

Penerapan Metode OCR dan NLP dalam Aplikasi Transliterasi Aksara Jawa Berbasis Android

1st Zerricho Helsa Bagus Givalle
Teknik Telekomunikasi
Telkom University
Bandung, Indonesia
zerrichohbg@student.telkomuniversity.
ac.id

2nd Obey Muhammad Abdillah
Teknik Telekomunikasi
Telkom University
Bandung, Indonesia
obeymuhammad@student.telkomuniver
sity.ac.id

3rd Hendra Aditya
Teknik Telekomunikasi
Telkom University
Bandung, Indonesia
hendraaditya@student.telkomuniversity
.ac.id

4th Inung Wijayanto
Teknik Telekomunikasi
Telkom University
Bandung, Indonesia
iwiwijayanto@telkomuniversity.ac.id

5th Sofia Saidah
Teknik Telekomunikasi
Telkom University
Bandung, Indonesia
sofiasaidahsfi@telkomuniversity.ac.id

Abstrak — Aksara Jawa merupakan salah satu warisan budaya Nusantara yang kian jarang digunakan dalam kehidupan sehari-hari, sehingga generasi muda mengalami kesulitan dalam membaca dan menulisnya. Minimnya pemanfaatan media digital untuk pembelajaran dan pelestarian aksara ini memperburuk kondisi tersebut. Penelitian ini bertujuan mengembangkan aplikasi transliterasi aksara Jawa berbasis Android yang memanfaatkan kecerdasan buatan untuk melakukan transliterasi secara otomatis, akurat, dan mudah diakses. Aplikasi memiliki dua fitur utama yaitu transliterasi teks Latin ke aksara Jawa menggunakan model *Natural Language Processing* berbasis Gemini AI, yang mampu memahami konteks bahasa secara generatif tanpa memerlukan pelatihan ulang lokal dan transliterasi berbasis gambar melalui metode *Optical Character Recognition* (OCR) dengan model *Convolutional Neural Network* (CNN) dari TensorFlow untuk mengenali pola visual karakter. Sistem dibangun dengan *backend* Flask dan diintegrasikan dengan antarmuka Android. Hasil pengujian menunjukkan transliterasi teks menghasilkan keluaran yang konsisten sesuai kaidah penulisan, sementara model OCR mencapai akurasi 96,25% dalam mengenali karakter Hanacaraka. Dengan kombinasi teknologi NLP dan OCR, aplikasi ini berpotensi menjadi media digital yang efektif, interaktif, dan mudah diakses untuk mendukung pelestarian aksara Jawa di era digital.

Kata kunci— Aksara Jawa, Transliterasi, OCR, NLP, CNN

I. PENDAHULUAN

Aksara Jawa merupakan salah satu warisan budaya bangsa yang memiliki nilai historis dan estetika tinggi, namun penggunaannya semakin menurun di era modern akibat dominasi huruf latin dalam kehidupan sehari-hari. Kondisi ini menimbulkan tantangan bagi generasi muda untuk mempelajari dan melestarikan aksara Jawa, terlebih karena minimnya media pembelajaran yang praktis dan interaktif [1][4].

Seiring berkembangnya teknologi kecerdasan buatan, berbagai penelitian telah dilakukan untuk mengembangkan sistem transliterasi aksara Jawa, baik berbasis teks maupun

pengenalan citra, dengan memanfaatkan metode seperti Optical Character Recognition (OCR) dan model Convolutional Neural Network (CNN) [10]i.

Beberapa penelitian sebelumnya telah menunjukkan hasil yang cukup baik, namun umumnya masih memiliki keterbatasan pada akurasi pengenalan aksara tulisan tangan, kecepatan pemrosesan, dan fleksibilitas penggunaan pada berbagai perangkat[5]-[7].

Berdasarkan permasalahan tersebut, penelitian ini mengusulkan pengembangan aplikasi transliterasi aksara Jawa yang mengintegrasikan model CNN untuk pengenalan karakter dan dukungan input berbasis teks serta gambar, sehingga pengguna dapat melakukan transliterasi secara cepat, akurat, dan mudah diakses melalui perangkat mobile. Tujuan utama pengembangan aplikasi ini adalah untuk menyediakan sarana digital yang efektif dalam mempelajari aksara Jawa sekaligus mendukung pelestarian warisan budaya di tengah perkembangan teknologi modern.

