International Journal of Engineering Technology Research Management
(IJETRM)
https://ijetrm.com/

ARTIFICIAL INTELLIGENCE POWERED GOVERNMENT-ISSUED DOCUMENT VERIFICATION SYSTEM IN UNIVERSIDAD DE MANILA USING RANDOM FOREST

Ronald B. Fernandez
ORCID: 0009-0007-0979-6315
Professor, College of Computing Studies, Universidad De Manila, Philippines

Marcus Riley A. Faustino
ORCID: 0009-0000-2910-7053
Undergraduate Student, College of Computing Studies, Universidad De Manila, Philippines

Naomi Trinity Aribuabo
ORCID: 0009-0001-8921-1052
Undergraduate Student, College of Computing Studies, Universidad De Manila, Philippines

Jaybert Brian D. Castillo
ORCID: 0009-0004-7249-8642
Undergraduate Student, College of Computing Studies, Universidad De Manila, Philippines

Jhenina B. Cadano
ORCID: 0009-0003-5399-4918
Undergraduate Student, College of Computing Studies, Universidad De Manila, Philippines

ABSTRACT

The rising number of student enrollments in institutions of higher learning has made it even more difficult to check official government documents. Manual checks are time-consuming, prone to error and susceptible to sophisticated forgery solutions, undermining institutional integrity and efficacy. The paper, entitled Artificial Intelligence-Powered Government Document Verification System in Universidad de Manila Using Random Forest, offers an approach to the process of document authentication that is automated by employing Optical Character Recognition (OCR) and machine learning. OCR was used to identify important characteristics in scanned government-provided documents like birth certificates, voter certificates, and licenses, and was particularly concerned with such issues as text alignment, font consistency, presence of watermarks, and signature verification. The features were then categorized with the help of the Random Forest algorithm that has proven to be more accurate, scalable and resistant to overfitting. If a document was identified as a suspicious one, it was reviewed manually, which guaranteed an extra security measure. The Agile Software Development Life Cycle had been taken up in the study, and, as a result, the requirements gathering, design, development, testing, and deployment stages were conducted in an iterative manner. The reliability, usability, and effectiveness of the system were tested in terms of functional, integration, performance, and user acceptance. Findings showed that processing speed, accuracy and fraud detection had improved significantly when compared to traditional manual practices. In the case of the Universidad de Manila, the system saves the administrative load, security in the authenticity of the student records and prevention of fraudulent admission. In addition to institutional advantages, the study will add to the continuous research and technical innovations in the field of document verification with AI. All in all, the proposed system offers a scalable, cost effective and smart solution which reinforces academic integrity, builds on institutional trust and operational efficiency in document management.

International Journal of Engineering Technology Research Management
(IJETRM)
https://ijetrm.com/

Keywords:
Artificial Intelligence (AI), Document Verification System, Random Forest Algorithm, Optical Character Recognition (OCR), Fraud Detection, Machine Learning, Government-Issued Documents, Automation in Verification.

INTRODUCTION

The integrity of documents has become a key issue to organizations that operate within the modern digital environment and require documents that are authentic as a way of protecting transparency, trust, and integrity. The rapid development of technology has eased the development and alteration of documents, but it has also created additional ways of forgery and counterfeiting. As fraud keeps on advancing, they indeed present significant threats to society, the corporate sector, research and learning institutions that rely on the credibility of the presented documents to accept admissions, secure employment, and establish their identity. The growing complexity of document falsification requires correspondingly sophisticated detection and validation methods, especially by using artificial intelligence (AI) and machine-learning technologies.

Machine Learning, paired with Artificial Intelligence and Optical Character Recognition, presents a new approach to the task of counteracting document fraud by automating the processes of identification and verification. OCR enables the segmenting and coding of text and images, and the machine-learning algorithms, namely the Random Forest, trains on the patterns that are present in legitimate and forged documents and produces accurate classifications. Random Forest algorithm is a model of ensemble (many decision-trees), which enhances verification, analyzing such features like font type, text alignment, watermark presence, and signatures. This multi-layered method additionally improves the detection accuracy, as well as reduces the possibility of human error, producing a quicker and more reliable verification procedure.

