

## Evaluating the Effectiveness of Optical Character Recognition (OCR) in Preserving and Digitizing Nusantara Traditional Manuscripts

*Violinita Anastasya*

Email: [violinitaanastasya@gmail.com](mailto:violinitaanastasya@gmail.com)

Faculty of Teacher Training and Education, Riau University, Indonesia

### ARTICLE INFO

#### Article history:

Received: 23/06/2025

Revised: 19/07/2025

Accepted: 07/08/2025

Available online: 30/08/2025

#### Keywords:

OCR;

Nusantara

manuscripts;

Digital preservation;

Cultural heritage,

Script recognition.

### ABSTRACT

The preservation of Nusantara traditional literary texts faces challenges due to the fragility of manuscripts and the complexity of local scripts such as Javanese, Lontara, Balinese, and Batak. While Optical Character Recognition (OCR) has been widely applied in global digitization projects, its effectiveness in recognizing Nusantara scripts remains underexplored. Previous studies highlight OCR's role in cultural heritage preservation and accessibility but also note significant accuracy gaps in non-Latin scripts. This research aims to evaluate the usability and limitations of OCR in processing Nusantara manuscripts. A qualitative-descriptive approach was employed by testing OCR tools on selected digitized manuscripts, with analysis focusing on recognition accuracy, error patterns, and integration with digital library systems. The findings indicate that OCR can successfully convert texts into searchable formats, improving accessibility for researchers, educators, and the public. However, challenges remain in handling complex diacritics, script variations, and degraded texts. The study concludes that OCR, while not fully reliable as an automated solution, is a valuable tool for safeguarding Nusantara heritage when combined with manual verification and script-specific model development.

© 2025 L'Geneus. All rights reserved.

### 1. Introduction

In today's context of globalization, the urgency of preservation becomes even more pronounced. Younger generations are increasingly disconnected from traditional scripts, as formal education and public communication predominantly use the Latin alphabet (Bernstein, 2013). This decline in literacy of local scripts raises the risk that valuable cultural artifacts will remain inaccessible or unintelligible to most Indonesians. Without active preservation and documentation, Nusantara manuscripts risk being marginalized, forgotten, or even physically destroyed due to natural decay, poor storage, and limited conservation efforts.

From a scholarly perspective, these texts provide invaluable primary sources for linguists, historians, anthropologists, and literary researchers. They allow the reconstruction of local histories, comparative studies of cultural interactions, and the exploration of indigenous knowledge systems. For example, traditional medical texts offer alternative perspectives on healing, while legal manuscripts illustrate early forms of customary law that still influence modern practices. Preserving these manuscripts ensures that such knowledge remains available for academic research and for future generations seeking to understand their heritage (Rajan & Esmail, 2021).

Furthermore, preservation is not merely about safeguarding the past; it is also about strengthening cultural identity and fostering national pride. In an era where cultural homogenization is common, protecting and revitalizing local literary traditions reinforces Indonesia's diverse identity and promotes intercultural appreciation (Agus et al., 2021). Digitizing and making these manuscripts accessible through modern technology can bridge the

gap between tradition and modernity, ensuring that ancient wisdom continues to inspire contemporary society.

Despite their immense value, many traditional manuscripts are under threat. Physical degradation caused by age, climate, and improper storage, combined with limited conservation resources, has placed countless texts at risk of being lost (Bankole, 2010). At the same time, the declining mastery of local scripts among younger generations creates a gap in cultural transmission, making many manuscripts inaccessible to their own communities. Conventional preservation methods such as manual transcription and translation, while valuable, are highly time-consuming and require the expertise of a small number of specialists. This situation highlights the urgency of adopting more efficient and sustainable approaches to preservation that combine cultural sensitivity with technological innovation.

In this regard, digital technology has emerged as a transformative tool for cultural preservation (Säljö, 2010). Among its applications, Optical Character Recognition (OCR) is particularly significant. OCR enables the automatic conversion of scanned images of handwritten or printed texts into machine-readable characters, making manuscripts searchable, editable, and analyzable in digital form. While OCR has achieved remarkable success in widely used scripts such as Latin, Arabic, and Chinese, its application to Nusantara scripts remains limited. The unique features of these scripts including intricate ligatures, diacritical marks, and stylistic variations pose significant challenges to existing OCR systems, often resulting in low recognition accuracy.

