

Sistem Konversi Tulisan Tangan Huruf Atau Angka yang Ditampilkan Menjadi Teks dan Suara Secara Jelas

Virgiawan Ramadhani¹, Dynu Anggoro Pangestu²

^{1,2}Teknik Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Nusantara PGRI Kediri

E-mail: ¹ibrahimibu84@gmail.com, ²dynuap23@gmail.com

Abstrak – Penelitian ini membahas pengembangan sistem konversi tulisan tangan menjadi teks digital dan suara menggunakan metode Optical Character Recognition (OCR) berbasis Tesseract serta integrasi text-to-speech Bahasa Indonesia. Latar belakang penelitian muncul dari kebutuhan digitalisasi data tulis tangan yang masih banyak digunakan di dunia pendidikan, administrasi, dan pencatatan manual. Sistem ini dirancang untuk mengenali huruf maupun angka yang ditulis tangan, kemudian hasil pengenalan tersebut ditampilkan sebagai teks dan dikonversi menjadi suara yang jelas. Tahapan penelitian mencakup preprocessing citra, perbaikan kemiringan (deskew), segmentasi, OCR, dan konversi suara. Pengujian dilakukan pada dataset tulisan tangan huruf A–Z dan angka 0–9 dengan berbagai variasi gaya tulisan. Hasil menunjukkan bahwa sistem mampu mengenali karakter tulisan tangan dengan baik dan menghasilkan suara yang sesuai dengan hasil teks. Sistem ini memberikan solusi praktis untuk membantu digitalisasi dokumen serta membantu pengguna yang membutuhkan akses informasi dalam bentuk audio. Penelitian ini diharapkan dapat dikembangkan lebih lanjut dengan integrasi model CNN untuk meningkatkan akurasi pengenalan..

Kata Kunci- OCR, tulisan tangan, konversi teks, text-to-speech, pengolahan citra

1. PENDAHULUAN

Perkembangan teknologi digital telah mendorong transformasi data dari bentuk manual menuju format elektronik. Salah satu tantangan dalam proses digitalisasi adalah konversi tulisan tangan yang memiliki variasi bentuk, ukuran, kemiringan, dan gaya penulisan[1]. Tulisan tangan (*handwriting*) merupakan bentuk input yang masih banyak digunakan dalam aktivitas sehari-hari, seperti formulir, catatan, arsip administratif, dan proses pembelajaran di sekolah. Namun, proses digitalisasi dokumen tulisan tangan sering membutuhkan waktu lama apabila dilakukan secara manual. Oleh karena itu, dibutuhkan suatu sistem otomatis yang mampu mengenali tulisan tangan dengan cepat dan akurat[2].

Metode Optical Character Recognition (OCR) menjadi salah satu pendekatan populer dalam mengenali karakter dari citra. Tesseract merupakan salah satu OCR engine yang banyak digunakan karena bersifat open-source dan mendukung berbagai bahasa termasuk Bahasa Indonesia[3]. Akan tetapi, kinerja OCR sangat dipengaruhi oleh kualitas citra, sehingga preprocessing seperti peningkatan kontras, penghilangan noise, dan perbaikan kemiringan sangat diperlukan[4].

Penelitian-penelitian sebelumnya menunjukkan bahwa integrasi OCR dengan text-to-speech (TTS) dapat meningkatkan aksesibilitas bagi pengguna yang memiliki keterbatasan penglihatan dan kebutuhan edukasi berbasis audio[5]. Beberapa studi tentang handwriting recognition telah memanfaatkan metode *machine learning* dan *deep learning*, namun implementasi sederhana berbasis Tesseract dengan perbaikan citra yang optimal tetap relevan untuk aplikasi skala ringan[6].

Melihat kebutuhan tersebut, penelitian ini mengembangkan sistem yang mampu mengonversi gambar tulisan tangan menjadi teks serta menghasilkan suara Bahasa Indonesia yang jelas. Sistem ini diimplementasikan berbasis aplikasi web menggunakan Python dan Streamlit..

Penelitian terkait pengenalan tulisan tangan juga telah banyak dikembangkan di lingkungan akademik nasional, termasuk pada jurnal-jurnal terbitan internal perguruan tinggi. Salah satunya adalah penelitian yang dipublikasikan pada jurnal STAINS Universitas Nusantara PGRI Kediri yang membahas penerapan teknik pengolahan citra dan OCR untuk meningkatkan akurasi pengenalan karakter pada dokumen tidak terstruktur. Hasil penelitian tersebut menunjukkan bahwa tahapan *preprocessing* citra memiliki peran signifikan dalam menekan tingkat kesalahan pengenalan karakter tulisan tangan[7].