The current research aims to build an AI-based document verification system in the case of the Universidad de Manila by using the random forest algorithm to automate and supplement the verification of academic and institutional documents. The system will be designed to work through a secure local server, with an interface based on Flutter to enable accessibility across platform. The system will have the potential to classify uploaded documents efficiently using OCR, AI, and Machine Learning to consider them either as valid or suspicious files and mark the suspicious files that need validity verification. With this innovation, the university will be able to simplify the administrative operations, minimize the risk of fraudulent submissions, and become overall more secure as an institution, thus adding to a more efficient and more technology-advanced academic setting.

OBJECTIVES

The main goal of the research is to plan and design an AI-based document verifying system, which uses the Random Forest machine learning algorithm, to improve the detection and check of forged student's government-issued documents at the Universidad de Manila. This is to enhance the accuracy of academic records and official university processes, reliability as well as the security of student's official records.

Also, the objectives are the following:

1) To develop and establish an AI-based verification system that would be able to analyze and confirm the authenticity of government documents offered by the students.
2) To use the Random Forest algorithm in classifying and detecting fraudulent or manipulated documents using learned training samples of genuine and fake samples.
3) To incorporate the developed system as a part of the current process of managing documents in the Universidad de Manila to automatize and simplify the verification process.
4) To compare the system performance with relation to accuracy, reliability, processing time and user satisfaction with the traditional manual verification techniques.
5) To have a safe and scalable structure that would allow the further development of the document verification process in other departments or organizations.

Related Studies

This study by [1] Surmieda (2022) developed an OCR-enhanced digital asset management system aimed at digitizing and organizing school documents. The system improved accessibility and minimized manual encoding errors, emphasizing how OCR technology can streamline document processing within academic institutions. This research demonstrates the efficiency of OCR in managing educational records and supports the

iJETRM
International Journal of Engineering Technology Research Management
(IJETRM)
https://ijetrm.com/

potential of AI-assisted document verification in universities.
This work by [2] Buctuanon et al. (2021) introduced "Docudile," a hybrid rule-based and OCR-supported document structure classification system designed for academic use. The system effectively categorized document layouts and formats, though it encountered limitations in handling irregularly formatted files. The work highlights the value of combining OCR and machine learning for document analysis and classification, contributing to more reliable academic record management.
This study by [3] Pino et al. (2022) focused on developing a block-level OCR system using Support Vector Machine (SVM) to transliterate Baybayin texts into the Latin alphabet. By integrating OCR with machine learning, the system achieved high recognition accuracy, showing the feasibility of AI-driven models in processing complex document scripts. This contributes to the foundation of document verification technologies that rely on feature recognition and pattern consistency.
This study by [4] Distor et al. (2021) examined the acceptance and use of Artificial Intelligence in local governance within the Municipality of Carmona, Cavite. Using frameworks from the Technology Acceptance Model and UTAUT, the researchers found that local officials displayed positive attitudes and behavioral intentions toward AI adoption. Their findings emphasize how AI can enhance efficiency and data security in public operations, paralleling its potential applications in academic document management.
This study by [5] Ranera (2024) presented a comparative analysis of natural language processing (NLP) embedding techniques on Philippine Supreme Court case decisions. The researcher compared embedding models like Doc2Vec and TF-IDF to improve document retrieval and classification accuracy. This research demonstrates how machine learning and text analytics can enhance document-based systems, serving as a local precedent for AI-driven verification models in academic institutions.

This study by [6] Seuret et al. (2020) proposed a forensic document examination system that used ensemble learning techniques, including Random Forest, to detect forged documents. Their findings showed that ensemble methods effectively identified variations in ink texture, text structure, and font arrangements. This study reinforces the Random Forest algorithm's reliability in distinguishing authentic from counterfeit documents, aligning closely with the focus of this research.

This study by [7] Patil and Maheshwari (2021) compared machine learning algorithms—Random Forest, Support Vector Machine, and K-Nearest Neighbors—for online signature verification. Results indicated that Random Forest achieved higher precision and consistency, particularly in detecting forgeries under varied handwriting patterns. This emphasizes Random Forest's strength in authentication tasks, validating its use in document verification systems.
This study by [8] Wang et al. (2022) applied deep learning-based OCR to extract and verify text from complex regulatory documents. By combining convolutional neural networks (CNN) with OCR, the model achieved high precision even with inconsistent layouts and image noise. This demonstrates the scalability of combining OCR and AI for advanced document recognition systems.

