Integration of Google ML Kit and OpenCV for Android-Based Document Digitization and Archive Management

Sugiyanto[1,*], Eka Bagus Surya[2], Achmad Wahyu Satrio Wibisono[3]
Department of Informatics Engineering, Faculty of Electrical Engineering and Information Technology, Adhi Tama Institute of Technology Surabaya

Email: [1,*]sugianto@itats.ac.id, [2]bs130408@gmail.com, [3]soratamotsu4@gmail.com

DOI: https://doi.org/10.31284/j.jtm.2026.v7i2.9264
Received July 17th 2026; Received in revised July 29th 2026; Accepted July 29th 2026; Available online July 31st 2026

Abstract

Digitalization of documents is an important necessity for supporting more efficient, practical, and easily accessible archive storage through mobile devices. However, document scanning using a smartphone camera still frequently produces suboptimal images due to lighting conditions, image-capture angle, and perspective distortion. This research develops the Android-based application Digital Portable Archive and Document Assistant (D-PANDA) by integrating the Google ML Kit Document Scanner API for automatic document detection and OpenCV for image quality enhancement, with a PHP Native backend and MariaDB database. Unlike commercial black-box solutions, D-PANDA provides full algorithmic transparency through its documented OpenCV-based post-processing pipeline, and offers a self-hosted backend alternative for organizations prioritizing data sovereignty. Testing was conducted using Black Box Testing to verify that all application functions operate according to requirements, as well as the User Experience Questionnaire Short (UEQ-S) with 15 respondents to evaluate the user experience. Black Box testing results indicate that all application features function correctly, with no functional errors detected. Meanwhile, the UEQ-S evaluation yielded scores of Pragmatic Quality = 2.232, Hedonic Quality = 2.286, and Overall = 2.259, all in the Excellent category, indicating that the application ranks among the top 10% of results according to the UEQ-S benchmark. The research findings demonstrate that D-PANDA effectively supports document digitalization and management while delivering a very positive user experience.

Keywords: *Blackbox Testing, Construction Project, Cost Control Management, ISO 9126, V-Model*

1. Introduction

Development of information technology has driven the transformation of document management processes from physical form to a more practical, efficient, and easily accessible digital archive. Document digitalization not only aims to reduce paper consumption but also enhances security, ease of search, and document storage effectiveness across various fields, such as education, government, and office administration [1]. However, document digitalization using a smartphone camera still faces various challenges, such as perspective distortion, uneven lighting, shadows, and poor image quality, which can reduce the readability of scanned documents [2], [3].

Numerous studies have addressed these problems, including through document detection techniques, perspective correction, and image quality enhancement. Document detection methods on mobile devices have enabled real-time recognition of document boundaries, making the scanning process more practical [4]. In addition, deep-learning-based document rectification techniques can correct perspective distortion and improve the quality of document digitalization results [5], [6]. In the field of image processing, edge detection and contour detection algorithms are widely used to assist in accurately identifying document boundaries. The Canny algorithm can detect object edges with high accuracy [7], whereas the Contour Detection method is used to obtain the shape of the document object based on edge detection results [8], [9]. To enhance the quality of scanned images, thresholding methods

such as Otsu and Sauvola are also commonly applied because they produce clearer and more readable document images [10], [11].

Google introduced the ML Kit Document Scanner API as a modern solution that provides automated document scanning capabilities on Android devices. This API supports document boundary detection, perspective correction, automatic cropping, and PDF document storage without requiring manual computer-vision algorithm implementation [12], [13]. Nevertheless, scanned results can still be improved by utilizing the OpenCV library as a supporting tool for image processing, resulting in optimal document quality.

In terms of data communication, Android applications typically use REST APIs as connectors between mobile applications and servers because they have a lightweight, flexible architecture that is easy to integrate with various platforms [14], [15]. In this research, the REST API is used as a communication medium between the Android application and the PHP Native-based backend for user authentication, document metadata storage, and management of scan-result histories using the MariaDB database.

Based on the above problems, this research develops the Digital Portable Archive and Document Assistant (D-PANDA), an Android-based document digitalization application that integrates the Google ML Kit Document Scanner API, OpenCV, REST API, and MariaDB to produce a more practical scanning process and to provide features for digital archive management. The application is equipped with features including registration, login, document scanning, PDF document storage, document history, and application settings, thereby helping users to perform document digitalization and management more efficiently.