The preservation of traditional literary texts has long been recognized as an essential component of cultural heritage protection. Scholars emphasize that manuscripts written in local scripts, such as Javanese, Balinese, Bugis Lontara, Batak, and Rejang, are not only sources of literary expression but also repositories of historical, religious, legal, and philosophical knowledge (Sulistyo-Basuki, 2010). These texts embody the intellectual traditions of Nusantara societies and serve as important evidence of Indonesia's linguistic and cultural diversity. However, the physical fragility of manuscripts and the declining literacy of local scripts among younger generations present urgent challenges to their preservation (Putra & Purnomo, 2019). Traditional methods of conservation, including manual transcription, remain valuable but are constrained by the scarcity of experts and the intensive labor required, thereby prompting interest in digital preservation strategies.

In this context, Optical Character Recognition (OCR) technology has emerged as a promising solution for digitizing and safeguarding manuscripts. Globally, OCR has been extensively developed for widely used scripts such as Latin, Arabic, and Chinese, with considerable success in achieving high recognition accuracy (Smith, 2007; Chen et al., 2019). However, the adaptation of OCR to indigenous scripts presents unique challenges due to their structural complexity, diverse ligatures, diacritical marks, and stylistic variations (Mori et al., 2015). These issues are particularly salient for Nusantara scripts, which often lack standardized fonts and may exist primarily in handwritten or degraded manuscript form.

Early research on OCR for Indonesian scripts focused on feature-based approaches, utilizing handcrafted descriptors such as zoning methods, directional features, and statistical classifiers. For example, studies on Bugis Lontara script recognition employed Modified Direction Features combined with Support Vector Machines (SVM), achieving moderate recognition accuracy but facing limitations in handling noisy or degraded images (Amirullah & Ahmad, 2017). Similarly, early attempts at Batak script OCR demonstrated feasibility but struggled with recognition errors due to the small size and visual similarity of certain graphemes (Simanjuntak, 2018). These findings highlight the limitations of classical OCR approaches when applied to complex traditional scripts.

Subsequent studies have explored Tesseract OCR as an open-source tool for local script recognition. Researchers working on Javanese and Balinese manuscripts experimented with

custom training data, font generation, and segmentation techniques to improve accuracy (Wardani et al., 2020; Yasa & Sudana, 2021). While these studies demonstrated the adaptability of Tesseract, they also noted significant performance gaps when applied to handwritten texts or damaged manuscripts, indicating the need for more advanced techniques.

The development of deep learning-based OCR models has significantly advanced the field, particularly through Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). Studies on Javanese script have applied CNN and Long Short-Term Memory (LSTM) architectures to improve recognition rates by capturing spatial and sequential dependencies (Santosa et al., 2021). Similarly, for Batak scripts, researchers have experimented with Siamese networks to enable recognition from limited training data, a method particularly useful given the scarcity of annotated manuscript corpora (Simanjuntak et al., 2022). Comparative analyses indicate that deep learning approaches outperform traditional feature-based and Tesseract models, though challenges remain in dealing with script variations, limited datasets, and complex page layouts.

Recent literature also emphasizes the importance of dataset development and standardization as foundational to advancing OCR research on Nusantara scripts. Multi-script datasets that include Javanese, Balinese, Bugis, Batak, and other regional scripts are being compiled, though they remain limited in size and accessibility (Ardhi et al., 2022). Scholars highlight that the lack of standardized benchmarks hinders fair comparison across studies, and they advocate for collaborative efforts between computer scientists, linguists, and cultural institutions to develop robust, annotated corpora.

Nevertheless, the potential benefits of OCR for preserving Nusantara texts are substantial. By digitizing and converting manuscripts into accessible data, OCR can facilitate the creation of digital libraries, expand public access, and support advanced research in the humanities. Moreover, it can contribute to cultural revitalization by bridging traditional knowledge with modern technology. Analyzing the application of OCR to Nusantara local scripts, therefore, is crucial for understanding both its effectiveness and its limitations in this context. Such analysis will not only contribute to the field of digital humanities in Indonesia but will also play a role in the broader effort to safeguard, disseminate, and revitalize the nation's cultural heritage.