This study by [9] Kuri et al. (2025) presented a legal document authentication system integrating OCR, Natural Language Processing, and Convolutional Neural Networks. The system examined both textual and visual document features to confirm authenticity, achieving high detection accuracy for falsified papers. Their research highlights the potential of combining multiple AI techniques for robust verification, closely paralleling the aim of the current study.

This study by [10] Lee et al. (2023) explored forensic document paper identification using hybrid feature extraction and machine learning classification. The researchers achieved high accuracy by combining texture and pattern-based features with AI classifiers, enabling precise differentiation between genuine and counterfeit documents. This work emphasizes the effectiveness of machine learning in forensic document authentication, reinforcing its relevance to the present study.

METHODOLOGY

The research design used in this paper was a developmental research design where a Document Verification System with AI is developed to verify government-issued documents in academic use at the Universidad de Manila. The system is developed with the help of machine learning, in particular, the Random Forest algorithm,

iJETRM
International Journal of Engineering Technology Research Management
(IJETRM)
https://ijetrm.com/

to classify and check documents with the features identified with the help of Optical Character Recognition (OCR) and image analysis. OCR technology was used to extract textual and visual features like font style, text layout, watermark, and signatures that were then processed by the Random Forest model to identify authenticity of a document. This approach has provided a dependable, speedy and automated verification system that mitigates human error, manual overload and strengthens institutional security against fraudulent documents.

System Development Model

Figure 1: Agile SDLC of Document Verification System

The proponents chose the Agile Methodology because of its iterative cycles and adaptability to user feedback, making it more convenient for a machine learning system since it has a dynamic process of development. This method of approach aligned with the continuous improvement of the identification system.

1) **Planning**
   Talk to the university registrar and or school faculty on their document verification challenges, the required features of the document verification, for instance, user authentication, document upload methods, and verification processes, etc. development
2) **Design**
   The extraction of the characters from the image by Optical Character Recognition (OCR) and then use of the Random Forest algorithm to verify documents. The design should be also easy to use and should comply with the user requirements.
3) **Development**
   The development and implementation of the user interfaces, backend structure, and integration of the artificial intelligence model were carried out through iterative development cycles. This approach allowed the system's features to be built, tested, and refined in incremental stages, enabling continuous improvement and timely modification based on performance and functionality outcomes.
4) **Testing**
   The continuous testing while developing the system and it involved unit testing, integration testing and user acceptance testing. These guaranteed the smooth functionality between the different elements, the functionality was satisfactory to the user and that the entire system produced correct and consistent verification outcomes.
5) **Deployment**
   Once functional, the system was deployed for user application. The implementation enabled the university registrar and faculty to embark on the utilization of the document verification features in practice. New updates and enhancements were released at the end of every sprint and provided users with finalized functionalities and improved performance with time.
6) **Evaluation and Refinement**
   After deployment, the results of every sprint were discussed with the stakeholders. Faculty and registrar feedback were made, and the features were optimized to enhance the usability and strengthen accuracy. The lessons that were acquiring during each of these cycles were used to make the adjustments toward the further improvements and the system became more efficient and responsive to the needs of the institution.

iJETRM
International Journal of Engineering Technology Research Management
(IJETRM)
https://ijetrm.com/

Data Gathering Techniques
Primary Data Collection

1. **Face to Face Interviews:** Conducted with the university's registrar and school faculty to assess document verification and authentication challenges through face-to-face and online mode

2. **Online Survey:** Conducted with the university's registrar and school faculty to assess document verification and authentication challenges through emails, messages, and online social media.
3. **Feedback:** The feedback obtained from these experts serves as an important basis for improving the accuracy and effectiveness of the proposed system.

Secondary Data Collection

1. **Review of Related Literature/Studies:** An extensive analysis of academic studies, journal articles, and research papers on the AI Powered Document Verification System was conducted. This review provided insights for the proponents into the challenges based on the literature and studies.
2. **Academic Research**: Usage of online research (e.g. Google Scholar) to assess helpful studies, articles, research, journals, etc. correlating to the study.