In addition to ensuring that all application functions operate correctly through Black Box Testing, this research also evaluates the user experience using the User Experience Questionnaire Short (UEQ-S). The UEQ-S is a user-experience evaluation instrument that measures two primary dimensions—Pragmatic Quality and Hedonic Quality—providing insight into the level of ease of use as well as the overall user experience quality with the application [16]. Data collection was performed using Google Forms as a distribution medium for the questionnaire to 15 respondents, making the data collection process easier and more efficient [17]. The evaluation results are expected to demonstrate that the D-PANDA application not only meets functional requirements but also delivers a very positive user experience.

The novelty of this research lies in three key aspects: (1) the integration of Google ML Kit's automated document detection with an OpenCV-based post-processing pipeline that explicitly implements Otsu thresholding and gradient-based edge refinement to enhance scan quality; (2) the provision of a self-hosted backend using PHP Native and MariaDB for document metadata management, offering an alternative to proprietary cloud ecosystems; and (3) the combination of scanning and archive management into a single, cost-free Android application with full algorithmic transparency, enabling replication and further development by the research community.

2. Method

This research employs the Research and Development (R&D) methodology, which aims to develop and evaluate the Android-based Digital Portable Archive and Document Assistant (D-PANDA) application. This approach was selected because the research not only produces software but also implements and tests the developed system. The research process begins with identification of user needs, system design, Android application implementation, backend integration using REST API, and ends with functional testing and user-experience evaluation. Figure 1 shows the research flow diagram.

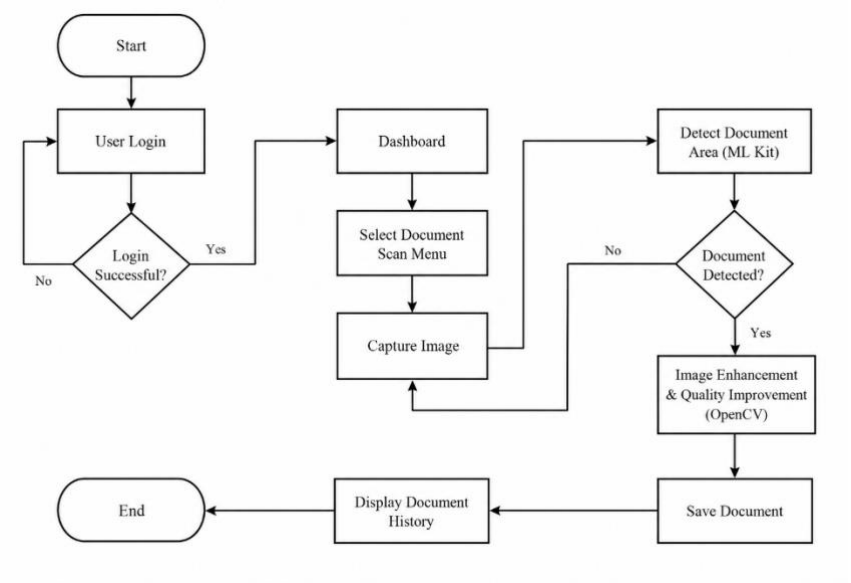

Figure 1. Research Flow Diagram

Based on the research flow diagram in Figure 1, the research process begins with user authentication via the login page. After successful authentication, the user is directed to the dashboard to select the document-scanning menu. The application then activates the device camera to capture document images, which are subsequently processed using the Google ML Kit Document Scanner API to automatically detect the document area. If the document is successfully recognized, the scanning results are reprocessed using OpenCV to enhance image quality before being saved in PDF format. Document metadata is sent via REST API to the server and stored in the database so that the document can be displayed again in the history menu.

The D-PANDA system architecture consists of two primary processes: user authentication and document digitalization. In the authentication process, the Android application communicates with the backend through a REST API built using PHP Native. Authentication validation is performed using JSON Web Token (JWT) before user data is saved to the MariaDB database. In the document digitalization process, the application uses the Google ML Kit to automatically detect the document area, whereas OpenCV is used to enhance the quality of the scanning results before the document is saved in PDF format. The overall system architecture is shown in Figure 2.

