Blockchain-Based Barangay Document Management System with OCR Data Extraction

Katherine D. Abubo[1*], Eldrin Dave A. Ramos[2], Faith A. Valencia[3]

[1]*Ariel Antwaun Rolando C. Sison,*
College of Information Systems and Technology Management
Pamantasan ng Lungsod ng Maynila (University of the City of Manila)
aarcsison@plm.edu.ph

[2] *Diony S. Abando*
College of Information Systems and Technology Management
Pamantasan ng Lungsod ng Maynila (University of the City of Manila)
dswabando@plm.edu.ph

[3] *Mark Christopher R. Blanco*
College of Information Systems and Technology Management
Pamantasan ng Lungsod ng Maynila (University of the City of Manila)
mcrblanco@plm.edu.ph

[4] *Vivien A. Agustin*
College of Information Systems and Technology Management
Pamantasan ng Lungsod ng Maynila (University of the City of Manila)
vaagustin@plm.edu.ph

Article Information

Received: 21-11-2024
Revised: 28-11-2024
Published: 5-12-2024

Keywords
Blockchain Technology; Document Management System; Optical Character Recognition (OCR); Data Security

****Correspondence Email:***
kdabubo2021@plm.edu.ph

Abstract
This study introduces a Blockchain-Based Barangay Document Management System enhanced with Optical Character Recognition (OCR) to address the challenges of manual document handling in Philippine barangays. The system utilizes Blockchain to create an encrypted, tamper-proof ledger, ensuring data integrity and transparency by decentralizing document storage and preventing unauthorized alterations. AES-256 encryption and hashing algorithms protect transactions, while OCR technology digitizes physical records with 96% accuracy, reducing processing times by 50% and minimizing manual errors. The integration of Merkle Tree ensures each document's uniqueness, and the system was evaluated against ISO 25010 standards, achieving high functionality, performance, and user satisfaction. This solution improves document security, reduces administrative workloads, and enhances service efficiency, fostering transparency and trust in local governance.

1. Introduction

This study explores the challenges of traditional paper-based document management in Barangay 494 and Barangay 99, which often result in inefficiencies, security risks, and data inaccuracies. To address these issues, a Blockchain-Based Barangay Document Management System incorporating Optical Character Recognition (OCR) was developed. The system ensures data security and integrity by creating a decentralized, immutable record using blockchain technology, while OCR reduces manual errors and optimizes document processing. The integration of Merkle tree hashing and AES-256 encryption guarantees tamper-proof records, and the OCR component, powered by Google Cloud Vision, achieves 96% accuracy in data extraction, cutting processing time by 50%. User evaluations indicated high satisfaction, with an average rating of 3.8 out of 4, highlighting the system's effectiveness in enhancing transparency, security, and efficiency in barangay document management. This solution offers a practical model for future e-governance projects at the local level.

1.1 Literature Review

The development of BALANGAY, a web-based system designed to centralize barangay services, demonstrates the effectiveness of digital platforms in enhancing accessibility for residents. This system integrates features like an incident heatmap, which assists barangay officials in formulating data-driven community programs. User feedback has been overwhelmingly positive, highlighting the benefits of centralized digital systems in managing barangay services. BALANGAY aims to streamline barangay processes by establishing centralized database and simplifying procedures, thereby enabling more efficient program development (Bautista et al., 2023).

Similarly, Balila et al. (2022) emphasizes the potential of digital transformation in barangay management with the Barangay Information Management System (BIMS), which scored an average of 4.58 in user evaluations. BIMS automates traditional record-keeping, ensuring the security of barangay records, processing pre-filled document requests, and safeguarding session details. This study focuses on the integration of blockchain and Optical Character Recognition (OCR) in barangay systems, emphasizing the advantages of enhanced security, data integrity, and automation. Blockchain's transparent and tamper-proof properties, along with OCR's accuracy in minimizing manual errors, address the common pitfalls of traditional paper-based systems.