II. KAJIAN TEORI

A. Transliterasi aksara Jawa

Transliterasi aksara Jawa merupakan proses konversi tulisan beraksara Latin ke aksara Jawa atau sebaliknya. Sistem ini umumnya menggunakan pendekatan *rule-based* melalui pemetaan fonem dan silabel untuk mengubah bentuk tulisan secara sistematis. Aturan tersebut mampu menangani aksara dasar (hanacaraka), pasangan, dan sandhangan, namun terbatas dalam konteks bahasa yang kompleks [2][3][9].

B. Optical Character Recognition (OCR)

OCR adalah teknik pengenalan karakter dari citra atau dokumen hasil pemindaian. Untuk aksara non-Latin seperti Jawa, tantangannya terletak pada bentuk karakter yang kompleks dan keterbatasan dataset. *Convolutional Neural Network* (CNN) menjadi arsitektur yang umum digunakan

karena kemampuannya mengekstraksi fitur spasial dan mengenali pola visual dengan akurasi tinggi [8].

C. Natural Language Processing (NLP)

NLP adalah cabang kecerdasan buatan yang berfokus pada pemrosesan bahasa alami. Dalam penelitian ini digunakan Gemini AI, model *Large Language Model* yang mendukung transliterasi teks Latin ke aksara Jawa secara kontekstual dan generatif, serta tidak memerlukan pelatihan ulang lokal [11].

D. Convolutional Neural Network (CNN)

CNN adalah model *deep learning* yang efektif untuk klasifikasi citra. Arsitektur CNN memanfaatkan lapisan konvolusi, *pooling*, dan *fully connected layer* untuk mengenali pola visual. Pada penelitian ini, CNN digunakan dalam modul OCR untuk mengenali karakter Hanacaraka dari gambar dengan tingkat akurasi tinggi [12].

E. Dataset Dan Praproses Data

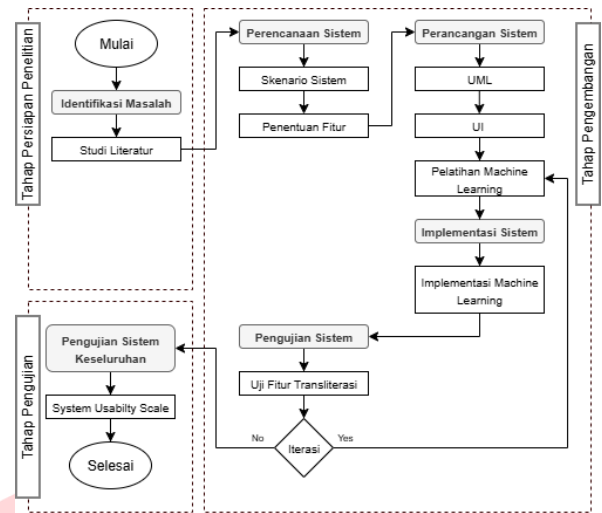
Dataset aksara Jawa terdiri dari data digital dan tulisan tangan, yang seimbang untuk melatih model agar mampu mengenali karakter dalam kondisi ideal maupun bervariasi. Praproses mencakup konversi *grayscale*, *thresholding*, segmentasi karakter, serta *normalize* dan *resize* citra agar seragam untuk proses pelatihan model [13].

III. METODE

Alur prosedur penelitian pada pengembangan aplikasi transliterasi Aksara Jawa ini terdiri dari tiga tahap utama, yaitu tahap persiapan penelitian, tahap pengembangan, dan tahap pengujian. Tahap persiapan penelitian diawali dengan identifikasi masalah dan studi literatur untuk memahami permasalahan serta solusi yang telah ada sebelumnya.