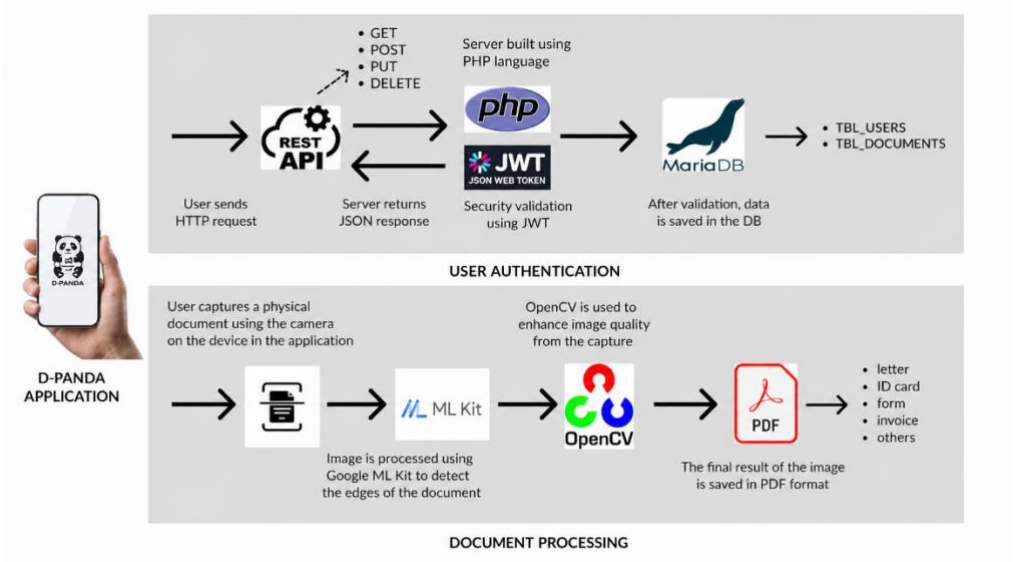

Figure 2. D-PANDA System Architecture

Based on the system architecture shown in Figure 2, the Android application serves as the primary user interface. Communication between the application and the server is performed using REST API in JSON data-exchange format. The backend is built using PHP Native, whereas MariaDB serves as the database for storing user information and document metadata. During development and testing, the local server is connected using Ngrok so that the Android application can access the backend services without requiring public hosting. The technologies used in this research are summarized in Table 1.

Table 1. Technologies Used In This Research

No	Technology	Function
1	Android Studio	Android application development environment
2	Java	Android application programming language
3	Gradle	Android application build system
4	PHP Native	Application backend
5	REST API	Communication between Android application and server
6	MariaDB	User and document database
7	Google ML Kit	Document detection and automatic scanning
8	OpenCV	Document image quality enhancement
9	JSON/JWT	User session authentication and security
10	Ngrok	Local server access during development process

The image-processing stages in this research include edge detection, document contour identification (contour detection), perspective correction, and thresholding to enhance the quality of the digitalization results. Although document detection is performed automatically by the Google ML Kit, the underlying concept of image processing still follows the gradient computation process of the image [9] as the basis for identifying the boundaries of the document object. The magnitude of the gradient is calculated using Equation (1). The gradient values are used to identify pixel intensity changes that indicate the presence of document edges.

$$G = \sqrt{G_x^2 + G_y^2} \quad (1)$$

where:
G = gradient magnitude
G_x = horizontal direction gradient
G_γ = vertical direction gradient

After edge detection, the image is processed using the Otsu Thresholding method [10] to automatically obtain the threshold value so that the document object can be separated from the background. The Otsu equation is shown in Equation (2). The thresholding method is used to increase the contrast of the image so that the digitalized document becomes clearer, more readable, and has better visual quality before being saved in PDF format.

$$\sigma_b^{2(t)} = \omega_{0(t)}\omega_{1(t)}^2\left(\mu_{0(t)} - \mu_{1(t)}\right) \quad (2)$$

where:
$\sigma^2_\beta(t)$ = between-class variance
$\omega_0(t)$ = probability of the background class
$\omega_1(t)$ = probability of the object class
$\mu_0(t)$ = average intensity of the background class
$\mu_1(t)$ = average intensity of the object class
t = threshold value

System testing is performed using two methods: Black Box Testing and the User Experience Questionnaire Short (UEQ-S). Black Box Testing is used to verify that all application functions operate

according to the functional requirements, regardless of the program code structure. The tests cover the registration, login, document scanning, document storage, history, and application settings features. Each feature is considered successful when the system output matches the expected result.

In addition to functional testing, this research evaluates user experience using the User Experience Questionnaire Short (UEQ-S). The UEQ-S instrument consists of eight attribute pairs that measure two main dimensions: Pragmatic Quality and Hedonic Quality. Data collection is performed through a Google Form sent to respondents who have already used the D-PANDA application. The data-collection flow is shown in Figure 3.