Jayaprakash et al. (2024) further support the shift to blockchain-based solutions by demonstrating how the enhanced Merkle hash tree method reduces encryption and decryption times. This advanced method targets the inefficiencies, errors, and security vulnerabilities of traditional barangay document management. The integration of blockchain technology ensures decentralized, secure data storage, minimizing the risks of data tampering and unauthorized access. Additionally, OCR enhances data accuracy and accelerates document processing, fostering more reliable and accessible services for barangay residents. This combination of technologies modernizes document management, alleviating the workload for barangay officials while boosting resident trust and engagement.

In a broader context, Clavin et al. (2020) explore the utility of blockchain technology in government operations, emphasizing its potential to enhance transparency and reduce corruption. This aligns with efforts to integrate blockchain into barangay systems, ensuring decentralized, transparent, and secure document management. Implementing blockchain with OCR not only aligns with local governance goals but also contributes to the overall drive toward digital innovation and improved efficiency in government operations.

Data integrity and security remain critical issues in digital governance, particularly with the rise of cloud storage solutions. A proposed hybrid encryption algorithm combining Advanced Encryption Standard (AES) and Elliptic Curve Cryptography (ECC) alongside the Merkle hash tree structure ensures the secure storage of user data on the cloud. This hybrid approach allows users to access and manage their data securely while

preventing unauthorized access and tampering. The algorithm's effectiveness was demonstrated through simulations, which showed high accuracy and efficient encryption and decryption times, confirming its reliability in maintaining data security within a blockchain environment.

Merkle hash trees are instrumental in handling large datasets efficiently, as they support privacy-preserving public auditing by utilizing hashtags for leaf nodes and organizing hash information for non-leaf nodes. This structure simplifies data mapping and supports secure digital storage, as demonstrated by Lackner, Mirhosseini, and Craß (2021). They propose a blockchain-based method to ensure data traceability and integrity, responding to the vulnerabilities associated with traditional digital storage systems. Zhu et al. (2023) further highlight the growing use of blockchain in securing log storage, mitigating the risks of illegal alterations in conventional systems, though challenges remain in managing large volumes of data efficiently.

The Barangay Document and Issuance System (BDIS), as evaluated by Balila et al. (2022), received high usability ratings, with a System Usability Scale score of 4.58, indicating strong user approval. Participants valued the system's efficiency, ease of use, and the reduction in administrative workload. By leveraging advanced ECC within the blockchain framework, the study demonstrated faster key generation, encryption, and decryption times, confirming the system's capability to enhance cloud security. These findings highlight the effectiveness of blockchain-integrated models in streamlining document processing and securing digital records, offering a viable solution for local governance and cloud security needs.

2. Research Methods

In this study, we assess the impact of digital transformation on local governance and document accessibility through the implementation of a barangay document management system in two Manila barangays: Barangay 494 in Sampaloc and Barangay 99 in Tondo. Our research incorporates a range of perspectives from residents and barangay officials, aiming to capture a comprehensive view of the system's effectiveness. Data collection was carried out using structured online questionnaires that adhere to ISO/IEC 25010 guidelines, ensuring a consistent and reliable evaluation of the system's quality and functionality.