Selain itu, beberapa penelitian terkini menunjukkan bahwa penggabungan metode OCR dengan pendekatan *deep learning*, khususnya *Convolutional Neural Network* (CNN), mampu meningkatkan kemampuan sistem dalam mengenali variasi bentuk tulisan tangan yang kompleks. CNN terbukti efektif dalam mengekstraksi fitur visual penting dari citra karakter, sehingga cocok diterapkan pada sistem pengenalan tulisan tangan berbasis citra digital[8]. Namun demikian, kompleksitas model CNN menuntut ketersediaan *dataset* yang memadai serta proses komputasi yang relatif lebih tinggi.

Penelitian lain juga menekankan pentingnya integrasi sistem pengenalan teks dengan teknologi *Text-to-Speech* (TTS) untuk meningkatkan aksesibilitas informasi, terutama bagi penyandang disabilitas penglihatan. Sistem yang mampu menyajikan informasi tidak hanya dalam bentuk teks, tetapi juga suara, dinilai lebih inklusif dan aplikatif dalam konteks pendidikan dan layanan publik[9]. Oleh karena itu, pengembangan sistem konversi tulisan tangan yang terintegrasi dengan TTS menjadi solusi yang relevan dengan kebutuhan pengguna saat ini.

Berdasarkan beberapa penelitian tersebut, masih diperlukan pengembangan sistem yang mampu mengombinasikan *preprocessing* citra, pengenalan tulisan tangan, serta konversi suara dalam satu platform yang terintegrasi dan mudah digunakan. Penelitian ini berfokus pada pengembangan sistem konversi tulisan tangan berbasis web yang menggabungkan OCR, CNN, dan *Text-to-Speech* untuk menghasilkan teks dan suara Bahasa Indonesia secara otomatis dan efisien[10].

2. METODE PENELITIAN

Metode penelitian ini mencakup perancangan sistem, analisis arsitektur, pemilihan dan penerapan algoritma, hingga perancangan alur kerja sistem.

2.1 Analisis Kebutuhan dan Permasalahan

1. Permasalahan utama yang diidentifikasi adalah variasi bentuk tulisan tangan yang menyebabkan sistem sulit mengenali karakter huruf atau angka secara akurat.
2. Kualitas gambar tulisan tangan yang kurang baik (buram, banyak *noise*, atau pencahayaan tidak merata) menghambat proses OCR.
3. Metode OCR tradisional belum optimal, sehingga dibutuhkan pendekatan berbasis CNN untuk meningkatkan akurasi.
4. Diperlukan sistem terintegrasi yang menggabungkan tahap *preprocessing* citra, pengenalan pola tulisan tangan menggunakan *Convolutional Neural Network* (CNN), proses konversi hasil pengenalan menjadi teks digital melalui mekanisme *Optical Character Recognition* (OCR), verifikasi hasil oleh pengguna, serta konversi teks menjadi suara menggunakan *Text-to-Speech* (TTS) dalam satu alur kerja yang efisien.

2.2 Arsitektur Sistem dan Metode

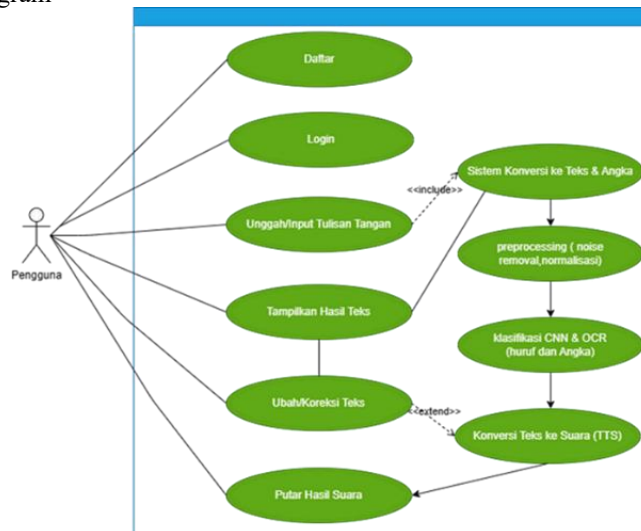
Arsitektur sistem dibangun di atas tiga modul utama: Antarmuka (UI), Modul OCR (Sistem Konversi), dan Modul TTS.