Conceptual Framework

The Input, Process, Output System Architectural model provides a clear and concise way to describe and represent a system's functionality. The diagram offers a perspective of how data flows through the system, transforming inputs into desired outputs. The simplicity of the diagram makes it simple to understand across technical and non-technical audiences.

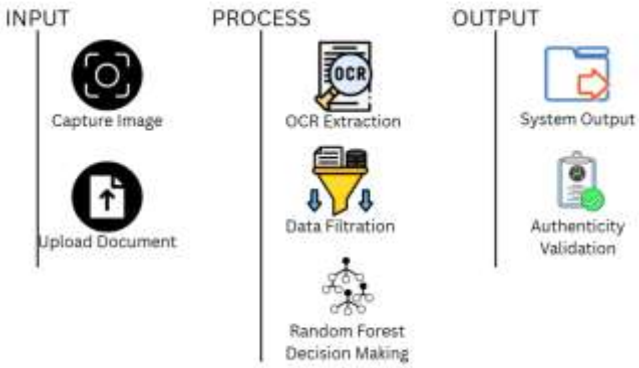

Figure 2: Conceptual Framework

Figure 2 sets out the relations between system entities that enable accurate, efficient, and secure verification of student documents in Universidad de Manila. User input, Optical Character Recognition, and the Random Forest algorithm interact within the study to provide reliable authentication results and confidence scores.

Input Phase – The input phase begins when users log in and upload scanned copies of their government-issued documents, such as birth certificates, voter's certificates, and licenses. These serve as the primary data that the study will analyze for authenticity.

Process Phase – In the process phase, Optical Character Recognition extracts key document features such as text alignment, font style, watermarks, and signatures. The Random Forest algorithm then analyzes these features to classify the document as authentic or suspicious. This ensures a systematic and automated verification procedure.

Output Phase – The output phase presents the results of the verification, showing whether the document is authentic or suspicious. A confidence score is provided for accuracy, and users are also given the option to re-upload documents if additional verification is needed.

iJETRM
International Journal of Engineering Technology Research Management
(IJETRM)
https://ijetrm.com/

Prototype

Figure 3: Login & Signup Screen

The system starts with the Log In & Sign-Up screen, where users either enter their credentials to access their accounts or register as new users. This step ensures that only authorized individuals can proceed with the verification process. Upon successful authentication, users are redirected to the homepage.

Figure 4: Home Screen

The Homepage serves as the main dashboard where users can start the document verification process. They are given options to either upload a document from their device's storage or take a new photo using their camera. This interface is designed for ease of navigation, allowing users to move forward quickly.

iJETRM
International Journal of Engineering Technology Research Management
(IJETRM)
https://ijetrm.com/

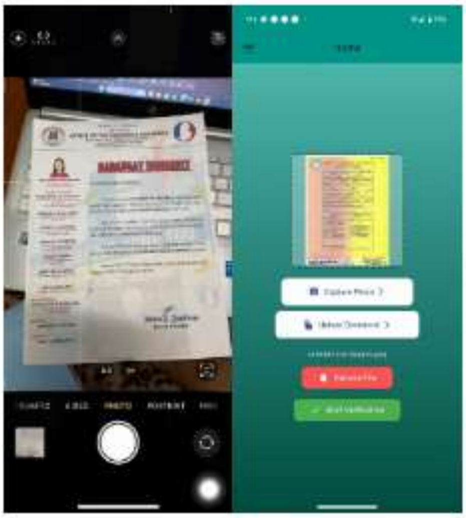

Figure 5: Gallery Selection and Camera

In the Gallery Selection and Camera frame, users browse their photo gallery and select a document or captures document using their camera in which they wish to verify. This step ensures that users can conveniently upload previously saved images instead of taking a new one. Once a document is selected, they proceed to the upload screen.

Figure 6: Side Bar Light Mode

After processing, users are directed to the Result Screen, where they receive the final verification outcome. The system displays a feedback ("Document is Authentic"), indicating the accuracy of the authentication process. If the document is successfully verified, users may choose to download a verification report. If verification fails or the document is flagged as suspicious, users may be given options to reattempt verification or contact support for further assistance.

iJETRM
International Journal of Engineering Technology Research Management
(IJETRM)
https://ijetrm.com/

RESULTS AND DISCUSSION

This section summarizes and discusses the evaluation results obtained from the ISO/IEC 25010-based questionnaire distributed to chosen respondents. The study focused on the AI Powered Document Verification System's major quality features, such as functional suitability, performance efficiency, compatibility, usability, reliability, security, maintainability, and portability. Each sub-criterion was scored on a five-point Likert scale, allowing users to identify both strengths and weaknesses based on their perceptions and interactions with the system.