A. Conceptual Framework

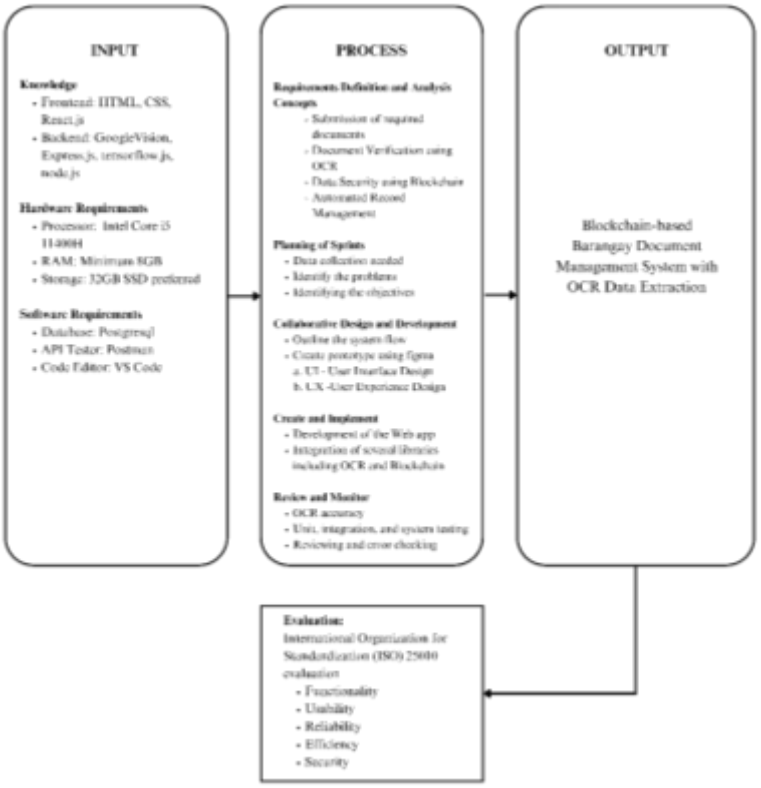

Fig 1. Conceptual Framework of the Study

Figure 1 illustrates the conceptual framework of the "Blockchain-Based Barangay Document Management System with OCR Data Extraction" using the Input-Process-Output approach. The input involves hardware requirements, with the system accessible via low-specification desktops, and software components integrating advanced web technologies and machine learning libraries within a Node.js backend, enabling data processing and user interaction through a web interface built with HTML, CSS, and React.js. The process phase includes prototype development, testing, and maintenance, while the output is the implementation of the blockchain-based system with OCR data extraction, evaluated against the ISO 25010:2011 software quality model in terms of functionality, usability, reliability, and security.

B. System Architecture

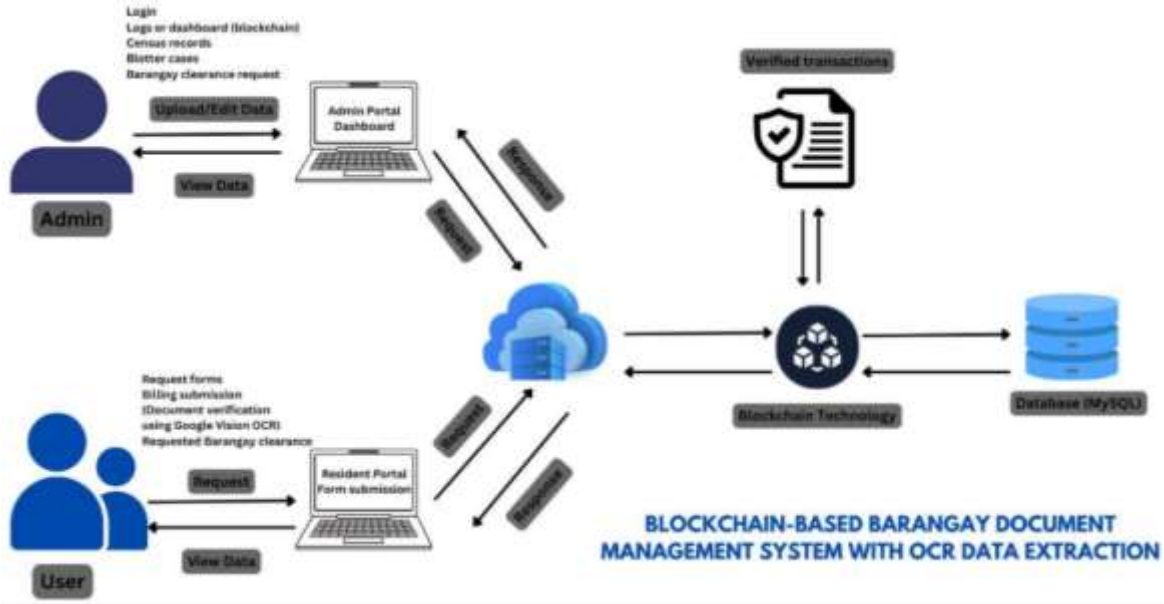

Fig 2. System Architecture of the Study

Figure 2. illustrates the Blockchain-Based Barangay Document Management System, which integrates blockchain technology with Optical Character Recognition (OCR) via Google Vision OCR. This system includes features for both admins and residents of the barangay 494 zone 49 in Sampaloc, Manila and barangay 99 in Tondo, Manila. Admins can monitor data and manage requests for barangay clearances, while residents can upload or edit their data, view personal information, and request clearances. The web server is developed using Node.js, React.js, HTML, CSS, and JavaScript within the VSCode environment. It serves as the central hub for communication, interfacing with a MySQL database for data storage and retrieval. Additionally, the system integrates blockchain technology for ensuring data integrity and immutability. The server handles requests from user interfaces, manages data processing, and ensures both efficient data management and verification through blockchain.