Selanjutnya, pada tahap pengembangan dilakukan perencanaan sistem yang mencakup penyusunan skenario sistem dan penentuan fitur. Proses ini dilanjutkan dengan perancangan sistem melalui pembuatan diagram UML, perancangan antarmuka (UI), pelatihan model *machine learning*, hingga implementasi sistem dan integrasi model *machine learning* ke dalam aplikasi.

Setelah implementasi, tahap pengujian dilakukan dengan menguji fitur transliterasi untuk memastikan fungsionalitas dan akurasi model. Jika ditemukan kekurangan, dilakukan iterasi pengembangan hingga hasil sesuai dengan yang diharapkan. Pengujian keseluruhan sistem menggunakan metode *System Usability Scale* (SUS) untuk mengukur tingkat kepuasan pengguna. Tahapan ini memastikan aplikasi yang dihasilkan memiliki kinerja yang baik dan memenuhi kebutuhan pengguna.



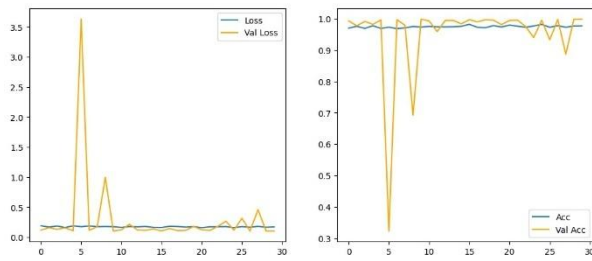
GAMBAR 1
(Prosedur Penelitian)

IV. HASIL DAN PEMBAHASAN

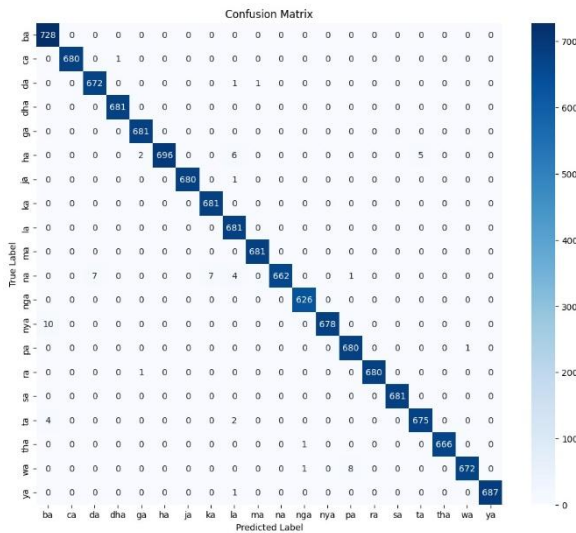
Implementasi aplikasi transliterasi aksara Jawa diawali dengan analisis kebutuhan dan perancangan antarmuka menggunakan Figma untuk menghasilkan wireframe dan mockup. Pengembangan dilakukan melalui tiga lingkungan utama, yaitu Android Studio untuk frontend, Visual Studio Code untuk backend berbasis Flask, dan Google Colab untuk pelatihan model *machine learning*. *Frontend* dibangun dengan Kotlin untuk mengelola tampilan dan logika aplikasi, sementara *backend* berfungsi sebagai server, menyediakan API endpoint, dan mengintegrasikan layanan Gemini AI untuk proses transliterasi.

Tahap pengolahan model mencakup pengumpulan dataset aksara Jawa, praproses dan augmentasi data, pelatihan model *Convolutional Neural Network* (CNN), evaluasi performa, dan konversi model ke format .h5. Setelah *frontend*, *backend*, dan model siap, dilakukan integrasi sistem serta pengujian menyeluruh untuk memastikan seluruh *pipeline* berjalan dari segmentasi dan normalisasi citra, klasifikasi karakter dengan CNN, hingga transliterasi berbasis teks menggunakan kecerdasan buatan.

Hasil penelitian menunjukkan bahwa aplikasi yang dikembangkan berfungsi sesuai rancangan, baik pada fitur transliterasi teks maupun pengenalan aksara berbasis citra. Pada transliterasi teks, model *Natural Language Processing* (NLP) berbasis Gemini AI berhasil mengubah teks Latin berbahasa Jawa menjadi aksara Jawa dengan keluaran yang konsisten secara struktur dan sesuai kaidah penulisan. Keunggulan model ini terletak pada kemampuannya memahami konteks bahasa secara generatif, sehingga transliterasi dapat mempertahankan makna dan ejaan yang benar tanpa memerlukan pelatihan ulang lokal.