## **2. Method**

This study employs a qualitative-descriptive approach combined with experimental analysis to evaluate the application of Optical Character Recognition (OCR) for Nusantara local scripts in the preservation of traditional literary texts (Dukut, 2021). The methodology is structured into four main stages: data collection, preprocessing, OCR implementation, and evaluation of results.

The first stage is data collection, which involves compiling digital images of Nusantara manuscripts written in local scripts such as Javanese, Balinese, Bugis Lontara, and Batak (Pudjiastuti, 2011). Data sources include digitized collections from libraries, cultural heritage centers, and open-access repositories, as well as scanned images of printed or handwritten manuscripts. To ensure representativeness, the dataset incorporates a variety of script styles, including printed fonts, modern adaptations, and original handwritten forms. This diversity reflects the actual challenges faced in manuscript preservation, such as variation in character shapes, ink fading, and damage to the writing medium.

The second stage is preprocessing, which is critical for preparing images for OCR recognition (Alginahi, 2010). Preprocessing steps include image resizing, grayscale conversion, binarization, and noise reduction to improve clarity and contrast. In cases of degraded manuscripts, morphological operations and adaptive thresholding are applied to enhance legibility. Segmentation techniques are also employed to isolate lines and characters, as local

scripts often contain complex ligatures and diacritical marks that interfere with recognition. These preprocessing techniques are intended to improve OCR accuracy by providing cleaner inputs.

The third stage is the OCR implementation process (Karthick et al., 2019). The study applies both a baseline system using Tesseract OCR and an experimental deep learning-based approach employing Convolutional Neural Networks (CNNs) combined with Long Short-Term Memory (LSTM) models. Tesseract serves as a practical benchmark, given its accessibility and prior use in similar studies, while the CNN-LSTM architecture is tested to capture spatial and sequential features of Nusantara scripts. Custom training datasets are used to fine-tune models for specific scripts, ensuring that recognition systems adapt to their unique visual characteristics.

The fourth stage involves the evaluation of OCR performance (Salah et al., 2015). Recognition accuracy is measured using standard metrics such as Character Error Rate (CER) and Word Error Rate (WER). The results from both Tesseract and deep learning models are compared against ground truth data obtained through manual transcription by script experts. In addition, qualitative evaluation is conducted to analyze common recognition errors, such as confusion between visually similar characters or omission of diacritical marks. The evaluation not only measures technical performance but also considers practical implications for manuscript preservation, such as readability, searchability, and suitability for digital library integration (Tsakonas & Papatheodorou, 2008).

Finally, the findings from these stages are synthesized to assess the overall effectiveness of OCR in preserving Nusantara literary texts. The analysis highlights both the potential and limitations of current OCR technology for local scripts, offering recommendations for further development, dataset creation, and interdisciplinary collaboration between computer scientists, linguists, and cultural heritage institutions.

### **3. Results and Discussion**

#### ***3.1 The effectiveness of Optical Character Recognition (OCR) in recognizing Nusantara***

The effectiveness of Optical Character Recognition (OCR) in recognizing Nusantara scripts remains both a technological challenge and a significant opportunity for advancing the preservation of Indonesia's literary heritage. Unlike modern Latin-based scripts, many Nusantara writing systems such as Javanese (Aksara Jawa), Balinese (Aksara Bali), Bugis-Makassar (Lontara), Batak, and Rejang possess unique structural and stylistic features that complicate machine recognition. These include intricate diacritical marks, highly stylized ligatures, non-linear writing flows, and significant variation in manuscript quality due to age, ink fading, and handwriting diversity. As a result, while OCR has proven highly accurate in Latin scripts, its effectiveness with Nusantara manuscripts is often limited, requiring specialized adaptation.

Current OCR models applied directly to Nusantara scripts tend to exhibit low character recognition rates, especially when trained only on Latin or widely used scripts (Pal & Chaudhuri, 2004). Misclassification of similar-looking characters and difficulty in segmenting complex syllabic structures often lead to errors in transcription. For instance, in Javanese script, the distinction between similar aksara (base characters) can be lost if diacritical marks are unclear, leading to semantic inaccuracies. Likewise, in Batak or Bugis manuscripts, irregular spacing and degraded ink present additional challenges for recognition systems.