2.1 Sampling

The sample for this study comprises 396 participants, with 337 residents and 59 barangay officials from Barangay 494 and Barangay 99. To ensure a well-rounded representation, stratified sampling was used, considering important demographic factors such as age, gender, and roles within the barangay. This sampling strategy captures diverse perspectives, covering both administrative personnel involved in document processing and residents who seek services like barangay clearances or file blotter cases.

2.2 Data Collection

Data collection for this study was conducted through structured online surveys targeting both barangay officials and residents. The surveys, designed to be concise yet comprehensive, focused on key dimensions such as system functionality, efficiency, usability, reliability, and security, in line with ISO/IEC 25010 standards. Separate versions were tailored for residents and officials, with the resident survey examining their experiences with requesting clearances and filing reports, and the official survey assessing the system's usability for managing data and ensuring record accuracy. The study follows the ISO/IEC 25010 framework to evaluate the system's effectiveness, using a four-point Likert

scale to assess aspects like functional suitability, performance efficiency, usability, reliability, and security. A descriptive table summarizes the ratings for each dimension, offering insights into user satisfaction and system performance.

To assess the effectiveness of the implemented Optical Character Recognition (OCR) solutions, a comparison is made between Google Vision OCR and Tesseract. The evaluation uses word count as a measure of accuracy, with results showing that Google Vision OCR outperforms Tesseract in terms of precision. Additionally, a blockchain-based system was developed to provide secure data encryption for processing barangay clearance requests and blotter report submissions. The system ensures that data submitted through these requests is encrypted and stored using hashing techniques, safeguarding sensitive information while maintaining a secure and accessible framework for managing critical data.

Fig 3. Comparison of Google Cloud Vision to Tesseract with ideal lighting

Figure 3 shows a comparison between Google Cloud Vision and Tesseract under ideal lighting conditions, where Google Cloud Vision achieved a high accuracy with a word count of 19, while Tesseract reached only 7 words.

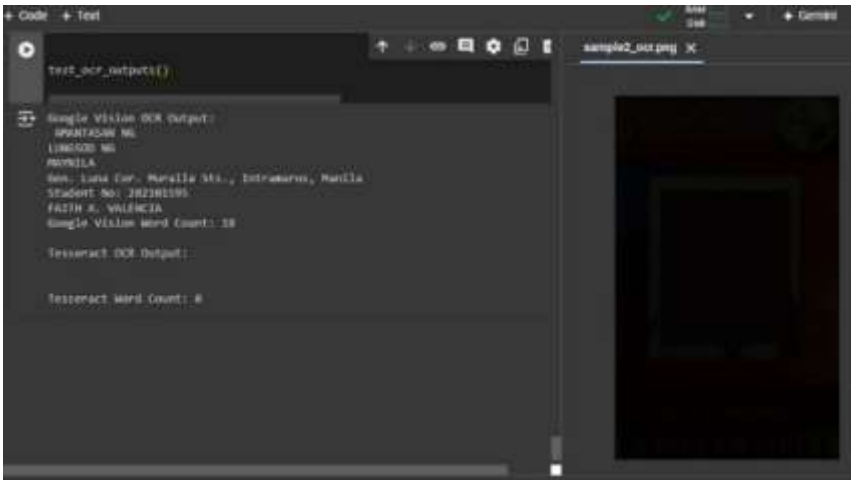

Fig 4. Comparison of Google Cloud Vision to Tesseract with low lighting

Figure 4 illustrates the comparison between Google Cloud Vision and Tesseract in low-light conditions. Google Cloud Vision managed an 18-word count, whereas Tesseract detected no words.