1. Pengenalan Karakter: Digunakan metode OCR berbasis *Convolutional Neural Network* (CNN). CNN dipilih karena unggul dalam mempelajari fitur-fitur visual yang kompleks dan dinamis dari tulisan tangan.
2. Tahap Pra-pemrosesan (*Preprocessing*): Diterapkan teknik *preprocessing* citra seperti noise removal dan normalisasi untuk meningkatkan kualitas gambar tulisan tangan, sehingga proses identifikasi karakter menjadi lebih optimal.
3. Konversi Suara: Integrasi dengan Modul TTS untuk mengonversi teks yang telah diverifikasi menjadi *output* suara yang jelas, sesuai dengan fungsi *Text-to-Speech*.

2.3 Perancangan dengan UML

Perancangan sistem di dokumentasikan menggunakan *Unified Modeling Language* (UML) untuk memvisualisasikan arsitektur dan alur kerja.

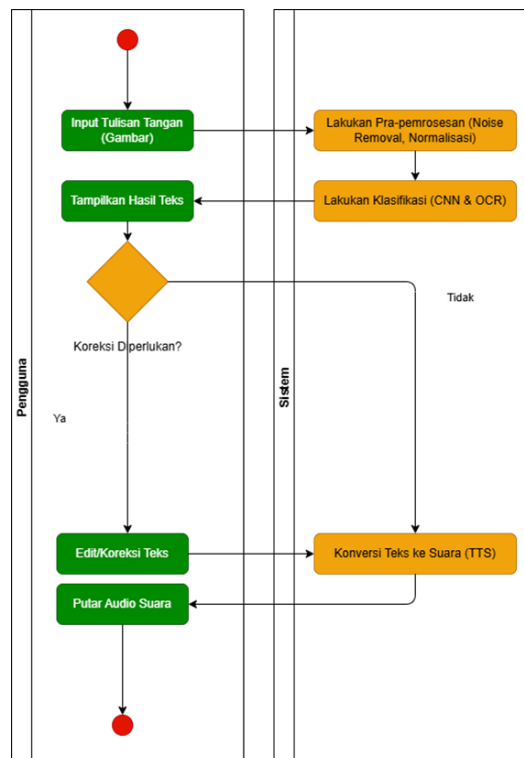
1. Use Case Diagram



Gambar 1. Use Case Diagram

Gambar 1 merupakan *Use Case* menggambarkan fungsionalitas sistem dari sudut pandang Pengguna. Diagram ini memvisualisasikan semua kapabilitas sistem dan interaksi antara Aktor (Pengguna) dan Fungsi Sistem (*Use Case*). Menunjukkan alur utama yang harus dijalankan pengguna untuk mencapai tujuan konversi tulisan tangan menjadi teks dan suara.

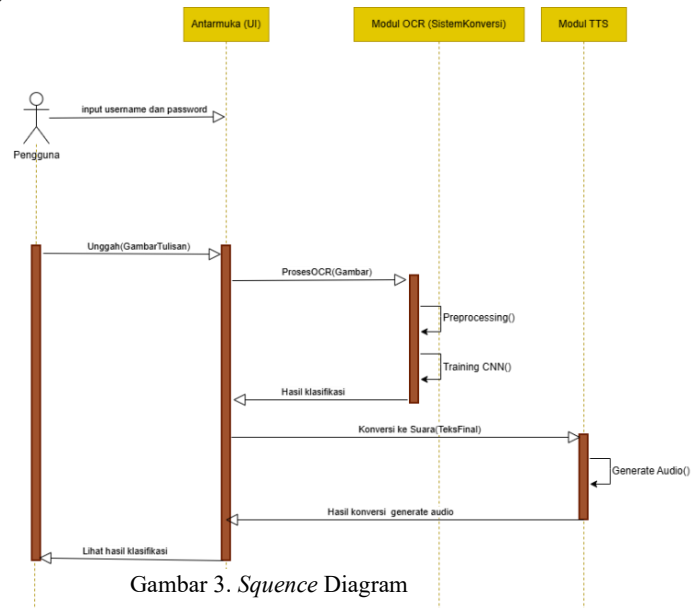
2. Activity Diagram



Gambar 2. Activity Diagram

Gambar 2 merupakan *Activity Diagram* yang menggambarkan proses pengenalan tulisan tangan dan konversinya menjadi teks serta suara, dibagi menjadi interaksi antara Pengguna dan Sistem. menunjukkan alur kerja sistem, dimulai dari Input Tulisan Tangan (Gambar) oleh pengguna, dilanjutkan dengan proses sistem Pra-pemrosesan (*Noise Removal*, Normalisasi) dan Klasifikasi (CNN & OCR). Hasilnya akan Tampilkan Hasil Teks kepada pengguna, yang memiliki titik keputusan untuk Edit/Koreksi Teks jika teks tidak sesuai, sebelum akhirnya dikonversi menjadi suara.