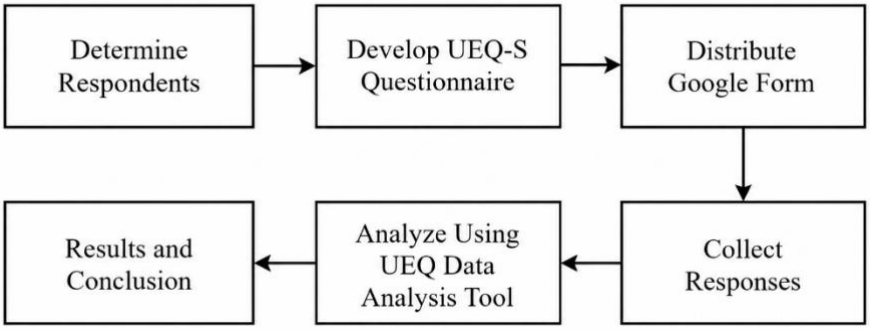

Figure 3. Data Collection Flow Diagram

The number of respondents is determined using Slovin's formula [18] as shown in Equation (3).

$$\boldsymbol{n} = \frac{\boldsymbol{N}}{\boldsymbol{1+Ne^2}} \tag{3}$$

where:
n = sample size
N = population size
e = error rate

With a population of 17 individuals and an error rate of 5%, the sample size is calculated as follows:

$$n = \frac{17}{1 + 17(0.05)^2} \approx 15$$

Thus, 15 respondents are used in this research. The user-experience value is calculated using the mean value from all respondent answers as shown in Equation (4).

$$\bar{\boldsymbol{x}} = \frac{\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \boldsymbol{x_i}}{\boldsymbol{n}} \tag{4}$$

where:
$\bar{x}$ = mean value
x_i = score of the i-th respondent's answer
n = number of respondents

The mean value is then analyzed using the UEQ-S Data Analysis Tool to obtain the Pragmatic Quality, Hedonic Quality, and Overall values. These results are subsequently compared with the UEQ-S benchmark to categorize them as Bad, Below Average, Above Average, Good, or Excellent. This analysis determines the level of user experience with the D-PANDA application, considering both usability and the app's appeal.

3. Results and Discussion

This research produces the D-PANDA Android application, which supports the document digitalization process through mobile devices. The application is developed using Java, PHP Native, and MariaDB, with integration of Google ML Kit and OpenCV. Figure 4 illustrates the implementation of the D-PANDA application interface, which consists of the login page, registration page, dashboard, document scan page, document history page, and account settings page. The login and registration pages are used for user authentication through OTP and JWT mechanisms. The dashboard serves as the central navigation hub for the app's main features. Document digitalization is performed on the document scan page, which utilizes Google ML Kit and OpenCV to detect and process document images. The scan results are saved and can be accessed again via the document history page, whereas the settings page is used to manage user accounts.