root_hash	when
183a4018e38e2f255b102376f58efb519778f08...	2024-10-19 18:48:38
0f1101221c93d4e168f6bdc8c4a270dac9a4ad7...	2024-10-19 18:59:52
c4608332098a7d8e557c564f96c7122fcc74dfef...	2024-10-19 19:01:27
921d997e2869429a927066a5636228e44875bd...	2024-10-19 19:22:14
9a522d3b8c86ceaca63b77584198664abfacebc...	2024-10-19 19:24:22
273fe70d4866dfc3b9aae1e99bd9353f18d6f55...	2024-10-19 19:25:18
569374d755d4de1c1124cceb5250ddff1e57856...	2024-10-19 19:29:53
949cb21a657fca2c1e8de70149e54249ad5a77d...	2024-10-19 21:22:56
537227c55ab799bab1f7e963a7e62ee5cc802f7...	2024-10-19 21:26:25
6d8a7440db63dfcd9a7ee42647deab2f3c15ea3...	2024-10-19 21:29:30
5994fe7f62987ba4325aa0a3a5dbfc34e68bdc0...	2024-10-19 21:30:35
c99ea10aa6cba0d9f159f35a4788653bf191704...	2024-10-19 21:31:46
75b07fcceb9acd1cccc1e2298cc9ab25f29308be...	2024-10-22 11:05:13
57a63e12c36ce0bb389fb4c6cf7132e5a47c6ee...	2024-10-22 15:31:27
fb2d6bb42b9ee1e6974fded4d51cb626449a5bd...	2024-10-22 16:28:02

Fig 5. Example of the requestor's root hash value

Figure 5 displays the root hash, or Merkle root, of the requestor. In a blockchain structure, the Merkle root represents the combined hash of all individual data entries below it. This root acts as a unique, consolidated identifier that securely reflects the integrity of the entire dataset. By summarizing each individual hash into a single output, the Merkle root allows for efficient verification of large data sets, where any alteration in the data will lead to a change in this root, ensuring data consistency and security.

2.3 Measures

The system uses the AES-256 for encryption and to get the value of randomness from IV which is a random number generator that is combined with the plaintext using XOR to get a random value so the formula would be for the initial step would be “Intermediate Block 1=Plaintext Block 1⊕IV”. The system would then the formula “Ciphertext Block 1 = AES (Intermediate Block n, Key)” which would give the full encryption for the first round, for the second and the rest of the other rounds we then replace the IV with the Ciphertext block making the formula look like this “Intermediate Block n = Plaintext Block n ⊕ Ciphertext Block (n-1)”. Each block is determined by the division of 16 bytes regardless of the content type whether it is a number, word, letter or character, or even a sentence it will still be divided into 16 bytes so one block could contain a part of a word it could be a full sentence. 16 bytes is where the AES operates specifically in a 4x4 finite field fully maximizing its ability to encrypt the data by following the procedure of shifting rows, mixing columns, substituting values, and adding the round key that are applied in each encryption round to scramble the data and create a secure ciphertext.