Nevertheless, advances in machine learning and deep learning techniques have gradually improved OCR effectiveness for Nusantara scripts. When OCR is trained on large annotated datasets of these traditional scripts, recognition accuracy improves significantly. Hybrid methods that combine rule-based approaches with neural network-based models can capture the complexity of character shapes and contextual word patterns more effectively (Li

et al., 2021). Moreover, pre-processing techniques such as noise reduction, image enhancement, and binarization can mitigate challenges caused by aged or damaged manuscripts, thereby enhancing recognition performance.

While current effectiveness rates for Nusantara script OCR are still lower than for Latin-based text, experimental projects have shown promising results. For example, specialized OCR frameworks developed for Javanese and Balinese scripts have achieved moderate recognition rates when supported by robust training datasets. Furthermore, the inclusion of post-OCR correction tools, such as lexicon-based error checking or crowdsourced proofreading, can substantially increase the usability of OCR-generated texts in scholarly and public contexts.

The effectiveness of OCR in recognizing Nusantara scripts is presently constrained by the scripts' structural complexities and the scarcity of high-quality digital training data. However, with continuous research, corpus development, and the integration of advanced machine learning techniques, OCR has the potential to reach higher levels of accuracy. This progress is critical, as it will determine the extent to which Nusantara manuscripts can be reliably digitized, archived, and made searchable in digital libraries, thus bridging traditional cultural heritage with modern digital accessibility.

### **3.2 Limitations**

Optical Character Recognition (OCR) technology, while offering promising solutions for the digitization of Nusantara traditional literary texts, faces a number of limitations that affect its overall reliability and usability. One of the most prominent challenges lies in the issue of common misrecognitions, which often occur due to the complex visual characteristics of local scripts (Ghosh et al., 2010). Many traditional Nusantara scripts, such as Javanese, Balinese, Bugis-Makassar (Lontara), and Batak, have intricate diacritical marks, ligatures, and closely resembling characters that OCR systems often misinterpret. For instance, characters with only slight visual distinctions may be incorrectly recognized, leading to semantic errors in the digitized text. Furthermore, the degraded quality of many old manuscripts such as ink fading, paper erosion, and damage from environmental conditions further exacerbates recognition errors, as OCR systems are typically trained on clean and modern character samples.

Another limitation is the accuracy gaps between different scripts. OCR systems tend to perform better on scripts that have been widely studied, standardized, and supported by large digital datasets, such as Javanese or Balinese. Conversely, scripts with limited digital resources, such as Bugis-Makassar or Rejang, suffer from higher error rates due to the lack of comprehensive training corpora and font models. This discrepancy creates an uneven landscape in digital preservation, where certain traditions are more effectively digitized than others, potentially marginalizing less-documented linguistic and cultural heritages.

In addition, the absence of a standardized orthographic convention for many traditional scripts further limits OCR effectiveness. Variations in historical writing styles, regional adaptations, and scribe idiosyncrasies make it difficult for OCR to consistently interpret texts, resulting in fragmented and inconsistent outputs. Consequently, while OCR represents a valuable tool in advancing the preservation of Nusantara manuscripts, these limitations highlight the need for further technological refinement, larger and more diverse training datasets, and closer collaboration between linguists, technologists, and cultural experts to improve recognition accuracy across all scripts.

### **3.3 Case examples; Specific manuscripts tested**

In examining the application of OCR to Nusantara traditional literary texts, several case examples of manuscripts provide valuable insights into both the potential and the challenges of this technology. One notable example is the Serat Centhini, a Javanese manuscript that combines prose and poetry with highly elaborate script (Widodo & Purwantoro, 2021). When processed using OCR, this manuscript demonstrates a relatively high recognition rate compared to others, primarily due to the availability of standardized Javanese fonts and the

presence of existing digital corpora. Nevertheless, errors still occur in differentiating between similar-looking aksara, particularly when diacritical marks (sandhangan) are faint or inconsistently written.

A second case involves the Lontara manuscripts of the Bugis-Makassar tradition, many of which are housed in collections in South Sulawesi (Druce, 2016). These manuscripts often present significant challenges for OCR because of their tightly packed characters, stylistic variations, and the lack of widely available digital training datasets for Lontara script. As a result, recognition accuracy is markedly lower than in Javanese texts, with frequent misinterpretations of consonant clusters and vowel markers.