The received data were processed following the formula of descriptive statistics, namely, means, and then presented in the form of graphical representations to clarify the performance of the system in relation to the research goals. The survey was conducted on 70 respondents who were selected purposely to provide significant feedback that constituted an empirical base on the evaluation of the system functionality, usability, and overall effectiveness.

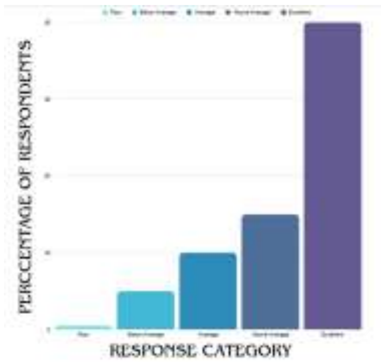

Figure 7: Bar Graph for Functional Suitability

The system scored an average mean of 4.78 (Excellent) on the Functional Suitability. The answer to this question was strongly affirmative with the respondents of the opinion that the system is doing its job accurately, automating the verification of official documents like birth certificates, voter IDs and licenses. They emphasized that the system is efficient in identifying major features such as text alignment, font consistency, watermarks, and digital signatures with the help of Optical Character Recognition (OCR) that is capable of giving accurate document authenticity results.

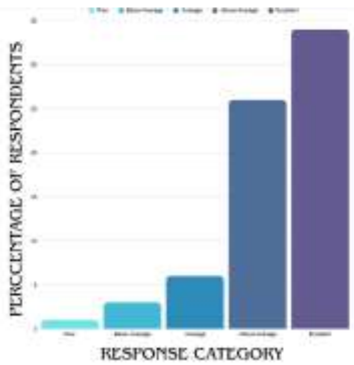

Figure 8: Bar Graph for Performance Efficiency

The system received a General Weighted Mean of 4.40 (Above Average) in Performance Efficiency. The respondents also indicated that the system is fast in processing and classifies scanned documents without any hesitation or crashing of the system and based on the Random Forest algorithm. It did not lose its stability in performance and real-time response even when dealing with several verification requests.

iJETRM
International Journal of Engineering Technology Research Management
(IJETRM)
https://ijetrm.com/

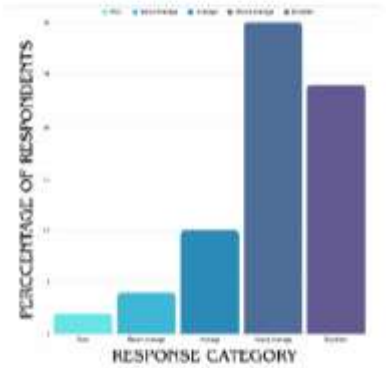

Figure 9: Bar Graph for Compatibility

Compatibility found a General Weighted Mean of 4.12(Above Average). The respondents affirmed that the application works reliably across the desktop and mobile platforms and integrates well with scanning and document input devices. Various document types (PDF, JPG, PNG) can be processed by the system without compatibility problems.

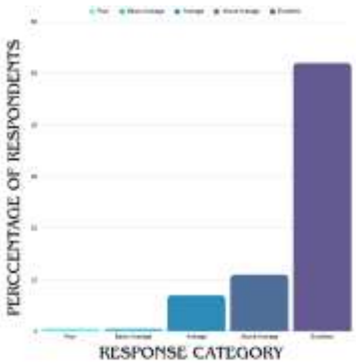

Figure 10: Bar Graph for Usability

Usability had an average mark of 4.50 (Excellent) against the system. The respondents were able to agree that the interface is easy to navigate, intuitive and visually structured. The verification process comes with less user training and the layout of the verification is user friendly as the verification results are clearly shown once the users scan, process and view the verification results.