GAMBAR 2
(Grafik Hasil *Training Loss* dan *Accuracy*)



GAMBAR 3
(*Confusion Matrix*)

Fitur pengenalan aksara berbasis gambar menggunakan *Optical Character Recognition* (OCR) dengan arsitektur *Convolutional Neural Network* (CNN) yang dilatih menggunakan dataset gabungan, yaitu 420 citra digital dan 650 citra tulisan tangan per karakter. Hasil pengujian terhadap model OCR menunjukkan tingkat akurasi sebesar 96,25%, yang mengindikasikan kinerja tinggi dalam mengenali karakter Hanacaraka pada berbagai kondisi visual. Analisis *confusion matrix* memperlihatkan bahwa sebagian besar kelas karakter dikenali dengan baik, meskipun terdapat sedikit kesalahan pada karakter dengan bentuk serupa, seperti “na” dan “tha” [14].

TABEL 1
(Hasil Training)

Training Result			
Aksara Jawa	Precision	Recall	f1-score
Ha	100%	98%	99%
Na	100%	97%	99%
Ca	100%	100%	100%
Ra	100%	100%	100%
Ka	99%	100%	99%
Da	99%	100%	99%
Ta	99%	99%	99%
Sa	100%	100%	100%
Wa	100%	99%	99%
La	98%	100%	99%
Pa	99%	100%	99%
Dha	100%	100%	100%
Ja	100%	100%	100%
Ya	100%	100%	100%
Nya	100%	99%	99%
Ma	100%	100%	100%
Ga	100%	99%	100%
Ba	98%	100%	99%
Tha	100%	100%	100%
Nga	100%	100%	100%
Accuracy			100%
Macro Avg	100%	100%	100%
Weighted Avg	100%	100%	100%

Evaluasi kinerja model juga dilakukan melalui metrik *precision*, *recall*, dan *F1-score* yang menunjukkan nilai rata-rata di atas 95%, menandakan keseimbangan antara ketepatan prediksi dan kelengkapan pengenalan. Hasil ini memperkuat temuan bahwa kombinasi CNN untuk pengenalan citra dan NLP untuk transliterasi teks memberikan solusi yang efektif dan adaptif terhadap variasi data masukan. Secara keseluruhan, integrasi teknologi ini menghasilkan sistem yang mampu mendukung pembelajaran, digitalisasi, dan pelestarian aksara Jawa dengan akurasi tinggi serta kemudahan akses melalui perangkat Android.

Penggunaan ketiga cara tersebut disesuaikan dengan jenis data dan sejauh mana deskripsi data akan dijelaskan. Misalnya, pada awal peneliti memaparkan narasi temuannya, kemudian didukung dengan sajian data dalam bentuk tabulasi, tabel atau grafik. Peneliti juga menyajikan data-data hasil penelitian, kemudian didukung grafik dilanjutkan deskripsi naratif.

TABEL 2
(*System Usability Scale*)