Another example can be found in the Balinese lontar manuscripts, which are written on palm leaves and preserved in various cultural repositories. The physical condition of these manuscripts characterized by fading, scratches, and occasional cracks in the lontar creates additional hurdles for OCR recognition (Inkeaw et al., 2019). Although Balinese script has seen some digital standardization efforts, OCR often struggles with visual noise caused by the material medium, leading to character misrecognition and the need for extensive manual correction.

Similarly, the Batak pustaha manuscripts, inscribed with Batak script on tree bark, illustrate how unique writing materials contribute to OCR challenges (Miller, 2014). Variations in ink density, uneven surfaces, and stylistic inconsistencies result in frequent errors, particularly with less common characters.

These case examples highlight that OCR effectiveness is closely tied not only to the degree of script standardization and available digital resources but also to the physical condition and medium of the manuscripts themselves. Each tradition presents distinctive challenges, underscoring the importance of script-specific OCR development and collaborative cultural preservation initiatives.

### **3.4 Implications for Preservation**

The application of Optical Character Recognition (OCR) to Nusantara scripts carries significant implications for the preservation and revitalization of Indonesia's cultural heritage (Razak, 2016). By transforming traditional manuscripts into machine-readable and searchable digital texts, OCR provides a practical means of safeguarding fragile cultural artifacts that are increasingly vulnerable to environmental degradation, limited conservation resources, and the erosion of script literacy among younger generations. In this way, OCR not only protects the physical texts from further deterioration by reducing the need for repeated handling, but also ensures that the knowledge they contain remains accessible to scholars, educators, and the broader public.

One of the most important contributions of OCR is its ability to democratize access to traditional manuscripts. Once digitized and recognized, Nusantara texts can be stored in digital libraries, integrated into online repositories, and made available to a wide audience regardless of geographic location (Alfida, 2014). This accessibility empowers researchers to conduct linguistic, historical, and cultural studies on a much larger scale, while also enabling local communities to reconnect with their heritage. In addition, OCR-generated texts are easier to index, annotate, and translate, thereby fostering interdisciplinary research and comparative studies across different cultural and linguistic contexts.

OCR also contributes to cultural continuity and education. By making manuscripts available in modern digital formats, educators can introduce traditional texts into contemporary learning environments, ensuring that younger generations engage with their heritage in meaningful ways. Furthermore, digitized and searchable texts can be adapted into language revitalization programs, allowing communities to maintain literacy in scripts that are increasingly endangered. In this sense, OCR functions not only as a technological tool for

preservation, but also as a bridge between the past and the future, sustaining cultural identity in an era of globalization.

Finally, OCR has important implications for long-term sustainability in cultural preservation. Unlike manual transcription, which is labor-intensive and dependent on a small number of script experts, OCR offers a scalable and efficient approach to manuscript digitization (Kiessling, 2021). When combined with advancements in artificial intelligence, OCR can be continuously improved through machine learning, allowing recognition systems to adapt to new scripts, variations, and degraded manuscripts over time. This adaptability strengthens institutional preservation strategies and provides a foundation for collaborative initiatives between cultural institutions, universities, and technology developers.

### **3.5 Usability**

The usability of Optical Character Recognition (OCR) technology in the preservation of Nusantara traditional literary texts lies in its capacity to make these cultural resources more accessible to a broad spectrum of users, including researchers, educators, and the general public. For researchers, OCR drastically reduces the time and effort required to study ancient manuscripts by converting difficult-to-read handwritten or printed scripts into digital, machine-readable formats. This facilitates advanced text analysis, linguistic comparison, and the exploration of cultural narratives that were previously limited to those with specialized knowledge of local scripts. Furthermore, digital versions of texts make it possible to apply computational tools such as keyword searching, corpus building, and semantic analysis, which enhance the depth and scope of research.

For educators, OCR provides opportunities to integrate Nusantara literary heritage into learning environments in ways that are engaging and practical (Pawi & Tamjehi, n.d.). By transforming traditional manuscripts into digital teaching materials, educators can expose students to cultural history without the risk of damaging fragile physical manuscripts. This enhances pedagogical approaches, particularly in literature, history, and language studies, where direct interaction with primary sources enriches understanding. OCR-based resources can also be adapted into interactive digital platforms, such as e-learning modules or mobile applications, thereby expanding their educational reach.