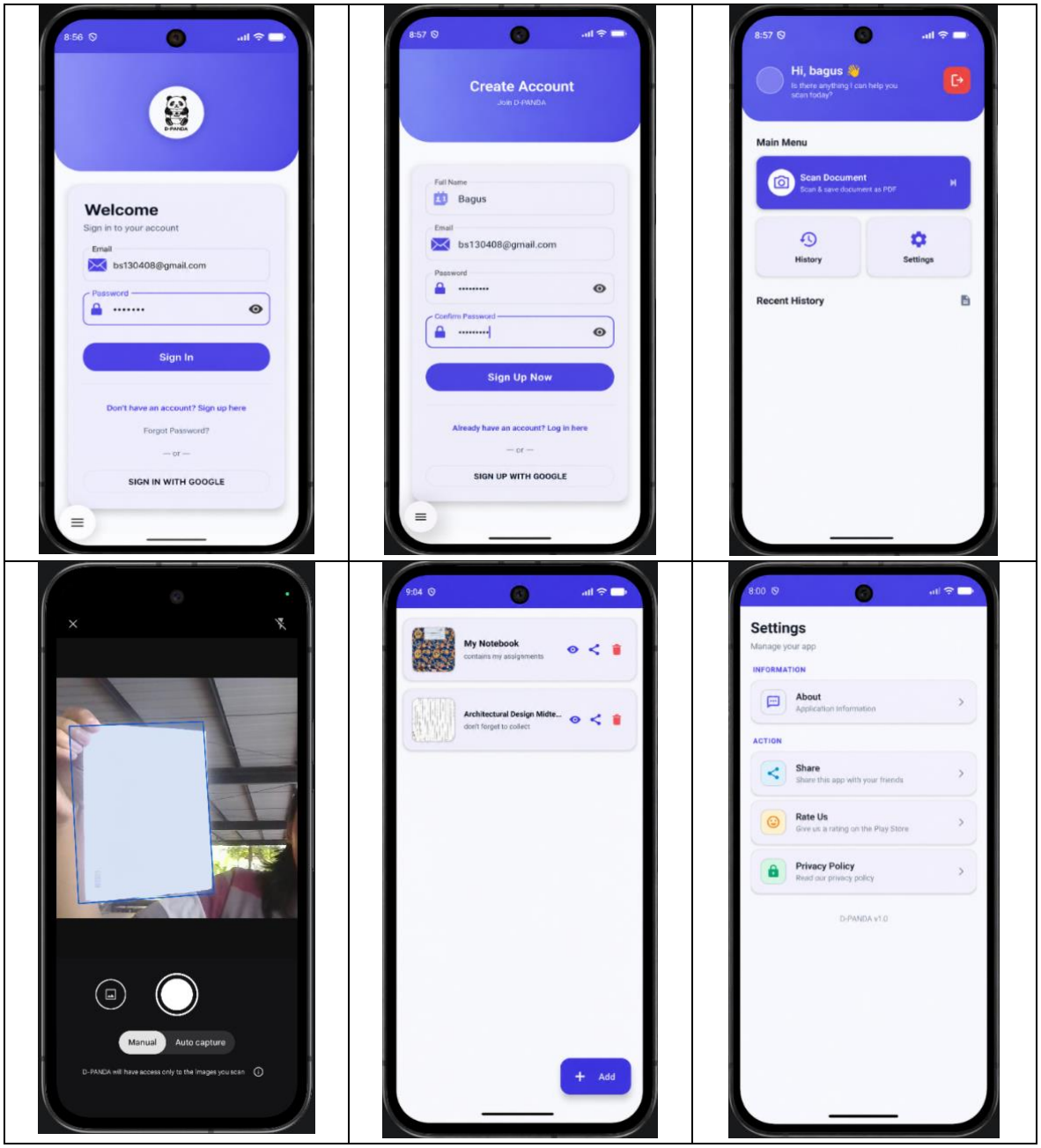

Figure 4. D-PANDA Application Interface

Based on the Black Box Testing results in Table 2, all main functions of the D-PANDA application operate according to the system requirements. The registration, login, document scanning, image processing using Google ML Kit and OpenCV, document storage, document history display, and logout processes all ran successfully without any functional errors. These test results demonstrate that the application meets the designed functional requirements and is ready for use as an Android-based digital document medium.

Table 2. Black Box Testing Results

No	Test Scenario	Expected Result	Actual Result	Status
1	The user performs registration by filling in valid data.	The system successfully stores user data and sends an OTP for account verification.	User data successfully saved and OTP successfully sent.	Success
2	The user performs login using a valid email and password.	The system successfully performs authentication and displays the dashboard page.	The user successfully logged into the dashboard page and obtained a JWT token.	Success
3	The user selects the Document Scan menu and captures a document image using the camera.	The system successfully automatically detects the document area using Google ML Kit.	Document area successfully detected and ready for processing.	Success
4	The system processes the document detection results using OpenCV.	The system generates a digital document with better quality.	Document successfully processed with perspective correction and image quality enhancement.	Success
5	The user saves the scanned document result.	The document is successfully saved to the server and database.	Document successfully saved and can be accessed again.	Success
6	The user opens the Document History menu.	The system displays all documents that have been saved.	Document list successfully displayed according to the user's data.	Success
7	The user selects the Settings menu and then performs logout.	The system ends the user's session and returns to the login page.	The user successfully exited the application and was directed to the login page.	Success

The document-detection test results shown in Table 3 indicate that Google ML Kit is able to detect document areas effectively across all sample documents used. OpenCV then successfully processed the images, producing a more tidy and readable digital document.

