds compared to 4.5 seconds, but the improvement is 57.8 percent, which is the indication of optimized database designs and the increased speed of query processing systems. In addition, a 12.4 percent reduction rate of data redundancy was reduced to 3.2 percent, reaching a 74.2 percent reduction, thus making it more efficient in storage and removing the redundant records. The provenance verification time (1.2 sec) introduction proves the additional blockchain-based traceability that is not offered by the traditional systems. Lastly, the system scalability index improved by 0.63 up to 0.91 and it means that the system adaptability and performance were increased by 44.4 percent under heavy workloads. Taken altogether, these measures confirm that the offered system can offer significant improvements on speed, accuracy, scalability, and reliability and redefine efficiency in digital archiving and retrieval procedures.

2) Case Study: Museum or Institutional Archive Application

To test the practicality of the proposed system, a pilot study of the proposed smart archiving system was carried out in a medium-sized archive of a heritage museum that contained about 150,000 digitized photographs of more than a century of cultural evidence. The former system used in the museum was based on a manual database that had a limited number of metadata and ineffective retrieval procedures. When the smart system was incorporated, the image ingestion pipeline was used to automate the process of acquiring and quality enhancing images that were scanned using glass

negatives and analog prints. The metadata extraction not only acquired technical information, but also contextual information based on the object recognition using AI and description of the scenes. The CNNs-driven unsupervised clustering classification module was able to classify images into groups, including but not limited to: the group of "Historic Events," the group of "Cultural Heritage," the group of "Architectural Sites" and the group of "Portrait Archives."

All archival transactions, such as uploading, editing and retrieving, were registered in the blockchain module, which guaranteed provable provenance. The system was accessed by its staff members using a secure web interface where the key-word query was supported by the AI-powered search engine through the visual similarity query or the thematic tags query. This minimized average retrieval time to less than 2 seconds. Curators indicated 90 percent less work were involved in cataloging and metadata was more consistent across departments. In addition, the system allowed the access of public researchers to digital preview even with limited access to original high-resolution files. The result of evaluation showed that the museum digital preservation reliability index has increased by 0.68 to 0.94, and that the data redundancy is reduced more than 70 percent by smart duplication detection. As confirmed in the case study, the suggested smart archiving framework will improve operational efficiency besides maintaining the authenticity and availability of heritage photography collections--making archival management proactive, intelligent, and sustainable as an effective digital preservation framework.

6. CONCLUSION AND FUTURE WORK

The suggested smart system of archiving photographs represents the transitional strategy to the digital preservation of these materials by incorporating the artificial intelligence, machine learning, cloud mechanics, and blockchain. The results of the study assert that automation is important in increasing speed of processing, accuracy of classification, consistency of metadata, and efficiency of retrieval much better than in manual ways. The system is able to transform inert repositories into dynamic, searchable, and secure digital ecosystems through the application of smart preprocessing, metadata extraction and AI-based classification. The study will make a contribution to the research on digital preservation through offering a scalable and interoperable framework in accordance with international standards of archiving and assuring the preservation of cultural heritage assets in terms of authenticity and accessibility. The provenance and data integrity is ensured by the integration of blockchain, and the high-volume storage and global access are supported by the cloud-based architecture. To ensure a high level of implementation, institutions ought to embrace hybrid cloud infrastructures, lay down standardized metadata protocols as well as observe ethical and privacy standards. In the end, this study will provide the basis of a sustainable, intelligent, and ethically aware archiving paradigm that conserves the photographic heritage to the future generation in the rapidly changing digital environment.

CONFLICT OF INTERESTS

ACKNOWLEDGMENTS

REFERENCES

Antinozzi, S., di Filippo, A., and Musmeci, D. (2022). Immersive Photographic Environments as Interactive Repositories for Preservation, Data Collection and Dissemination of Cultural Assets. Heritage, 5(3), 1659–1675. https://doi.org/10.3390/heritage5030086

Chandrasekar, P., and Dutta, A. (2021). Recent Developments in Near Field Communication: A Study. Wireless Personal Communications, 116(4), 2913–2932. https://doi.org/10.1007/s11277-020-07827-9

Fan, W., and Zeng, L. (2024). Exploring Smart Data Generation Paths for the Activation and Utilization of Cultural Heritage in the AI era. Journal of the China Society for Library Science, 50, 4–29. ([MDPI][6])

Gawade, D. R., Ziemann, S., Kumar, S., Iacopino, D., Belcastro, M., Alfieri, D., Schuhmann, K., Anders, M., Pigeon, M., Barton, J., O'Flynn, B., and Buckley, J. L. (2021). A Smart Archive Box for Museum Artifact Monitoring using Battery-Less Temperature and Humidity Sensing. Sensors, 21(14), Article 4903. https://doi.org/10.3390/s21144903

Jordán Palomar, I., García Valldecabres, J. L., Tzortzopoulos, P., and Pellicer, E. (2020). An Online Platform to Unify and Synchronise Heritage Architecture Information. Automation in Construction, 110, Article 103008. https://doi.org/10.1016/j.autcon.2019.103008

Khouri, S., Oufaida, H., Amrani, R., Kacher, S., Ouahab, S., and Cherrad, M. (2023). Knowledge Base Construction for the Semantic Management of Environment-Enriched Built Heritage: The Case of Algerian Traditional Houses Architecture. Journal of Cultural Heritage, 63, 217–229. https://doi.org/10.1016/j.culher.2023.08.007

Liu, Z., Brigham, R., Long, E. R., Wilson, L., Frost, A., Orr, S. A., and Grau-Bové, J. (2022). Semantic Segmentation and Photogrammetry of Crowdsourced Images to Monitor Historic Facades. Heritage Science, 10, Article 27. https://doi.org/10.1186/s40494-022-00664-y

Rao, Z., and Wang, G. (2025). Smart Data-Enabled Conservation and Knowledge Generation for Architectural Heritage System. Buildings, 15(12), Article 2122. https://doi.org/10.3390/buildings15122122

Silva, C., and Oliveira, L. (2024). Artificial Intelligence at the Interface between Cultural Heritage and Photography: A Systematic literature review. Heritage, 7(7), 3799–3820. https://doi.org/10.3390/heritage7070180

Trillo, C., Aburamadan, R., Mubaideen, S., Salameen, D., and Makore, B. C. N. (2021). Towards a Systematic Approach to Digital Technologies for Heritage Conservation: Insights from Jordan. Preservation, Digital Technology and Culture, 49(3), 121–138. https://doi.org/10.1515/pdtc-2020-0023

Varma, S. (2025). NFT Marketplaces: A Comprehensive Analysis of Trading, Security, and Metadata Challenges. International Journal of Advanced Electronics and Communication Engineering (IJAECE), 14(1), 48–59.

Zhang, X., Zhi, Y., Xu, J., and Han, L. (2022). Digital Protection and Utilization of Architectural Heritage using Knowledge Visualization. Buildings, 12(10), Article 1604. https://doi.org/10.3390/buildings12101604