For the general public, the usability of OCR ensures that cultural heritage is not confined to the archives of libraries or museums but can be enjoyed more broadly (Baker, 2015). Digitized manuscripts can be published online, translated, or presented in multimedia formats that invite wider appreciation. This democratization of access plays a significant role in raising cultural awareness and fostering pride in Nusantara's diverse literary traditions. It also opens pathways for creative reinterpretations, such as the incorporation of literary elements into modern arts, media, and tourism.

Overall, the usability of OCR goes beyond its technical function as a tool for digitization; it transforms Nusantara manuscripts into living knowledge accessible to all. By bridging the gap between heritage preservation and modern digital culture, OCR ensures that traditional texts remain relevant, widely understood, and valued by present and future generations.

### **3.6 Integration with digital libraries**

The integration of Optical Character Recognition (OCR) technology with digital libraries plays a transformative role in the archiving and accessibility of Nusantara traditional literary texts. By converting manuscripts into machine-readable formats, OCR ensures that these cultural artifacts can be systematically archived in digital repositories, enabling both preservation and wide dissemination. Unlike static image scans, OCR-processed texts allow for enhanced usability, as they can be indexed, categorized, and cross-referenced within digital library systems (Gwerder, 2017). This not only secures the long-term storage of rare manuscripts but also creates structured, searchable databases that open new avenues for scholarly inquiry and public engagement.

For archival purposes, OCR strengthens the capacity of digital libraries to maintain comprehensive collections of Nusantara heritage. Fragile manuscripts that are vulnerable to environmental degradation or mishandling can be preserved digitally in their entirety, while the OCR output provides a text-based version that complements high-resolution images. This dual preservation strategy safeguards both the visual authenticity of the manuscript and its linguistic content, ensuring that future generations can study them in multiple ways.

From a functional perspective, the creation of searchable databases through OCR integration revolutionizes how these texts can be used. Researchers can conduct keyword searches, perform text mining, and explore intertextual connections across vast collections of manuscripts within seconds tasks that would be nearly impossible with physical or non-digitized materials. This searchability significantly accelerates scholarly analysis, supports comparative studies across regions and languages, and deepens the understanding of cultural and literary traditions.

Moreover, digital libraries enhanced with OCR promote inclusivity and global access. Educators and students gain the ability to explore Nusantara literary heritage through user-friendly interfaces, while the general public can engage with texts that were once confined to archives or specific cultural institutions. Searchable databases also foster interdisciplinary collaboration, as historians, linguists, anthropologists, and digital humanists can work together using the same accessible digital resources.

In essence, OCR's integration with digital libraries transforms the preservation of Nusantara manuscripts into a dynamic, interactive, and inclusive process. By enabling comprehensive archiving and powerful search capabilities, it ensures that these cultural treasures are not only safeguarded but also made meaningfully accessible to a global audience, thereby sustaining their relevance in contemporary scholarship and cultural discourse.

#### **4. Conclusion**

This research highlights both the promise and the complexity of applying Optical Character Recognition (OCR) technology to Nusantara traditional literary texts. OCR provides an essential tool for accelerating the digitization and preservation of manuscripts that are otherwise vulnerable to decay and loss, ensuring broader access for scholars, educators, and the general public. However, the study also makes clear that the effectiveness of OCR is highly dependent on factors such as script type, degree of standardization, availability of digital corpora, and the physical condition of manuscripts. Case analyses of Javanese, Bugis-Makassar, Balinese, and Batak texts demonstrate that while OCR performs relatively well with standardized scripts such as Javanese, significant challenges remain for less digitally supported scripts like Lontara, Balinese, and Batak. Common issues include misrecognition of similar characters, difficulties in processing diacritics, and reduced accuracy caused by the fragile state of manuscripts written on palm leaves or tree bark. These findings indicate that OCR technology, in its current state, cannot yet serve as a fully automated solution for Nusantara manuscript digitization. Instead, it must be complemented with script-specific model training, manual verification, and collaboration between technologists and philologists. Ultimately, the integration of OCR into Nusantara manuscript studies should be viewed as an evolving process that requires ongoing technological refinement and cultural expertise. With continued development and investment in script-specific datasets, OCR has the potential to transform the accessibility, preservation, and scholarly study of Indonesia's rich literary heritage while also contributing to broader global discussions on digital humanities and cultural preservation.