No	Nama	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	Total
1	Bambang	2	4	4	4	3	4	4	4	4	4	92,5
2	Wami	4	4	4	4	3	4	3	4	3	3	90
3	Sulis	2	4	4	4	4	4	4	4	4	4	90
4	Edi	3	3	3	4	4	4	4	4	3	4	85
5	Agus	4	3	4	3	4	4	4	4	4	3	87,5
6	Andri	4	4	4	4	3	4	4	3	4	3	92,5
7	Suci	4	3	3	4	3	3	3	3	4	3	82,5
8	Rini	4	3	4	3	3	4	3	3	4	4	87,5
9	Yayuk	4	3	4	3	3	3	4	3	4	4	87,5
10	Ulfa	4	3	3	4	3	3	4	3	4	3	85
11	Rizal	4	4	4	4	3	4	4	3	4	4	95
12	Tosi	4	4	3	4	3	4	4	4	4	4	95
13	Rizqi	3	3	3	3	4	4	4	4	4	3	87,5
14	Rian	4	4	3	3	3	3	3	3	4	4	85
15	Wahyu	4	4	4	4	3	4	4	4	4	4	97,5
16	Yudha	3	3	3	3	3	3	3	3	4	4	80
17	Ridho	4	3	4	3	4	3	4	3	4	4	90
18	Edy	4	4	3	3	4	4	3	3	4	4	90
19	Sutrisno	4	3	4	3	4	3	4	3	4	3	87,5
20	Adam	3	3	3	4	4	4	3	3	3	4	85
21	Rohmat	3	4	3	4	3	4	3	4	3	4	87,5
22	Indah	3	3	3	3	3	4	4	4	4	4	87,5
23	Nabila	4	4	4	4	3	4	3	3	3	4	90
24	Sherly	3	3	3	4	4	4	4	4	4	4	92,5
25	Devita	4	3	4	3	4	3	4	3	4	3	87,5

TABEL 3
(Daftar Pertanyaan)

No	Pertanyaan
1	Saya merasa akan sering menggunakan aplikasi Transliterasi Aksara Jawa ini
2	Saya merasa aplikasi ini tidak terlalu rumit untuk digunakan.
3	Saya merasa aplikasi ini mudah digunakan untuk proses transliterasi aksara Jawa.
4	Saya merasa tidak membutuhkan bantuan teknis untuk dapat menggunakan aplikasi ini.
5	Saya merasa aplikasi ini menghasilkan output yang konsisten.
6	Saya merasa sebagian besar orang akan dapat belajar menggunakan aplikasi ini dengan cepat.
7	Saya tidak merasa bingung untuk menggunakan aplikasi ini.
8	Saya merasa terbantu saat menggunakan aplikasi transliterasi Aksara Jawa ini.
9	Saya tidak perlu mempelajari banyak hal sebelum dapat menggunakan aplikasi ini dengan baik.
10	Saya merasa fitur-fitur yang ada pada aplikasi ini terintegrasi dengan baik.

Pengujian System Usability Scale (SUS) digunakan untuk mengevaluasi tingkat kemudahan penggunaan aplikasi, dengan tujuan mengukur efektivitas, efisiensi, dan pengalaman pengguna dalam berinteraksi dengan sistem. Pengujian dilakukan melalui survei daring menggunakan Google Form yang berisi 10 pertanyaan berskala Likert, diisi oleh 25 responden pengguna Android yang memahami penggunaan aksara Jawa. Skor dari setiap respon dikonversi ke rentang 0–100 dengan mengalikan hasil penilaian akhir dengan faktor 2,5.

V. KESIMPULAN

Penelitian ini berhasil mengembangkan aplikasi transliterasi aksara Jawa berbasis Android dengan mengintegrasikan teknologi *Natural Language Processing* (NLP) dan *Optical Character Recognition* (OCR) untuk mendukung pelestarian warisan budaya di era digital.

Fitur transliterasi teks yang memanfaatkan model Gemini AI mampu menghasilkan keluaran aksara Jawa yang konsisten dan sesuai kaidah penulisan tanpa memerlukan pelatihan ulang lokal. Sementara itu, fitur OCR berbasis *Convolutional Neural Network* (CNN) yang dilatih dengan dataset digital dan tulisan tangan mencapai akurasi pengenalan karakter Hanacaraka sebesar 96,25%, menunjukkan kinerja yang andal dalam mengenali pola visual dari gambar.

Kombinasi kedua metode ini menghasilkan sistem yang interaktif, efektif, dan mudah diakses oleh masyarakat, baik untuk keperluan pembelajaran maupun pelestarian aksara Jawa. Dengan demikian, aplikasi ini tidak hanya memberikan solusi teknis untuk transliterasi, tetapi juga memiliki potensi sebagai media edukasi dan promosi budaya yang relevan dengan perkembangan teknologi masa kini.