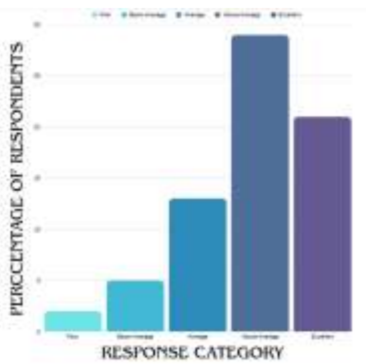

Figure 11: Bar Graph for Reliability

iJETRM
International Journal of Engineering Technology Research Management
(IJETRM)
https://ijetrm.com/

The Reliability aspect had a General weighted mean of 4.00 (Above average). The respondents were very confident that the system could do continuous verification without any mishaps and loss of information. The system is effective in preserving data integrity and maintaining verification logs and results of the classification as always accurate.

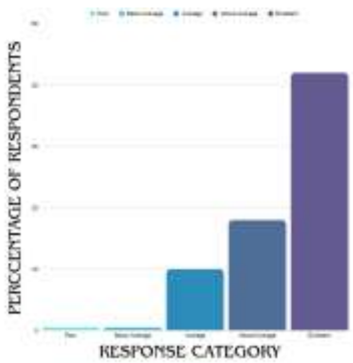

Figure 12: Bar Graph for Security

Security had the best mean of 4.37 (Above Average). Respondents commended the secure authentication system and good security of sensitive document information in the system. The platform is encrypted in the transmission and access control, and only authorized staff can give or access the document results.

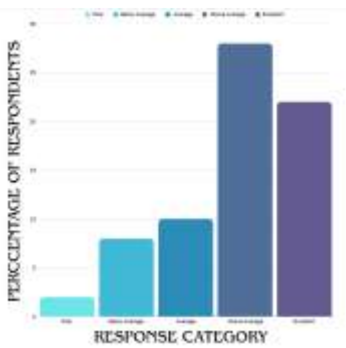

Figure 13: Bar Graph for Maintainability

In case of Maintainability, the system received a General Weighted Mean of 4.25 (Above Average). The respondents showed that the modular design of the system facilitates an easy updating, debugging and subsequent improvement. Minimal effort can be used to accomplish the tasks of maintenance, adding new document types, or changing the OCR settings

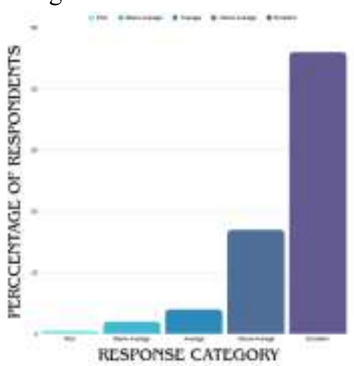

Figure 14: Bar Graph for Portability

iJETRM
International Journal of Engineering Technology Research Management
(IJETRM)
https://ijetrm.com/

The system scored 4.59 (Excellent) on General Weighted Mean on Portability. Respondents acknowledged that such software is easy to install and it can be properly operated on different devices and operating systems. The system also is compatible with a wide range of hardware configurations without performance or functionality loss.

Figure 15: Overall Summary Survey

Altogether, the survey outcomes indicate that the presented system is effective in serving the users and fitting their expectations as it offers a functional, portable, and secure system to check the government-issued documents and even denotes the opportunities to optimize the work further in terms of performance and compatibility.

ACKNOWLEDGEMENT

We the proponents, we would like to acknowledge that we owe our Capstone Adviser, Dr. Ronald B. Fernandez, the firm hand, constant support, and the mentorship commitment during our study development. His experience, support, and feedback have played a significant role in ensuring that our capstone project, which is the Artificial Intelligence-based Government-powered system of verifying documents is a successful one. We also wish to express our warmest gratitude to the faculty members, university personnel, and students of Universidad de Manila that took part in our survey and appraisals as their constructive criticism greatly contributed to the quality of our system and the building of the manuscript as well. Lastly, we would like to express our appreciation to our families and friends whom we have relied on throughout this entire project to be patient, loving and supporting and this is what helped to push on and, in the end, complete this project.