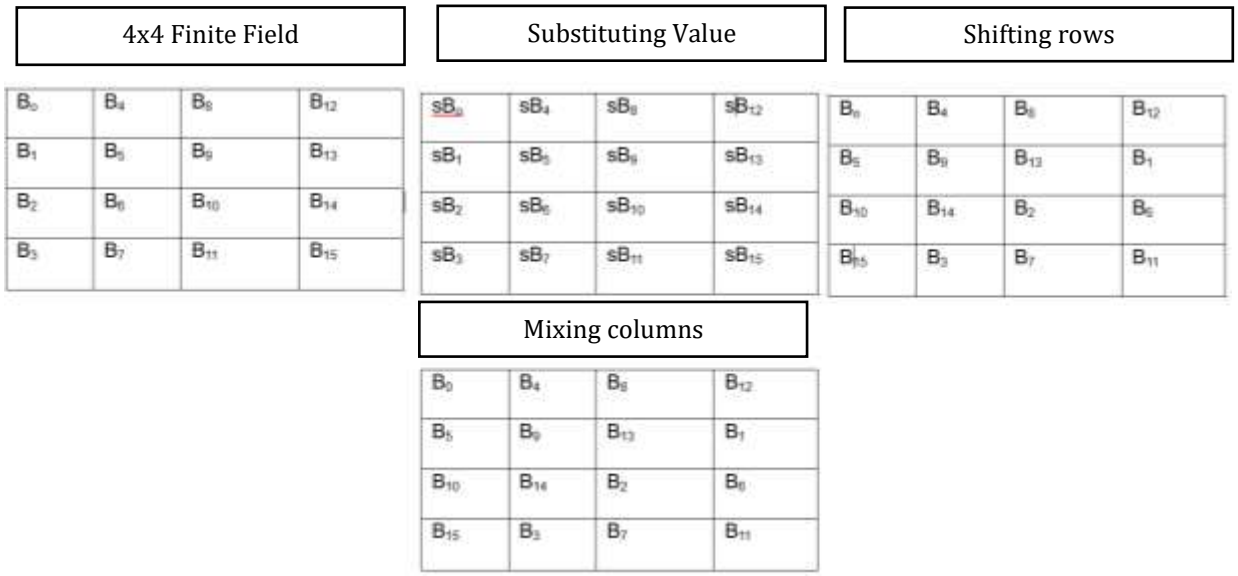

Fig. 6 AES Process Diagram

The SHA-256 hashing algorithm ensures data integrity and security by creating a unique 256-bit hash based on encrypted input values. Initially, eight constants derived from prime numbers are used to initialize the hash, and the data is divided into 512-bit blocks, with padding applied if necessary. Each block undergoes 64 rounds of transformations, where working variables are updated using logical operations, modular additions, and constants. Functions like "choice" and "majority" help thoroughly scramble the data. The hash value for each block is combined, and these values are added together to form the final hash, or Merkle root. The resulting hash changes completely if the data is altered, making it a reliable method for verifying data integrity and ensuring that the encrypted information remains unmodified.

3. Results and Discussion

The mean satisfaction ratings for the system were consistently above 3.5 out of 4, indicating strong support for its ability to reduce manual processes. These results suggest that the system effectively meets core functional requirements, improving accessibility for residents and administrative efficiency for barangay officials. In terms of performance and usability, the system was highly rated for its efficiency in managing large volumes of data and handling multiple user requests simultaneously. Respondents praised the system's speed, reliability, and its capacity to provide essential reports. Usability scores were also favorable, with high ratings for the clarity of the interface, ease of access, and alignment with user needs. The system's design was intuitive, ensuring smooth interactions for both residents requesting documents and officials handling administrative tasks. Additionally, the system's reliability was demonstrated through its high uptime and quick recovery from minor outages, confirming its suitability for daily operations. Testing and validation processes showed that the OCR and blockchain components were effective. Google Cloud Vision OCR was preferred for its higher accuracy in text recognition, particularly in low-light conditions, surpassing Tesseract OCR. Blockchain testing verified the system's ability to maintain data integrity through secure, tamper-proof hash generation and validation. The security features, including user-specific access controls, were well received, with respondents expressing confidence in the system's ability to protect sensitive data.

Table 1. Overall Mean Average Across all Criterias

Criteria	Mean	Verbal interpretation
Functional suitability	3.62	Satisfied
Performance efficiency	3.58	Satisfied
Usability	3.61	Satisfied
Reliability	3.57	Satisfied
security	3.6	Satisfied

4. Conclusions

In alignment with the research's objectives and evaluations, the following conclusions were drawn:

1. The developers successfully designed a user-friendly web-based portal system for barangay document management system. The developers successfully integrated a comprehensive dashboard that displays detailed statistical records and reports. This includes the census records, list of blotter reports, list of blotter case log, barangay clearance number of requesters by purpose and summary of barangay clearance requests. The dashboard effectively tracks and monitors this information by presenting totals, lists, and charts categorized for easy access and accurate reporting.
2. The developers successfully created a blockchain-based barangay document management system, ensuring secure data encryption for processing both barangay clearance requests and blotter report cases.
3. The developers integrated Optical Character Recognition (OCR) technology using Google Vision's API for data extraction and verification, allowing the system to accurately validate barangay resident information from forms and uploaded billing documents. With an accuracy rate of 96% to 98%, the OCR feature has greatly optimized the manual processes for barangay clearance requests and blotter case reports, minimizing human error and enhancing overall efficiency.
4. The research survey was distributed online and garnered a total of 396 respondents, including 337 barangay residents and 59 barangay officials. While this sample provided valuable insights, a more diverse representation of residents and officials from Barangay 494 and Barangay 99 could yield more generalized results and enhance the reliability of the findings.
5. The system's functional suitability, performance efficiency, usability, reliability, and security were evaluated according to ISO/IEC 25010 standards. In each category, the majority of respondents, consisting of barangay residents and barangay officials, rated the system as "Very Satisfied".

In alignment with the research evaluations, the following recommendations to further improve the Blockchain-Based Barangay Document Management System with OCR Data Extraction were drawn:

1. Enhance user training and support for barangay officials and residents to effectively utilize the system's features, including OCR data extraction.
2. Integrate Robotic Process Automation (RPA) to automate repetitive tasks like data entry and document routing, improving efficiency and minimizing human error.
3. Enhance blockchain security through the implementation of multi-signature authentication for sensitive transactions, while also addressing the scalability and decentralization of various barangays.
4. Develop a feedback mechanism within the system, using chatbots to enable users to report issues and suggest improvements, fostering continuous enhancement based on user experience.

5. References

Bautista, M. T. T., Bolonos, A. K. B., Camba, U. J., De Guzman, L. M., Goh, M. L. I., Raguro, M. C. F., & Tolentino, L. S. (2023b). BALANGAY: A Web-Based Document Request and Incident Reporting System with Decision Support for Barangay Program Development.

Carpio, C. O. (2020). Barangay management system. *International Journal of Multidisciplinary Research and Publications (IJMRAP)*, *3*(2), 26-32.

Clavin, J., Duan, S., Zhang, H., Janeja, V. P., Joshi, K. P., Yesha, Y., ... & Li, J. D. (2020). Blockchains for government: use cases and challenges. Digital Government: Research and Practice, 1(3), 1-21.

Gangadharaiah, S., & Shrinivasacharya, P. (2024c). Secure and efficient public auditing system of user data using hybrid AES-ECC crypto system with Merkle hash tree in blockchain. *Multimedia Tools and Applications*. https://doi.org/10.1007/s11042-024-18363-0

Jayaprakash, J. S., Balasubramanian, K., Sulaiman, R., Hasan, M. K., Parameshachari, B. D., & Iwendi, C. (2022). Cloud data encryption and authentication based on enhanced Merkle hash tree method. Computers, Materials & Continua, 72(1).

Lackner, A., Mirhosseini, S. A. M., & Craß, S. (2022, August). Securing File System Integrity and Version History Via Directory Merkle Trees and Blockchains. In *International Conference on Database and Expert Systems Applications* (pp. 294-304). Cham: Springer International Publishing.

Li, W., Feng, Y., Liu, N., Li, Y., Fu, X., & Yu, Y. (2024). A secure and efficient log storage and query framework based on blockchain. *Computer Networks*, *252*, 110683.

M. Balilia, C., A. Acoba, J., I. Pagulayan, L., C. Mebaña, J., & Dela Cruz, L. (2022). Windows-Based Information Management System with Project Monitoring for Barangay (Vol. 2). (EnRB) Engineering Research Bulletin. http://119.92.172.179/journals/beai/papers/vol2/vol%202%20series%202022-69-78.pd

Sathya, A. R., & Banik, B. G. (2020). A comprehensive study of blockchain services: future of cryptography. *International Journal of Advanced Computer Science and Applications*, *11*(10).

Sharma, D. M., Shandilya, S. K., & Satapathy, S. C. (2023). Maximizing blockchain security: Merkle tree hash values generated through advanced vectorized elliptic curve cryptography mechanisms. Concurrency and Computation Practice and Experience, 35(23). https://doi.org/10.1002/cpe.7829