Table 3. Document Detection Test Results

No	Document Type	Result
1	Letter	Success
2	Form	Success
3	Invoice	Success
4	A4 Document	Success

User-experience evaluation was performed using the User Experience Questionnaire Short (UEQ-S) with 15 respondents. The questionnaire consists of eight attribute pairs that represent two main dimensions: Pragmatic Quality and Hedonic Quality. The evaluation results shown in Table 4 indicate that Pragmatic Quality achieved a mean value of 2.232, whereas Hedonic Quality achieved a value of 2.286. The overall value (Overall) was 2.259. Based on the UEQ-S benchmark shown in Figure 5, all three scales fell into the Excellent category, which indicates that the application is among the top 10 % of applications according to the UEQ-S benchmark database.

Table 4. UEQ-S Analysis Results

Scale	Mean	Benchmark	Interpretation
Pragmatic Quality	2.232	Excellent	In the range of the 10% best results
Hedonic Quality	2.286	Excellent	In the range of the 10% best results
Overall	2.259	Excellent	In the range of the 10% best results

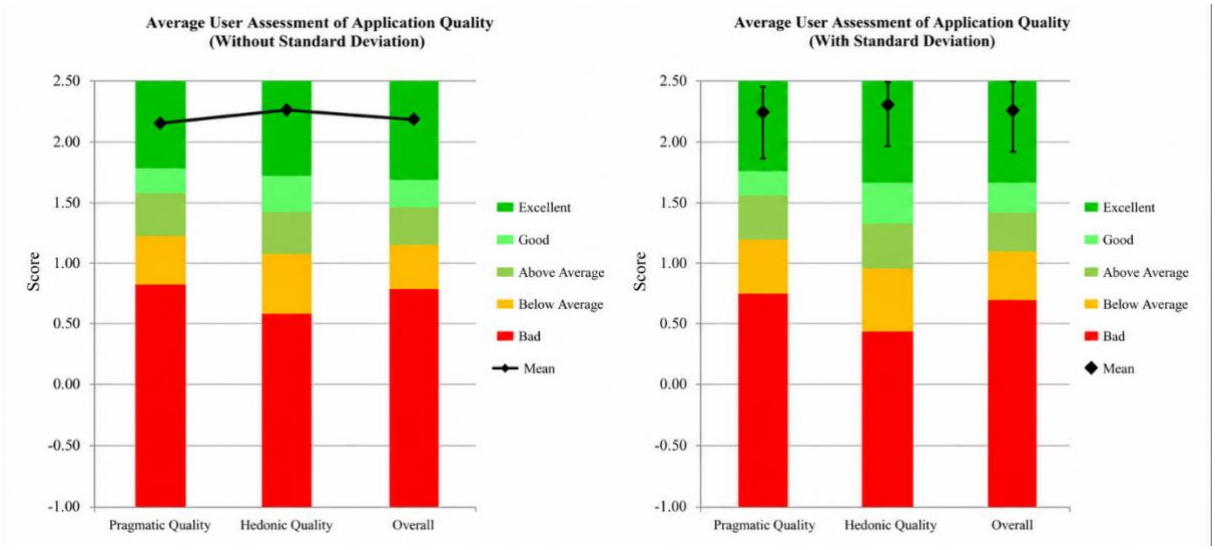

Figure 5. UEQ-S benchmark

The Pragmatic Quality score shows that the application is easy to learn, easy to use, and capable of helping users perform document digitalization efficiently. In contrast, the Hedonic Quality score shows that the interface appearance and user interaction experience are perceived as attractive and provide a positive impression for users. Overall, the UEQ-S evaluation results confirm that the D-PANDA application not only meets the functional aspects but also provides a very good user experience.

The high Pragmatic Quality score can be attributed to several factors. First, the integration of Google ML Kit Document Scanner API eliminates the need for manual document boundary adjustment, which is a common pain point in conventional mobile scanning. Second, the OpenCV-based post-processing pipeline (specifically Otsu thresholding and perspective correction) produces clean, readable outputs that users associate with professional scanning quality. Third, the seamless workflow from capture to PDF storage reduces cognitive load and task completion time.

The high Hedonic Quality score suggests that users found the application visually appealing and engaging. This may be due to the clean, intuitive interface design and the satisfaction derived from obtaining high-quality scan results with minimal effort. Previous studies have shown that perceived ease of use and visual aesthetics are strong predictors of continued technology adoption [16], which bodes well for D-PANDA's practical utility in real-world settings.