CONCLUSION

In Conclusion, the Artificial Intelligence-Powered Government-Issued Document Verification System developed by the Universidad de Manila in the current study achieved its major goal of providing a more efficient, quicker, and trustworthy system of document verification. The use of traditional manual inspection means can take long and is prone to human error and, therefore, increases the chances of signing forged or falsified documents. In its turn, the addition of Optical Character Recognition (OCR) and the Random Forest algorithm allowed the system to scan textual and visual document characteristics in a few seconds, ensuring that the verification is much faster and more precise. The automation significantly reduced the responsibilities of the university staff, as well as, enhanced the safety and integrity of academic transactions.

As a result of the evaluation conducted on seventy respondents based on the ISO 25010 quality characteristics, the data showed a high number of user satisfaction with Functionality (4.78) scoring the highest, Portability (4.59) and Usability (4.50) scoring the highest. The general average score of 4.37 that surpasses the average score is indicative of the fact that the system is functioning as expected, is reliable, and meets the demands of the expected users. These results demonstrate that machine learning and automation have a synergistic opportunity to modernize institutional procedures, and the result will be significantly better than traditional verification procedures.

IJETRM
International Journal of Engineering Technology Research Management
(IJETRM)
https://ijetrm.com/

Amidst such successes the system does have some limitations such as limited variety of document types, sensitivity of OCR to low-resolution scans, and lacks real-time links to national databases to do cross-validation. Future improvements will need to explore the integration of blockchain technology with deep learning systems and real-time API integrations, to enhance the system in terms of precision, scalability, and reliability. Altogether, the study highlights the ability of AI-based verification systems to transform the existing document authentication process to make it more effective, secure, and trustworthy both at the Universidad de Manila and beyond.

REFERENCES

[1] Surmieda, J. (2022). *OCR-enhanced digital asset management system: Prototype design and construction. Philippine Journal of Librarianship and Information Studies, 42*(2), 33–43. https://journals.uplb.edu.ph/index.php/pjlis/article/view/1348

[2] Buctuanon, M. M., et al. (2021). *Docudile: A rule-based pattern and OCR-based system for document structure classification. Mindanao Journal of Science and Technology, 19*(1), 17–39. https://www.msug.edu.ph/mjst/article/view/53

[3] Pino, R., Mendoza, M., et al. (2022). *Block-level optical character recognition system for automatic transliterations of Baybayin texts using support vector machine. Philippine Journal of Science, 151*(1), 303–315.
https://philjournalsci.dost.gov.ph/vol-151-no-1

[4] Distor, M. R., et al. (2021). *Acceptance and use of Artificial Intelligence in local governance: A study in the Municipality of Carmona, Cavite. Philippine Journal of Public Administration, 65*(2), 45–60. https://www.up.edu.ph/acceptance-and-use-of-ai-in-local-governance/

[5] Ranera, L. T. B. (2024). *A comparative analysis and evaluation of natural language processing document embedding techniques on Philippine Supreme Court case decisions. Philippine Engineering Journal, 45*(1). https://journals.upd.edu.ph/index.php/pej/article/view/3457

[6] Seuret, J.-P., Margner, V., & Heutte, L. (2020). *Forensic document examination system using boosting and bagging methodologies. Soft Computing, 24*(10), 5409–5426. https://link.springer.com/article/10.1007/s00542-020-05777-4

[7] Patil, R., & Maheshwari, A. (2021). *Comparison of SVM and Random Forest methods for online signature verification. International Journal of Intelligent Human-Computer Interaction, 12*(4), 211–220. https://content.iospress.com/articles/international-journal-of-intelligent-human-computer-interaction/ihci200223

[8] Wang, Y., Li, Z., & Xu, F. (2022). *Deep learning-based text recognition applied to complex regulatory documents.* In *Advances in Computational Collective Intelligence* (pp. 203–214). Springer. https://link.springer.com/chapter/10.1007/978-3-030-45398-1_19

[9] Kuri, M., Bokare, P., Dhuri, S., Inamdar, A., Gavande, P., & Gupta, A. (2025). *Legal document authentication and verification system leveraging OCR, NLP, and CNN for comprehensive document authentication. International Journal for Research in Applied Science and Engineering Technology (IJRASET).* https://www.ijraset.com/research-paper/legal-document-authentication-and-verification-system

[10] Lee, J., Kim, S., & Choi, D. (2023). *Identification of document paper using hybrid feature extraction. Journal of Forensic Sciences, 68*(2), 540–552. https://onlinelibrary.wiley.com/doi/10.1111/1556-4029.15099