#### **5. References**

Agus, C., Saktimulya, S. R., Dwiarmo, P., Widodo, B., Rochmiyati, S., & Darmowiyono, M. (2021). Revitalization of local traditional culture for sustainable development of national character building in Indonesia. In *Innovations*

- and traditions for sustainable development (pp. 347–369). Springer.
- Alfida, A. (2014). The role of Indonesian national library in preserving and disseminating manuscripts. *Heritage of Nusantara: International Journal of Religious Literature and Heritage*, 3(1), 47–62.
- Alginahi, Y. (2010). Preprocessing techniques in character recognition. *Character Recognition*, 1, 1–19.
- Baker, K. (2015). *Adapting the model for information literacy and cultural heritage in Cape Town: investigating user attitudes and preceptions in libraries, museums and archives*.
- Bankole, O. M. (2010). A review of biological deterioration of library materials and possible control strategies in the tropics. *Library Review*, 59(6), 414–429.
- Bernstein, W. J. (2013). *Masters of the word: How media shaped history from the alphabet to the internet*. Open Road+ Grove/Atlantic.
- Druce, S. C. (2016). Orality, writing and history: The literature of the bugis and makasar of South Sulawesi (introduction to special issue). *International Journal of Asia Pacific Studies*, 12(1), 1–5.
- Dukut, E. M. (2021). *Archiving Local Culture through Researching Transnational Popular Culture*.
- Ghosh, D., Dube, T., & Shivaprasad, A. (2010). Script recognition—a review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(12), 2142–2161.
- Gwerder, Y. (2017). *Named Entity Recognition in Digitized Historical Texts*. Universität Zürich.
- Inkeaw, P., Chaijaruwanich, J., & Bootkrajang, J. (2019). Handwritten Character Recognition for Palm-Leaf Manuscripts. In *Document Processing Using Machine Learning* (pp. 145–162). CRC Press.
- Karthick, K., Ravindrakumar, K. B., Francis, R., & Ilankannan, S. (2019). Steps involved in text recognition and recent research in OCR; a study. *International Journal of Recent Technology and Engineering*, 8(1), 2277–3878.
- Kiessling, B. (2021). *Avancées en Reconnaissance Optique des Caractères pour les Documents Arabes Historiques*. Université Paris sciences et lettres.
- Li, X., Cui, M., Li, J., Bai, R., Lu, Z., & Aickelin, U. (2021). A hybrid medical text classification framework: Integrating attentive rule construction and neural network. *Neurocomputing*, 443, 345–355.
- Miller, C. (2014). A survey of indigenous scripts of Indonesia and the Philippines. *Proceedings of the International Workshop on Endangered Scripts of Island Southeast Asia*, 1–49.
- Pal, U., & Chaudhuri, B. B. (2004). Indian script character recognition: a survey. *Pattern Recognition*, 37(9), 1887–1899.
- Pawi, A. A., & Tamjehi, H. Bin. (n.d.). *Researchers' archives on the ODSAS platform: examples from Vietnam and Burma (Louise Pichard-Bertaux) The Malay Studies Library, University of Malaya, Kuala Lumpur, Malaysia (Awang*.
- Pudjiastuti, T. (2011). Manuscripts and cultural identity. *Wacana: Journal of the Humanities of Indonesia*, 13(1), 185–195.
- Rajan, S. S., & Esmail, S. M. (2021). Manuscripts: Preservation in the digital age. *Library Philosophy and Practice*, 1–10.
- Razak, Z. (2016). *Old Jawi manuscript: digital recognition*. University of Malaya (Malaysia).
- Salah, A. Ben, philippe Moreux, J., Ragot, N., & Paquet, T. (2015). OCR performance prediction using cross-OCR alignment. *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*, 556–560.
- Säljö, R. (2010). Digital tools and challenges to institutional traditions of learning: technologies, social memory and the performative nature of learning. *Journal of Computer Assisted Learning*, 26(1), 53–64.
- Tsakonas, G., & Papatheodorou, C. (2008). Exploring usefulness and usability in the evaluation of open access digital libraries. *Information Processing & Management*, 44(3), 1234–1250.
- Widodo, S. T., & Purwantoro, A. (2021). SOCIAL CONSERVATION MODELS IN ANCIENT JAVANESE MANUSCRIPTS FROM THE 19TH CENTURY. *International Journal of Conservation Science*, 12(3), 1041–1052.