Several prior studies have addressed mobile document digitization using various approaches. Skoryukina et al. [4] demonstrated real-time document detection on mobile devices using conventional computer vision techniques, achieving detection rates of approximately 85-90% under controlled conditions. However, their approach required manual parameter tuning and was sensitive to lighting variations. In contrast, D-PANDA leverages Google ML Kit's deep learning-based detection, which is more robust to diverse environmental conditions and requires no manual calibration. Feng et al. [5, 6] proposed deep learning-based document rectification methods that achieve high-quality perspective correction. While these methods produce excellent results, they require substantial computational resources and are typically implemented on server-side infrastructure. D-PANDA takes a different approach by combining ML Kit's on-device detection with OpenCV's lightweight post-processing, achieving a balance between quality and computational efficiency that is suitable for resource-constrained mobile devices.

Compared to commercial applications such as Adobe Scan, Microsoft Lens, and CamScanner, D-PANDA offers several distinct advantages. First, it provides full algorithmic transparency—the image processing pipeline using Otsu thresholding and gradient-based edge detection is explicitly documented and reproducible. Second, D-PANDA supports a self-hosted backend using PHP Native and MariaDB, allowing organizations to maintain data sovereignty rather than relying on proprietary cloud ecosystems. Third, D-PANDA is a cost-free solution with no

recurring subscription fees, making it particularly suitable for educational institutions, government offices, and small enterprises with limited budgets. A key technical contribution of D-PANDA is the complementary integration of ML Kit and OpenCV. ML Kit excels at automated document boundary detection and initial perspective correction, but its output can still benefit from additional enhancement. OpenCV fills this gap by applying adaptive thresholding (Otsu's method) and edge refinement to produce cleaner, more readable document images. This hybrid approach (combining the strengths of deep learning-based detection with traditional computer vision post-processing) represents a pragmatic and effective solution for mobile document digitization.

The findings of this research have several practical implications. First, D-PANDA demonstrates that high-quality document digitization is achievable on standard Android devices without requiring specialized hardware or cloud-based processing. This democratizes access to professional-grade document scanning, particularly in developing regions where smartphone penetration is high but access to flatbed scanners and commercial software is limited. Second, the integrated archive management feature addresses a critical gap in existing mobile scanning solutions. Most commercial applications focus exclusively on scanning and storage in proprietary clouds, whereas D-PANDA provides users with control over their document metadata and storage infrastructure. This is particularly relevant for organizations handling sensitive information, where data sovereignty and compliance with privacy regulations are paramount. Third, the positive UEQ-S results suggest that users are receptive to the integrated scanning-and-management workflow. This indicates potential for broader adoption of self-hosted, privacy-preserving document management solutions, especially as awareness of data privacy issues continues to grow.

4. Conclusion

The Digital Portable Archive and Document Assistant (D-PANDA) application was successfully developed as an Android-based document digitalization solution by utilizing the Google ML Kit Document Scanner API and OpenCV to generate digital documents in PDF format. The Black Box Testing results show that all application functions operate according to the system requirements. In addition, UEQ-S evaluation yielded Pragmatic Quality of 2.232, Hedonic Quality of 2.286, and Overall of 2.259, all of which fall into the Excellent category. These results indicate that the application has good functional quality and is able to deliver a very satisfying user experience.

The hybrid integration of Google ML Kit and OpenCV provides a balanced approach that combines automated detection with documented image enhancement techniques. However, the dependency on Google Play Services, device camera variability, and lack of end-to-end encryption represent important limitations that should be addressed in future work. Despite these limitations, D-PANDA offers a viable alternative to commercial solutions for organizations prioritizing data sovereignty and cost efficiency. Future research can focus on extending D-PANDA to support real-time multi-document batch scanning, integrate with cloud-based storage for larger archives, and incorporate advanced AI models such as deep learning for automated document classification and extraction to further enhance processing speed and accuracy.

References

[1] K. Kyamakya, A. Haj Mosa, F. A. Machot, and J. C. Chedjou, "Document-Image Related Visual Sensors and Machine Learning Techniques," *Sensors*, vol. 21, no. 17, p. 5849, Jan. 2021, doi: 10.3390/s21175849.

[2] R. Bernardino, R. D. Lins, and R. da S. Barboza, "A Quality, Size and Time Assessment of the Binarization of Documents Photographed by Smartphones," *J. Imaging*, vol. 9, no. 2, p. 41, Feb. 2023, doi: 10.3390/jimaging9020041.