REFERENSI

- [1] S. A. Nugroho and R. Wulandari, "Pengembangan media pembelajaran aksara Jawa interaktif berbasis multimedia (studi kasus SD Mardi Rahayu 01 Ungaran)," *Jurnal Ilmiah Komputer Grafis*, vol. 13, 2020.
- [2] Abdul Robby, G., Tandra, A., Susanto, I., Harefa, J., & Chowanda, A. (2019). Implementation of optical character recognition using tesseract with the javanese script target in android application. *Procedia Computer Science*, 157, 499-505.
- [3] Sari, C. A., Kuncoro, M. W., Setiadi, D. R. I. M., & Rachmawanto, E. H. (2018). Roundness and eccentricity feature extraction for Javanese handwritten character recognition based on K-nearest

neighbor. In 2018 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI) (pp. 5-10).

- [4] M. I. Maulana, "Menurunnya Minat Tulisan Aksara Jawa di Lingkup Mahasiswa," *Heuristik: Jurnal Pendidikan Sejarah*, vol. 3, no. 2, pp. 62-76, Agustus 2023
- [5] Nulis Aksara Jawa. (n.d.). Google Play Store. Retrieved from https://play.google.com/store/apps/details?id=com.mu_hfau.nulisjawa
- [6] Nulis Aksara. (n.d.). Google Play Store. Retrieved from https://play.google.com/store/apps/details?id=com.gar_nesapps.aksara
- [7] Hana. (n.d.). Google Play Store. Retrieved from https://play.google.com/store/apps/details?id=dansvel_hana
- [8] R. Hidayatullah, H. C. Nugroho, dan R. Arifin, "Pengenalan Karakter Optikal Aksara Jawa Menggunakan Connectionist Temporal Classification (CTC)," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 8, no. 1, pp. 171–177, Jan. 2024, doi: 10.29207/resti.v8i1.5750. <https://jurnal.umpwr.ac.id/intek/article/view/5750/2416>
- [9] S. Sutarman and E. Yulianto, "Transliterasi Aksara Jawa ke Latin Menggunakan Metode Rule-Based," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 4, no. 2, pp. 123–130, 2017. https://scholar.google.com/scholar_lookup?title=Transliterasi+Aksara+Jawa+ke+Latin+Menggunakan+Metode+Rule-Based&author=Sutarman&publication_year=2017
- [10] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *Journal of Big Data*, vol. 6, no. 1, pp. 1–48, 2019. <https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0197-0>
- [11] Google DeepMind. (2023). Gemini: Our most capable and general AI model.
- [12] Kesaulya, G. N. A., Fariza, A., & Karlita, T. (2022). Javanese Script Text Image Recognition Using Convolutional Neural Networks. In IES 2022 - 2022 International Electronics Symposium: Energy Development for Climate Change Solution and Clean Energy Transition, Proceeding (pp. 534-539). IEEE. <https://doi.org/10.1109/IES55876.2022.9888527>

- [13] A. S. Putra, I. S. Wibowo, dan D. A. Nugroho, "Penerapan Optical Character Recognition untuk Pengenalan Karakter," *Jurnal Buana Informatika*, vol. 15, no. 1, pp. 69-78, Apr. 2024. <https://ojs.uajy.ac.id/index.php/jbi/article/view/9159/3521>
- [14] A. P. Santoso dan R. H. Nugroho, "Pengklasifikasian Aksara Jawa Metode Convolutional Neural Network," *Jurnal Teknologi Informasi dan Sistem Informasi*, vol. 8, no. 1, 2020. <https://jurnal.mdp.ac.id/index.php/jatasi/article/download/2177/961>
- [15] T. L. M. Suryanto, W. N. Simarmata, dan A. Farqi, "System Usability Scale (SUS) sebagai Metode Pengujian Kegunaan pada Situs Program Studi," *Prosiding Seminar Nasional Teknologi Informasi*, pp. 285–290, 2022. <https://sitasi.upnjatim.ac.id/index.php/sitasi/article/download/314/61>
- [16] T. W. Ramdhani, I. Budi, dan B. Purwandari, "Optical Character Recognition Engines Performance Comparison in Information Extraction," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 8, pp. 120-127, 2021. <https://scispace.com/pdf/optical-character-recognition-engines-performance-comparison-2uucfxppfp.pdf>