[3] H. Feng, W. Zhou, J. Deng, Q. Tian, and H. Li, "DocScanner: Robust Document Image Rectification with Progressive Learning," Dec. 24, 2022, *arXiv*: arXiv:2110.14968. doi: 10.48550/arXiv.2110.14968.

[4] N. Skoryukina, D. P. Nikolaev, A. Sheshkus, and D. Polevoy, "Real time rectangular document detection on mobile devices," in *Seventh International Conference on Machine Vision (ICMV 2014)*, SPIE, Feb. 2015, p. 458. doi: 10.1117/12.2181377.
[5] X. Li, B. Zhang, J. Liao, and P. V. Sander, "Document Rectification and Illumination Correction using a Patch-based CNN," Sep. 20, 2019, *arXiv*: arXiv:1909.09470. doi: 10.48550/arXiv.1909.09470.
[6] H. Feng, W. Zhou, J. Deng, Y. Wang, and H. Li, "Geometric Representation Learning for Document Image Rectification," Oct. 15, 2022, *arXiv*: arXiv:2210.08161. doi: 10.48550/arXiv.2210.08161.
[7] J. Canny, "A Computational Approach to Edge Detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986, doi: 10.1109/TPAMI.1986.4767851.
[8] P. Arbeláez, M. Maire, C. Fowlkes, and J. Malik, "Contour Detection and Hierarchical Image Segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 898–916, May 2011, doi: 10.1109/TPAMI.2010.161.
[9] N. A. Rehman and F. Haroon, "Adaptive Gaussian and Double Thresholding for Contour Detection and Character Recognition of Two-Dimensional Area Using Computer Vision," *Eng. Proc.*, vol. 32, no. 1, p. 23, 2023, doi: 10.3390/engproc2023032023.
[10] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," *IEEE Trans. Syst. Man Cybern.*, vol. 9, no. 1, pp. 62–66, Jan. 1979, doi: 10.1109/TSMC.1979.4310076.
[11] J. Sauvola and M. Pietikäinen, "Adaptive document image binarization," *Pattern Recognit.*, vol. 33, no. 2, pp. 225–236, Feb. 2000, doi: 10.1016/S0031-3203(99)00055-2.
[12] "ML Kit," Google for Developers. Accessed: Jun. 18, 2026. [Online]. Available: https://developers.google.com/ml-kit/vision/doc-scanner/android
[13] "Android Developers Blog: Easily add document scanning capability to your app with ML Kit Document Scanner API." Accessed: Jun. 18, 2026. [Online]. Available: https://android-developers.googleblog.com/2024/02/ml-kit-document-scanner-api.html
[14] A. Neumann, N. Laranjeiro, and J. Bernardino, "An Analysis of Public REST Web Service APIs," *IEEE Trans. Serv. Comput.*, vol. 14, no. 4, pp. 957–970, Jul. 2021, doi: 10.1109/TSC.2018.2847344.
[15] A. Belkhir, M. Abdellatif, R. Tighilt, N. Moha, Y.-G. Guéhéneuc, and É. Beaudry, "An Observational Study on the State of REST API Uses in Android Mobile Applications," in *2019 IEEE/ACM 6th International Conference on Mobile Software Engineering and Systems (MOBILESoft)*, May 2019, pp. 66–75. doi: 10.1109/MOBILESoft.2019.00020.
[16] "User Experience Questionnaire Handbook." Accessed: Jun. 27, 2026. [Online]. Available: https://www.researchgate.net/publication/281973617_User_Experience_Questionnaire_Handbook
[17] "Sosialisasi Dan Pelatihan Aplikasi Google Form Sebagai Kuisioner Online Untuk Meningkatkan Kualitas Pelayanan." Accessed: Jun. 27, 2026. [Online]. Available: https://www.researchgate.net/publication/335500269_Sosialisasi_Dan_Pelatihan_Aplikasi_Google_Form_Sebagai_Kuisioner_Online_Untuk_Meningkatkan_Kualitas_Pelayanan
[18] S. Ellen, "Slovin's Formula Sampling Techniques," Sciencing. Accessed: Jun. 30, 2026. [Online]. Available: https://www.sciencing.com/slovins-formula-sampling-techniques-5475547/

How to cite this article:
Sugiyanto, Surya E B, Wibisono A W S. Integration of Google ML Kit and OpenCV for Android-Based Document Digitization and Archive Management. *Jurnal Teknologi dan Manajemen.* 2026 July; 7(2): 195-204. DOI: 10.31284/j.jtm.2026.v7i2.9264