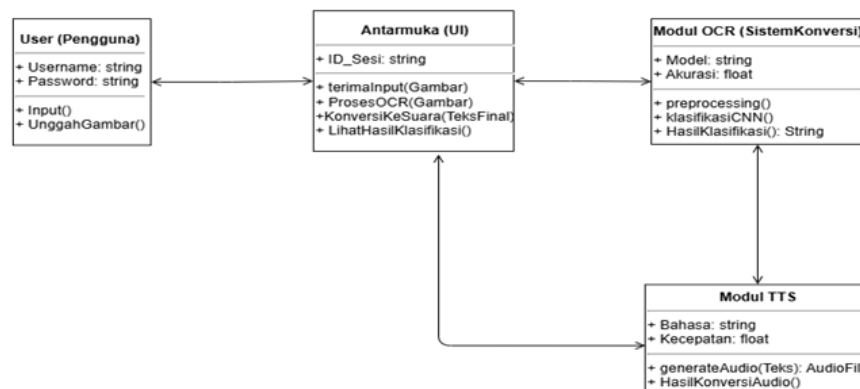
3. Sequence Diagram



Gambar 3. Sequence Diagram

Gambar 3 merupakan *Sequence Diagram* yang membagi alur proses menjadi interaksi dinamis antara Antarmuka (UI), Modul OCR (Sistem Konversi), dan Modul TTS³¹. Antarmuka (UI) bertindak sebagai penghubung, menerima input Unggah(Gambar Tulisan) dari pengguna dan meneruskannya sebagai perintah Proses OCR(Gambar) ke Modul OCR³². Modul OCR melakukan langkah *preprocessing()* dan klasifikasi *CNN()* secara internal sebelum mengembalikan Hasil Klasifikasi³³. Setelah verifikasi oleh pengguna, UI memicu Modul TTS dengan pesan Konversi ke Suara (TeksFinal) untuk menghasilkan audio³⁴.

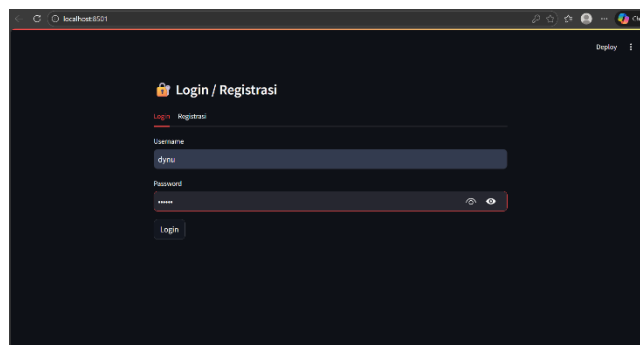
4. Class Diagram



Gambar 4. Class Diagram

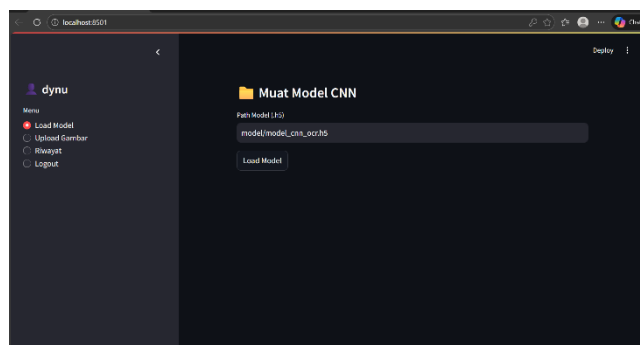
Gambar 4 *Class Diagram* menjelaskan Struktur data sistem yang dimodelkan dan memisahkan tanggung jawab menjadi empat kelas inti: User, GambarTulisan, SistemKonversi (Modul OCR), dan HasilKonversi³⁵. Relasi asosiasi yang jelas menjamin aliran data yang logis³⁶. Modul OCR (SistemKonversi) memiliki fungsi inti *preprocessing()* dan klasifikasi *CNN()*³⁷. Akurasi pengenalan model CNN dapat dievaluasi menggunakan persamaan seperti *Mean Squared Error (MSE)* dalam proses pelatihan.

2.4 Implementasi Sistem



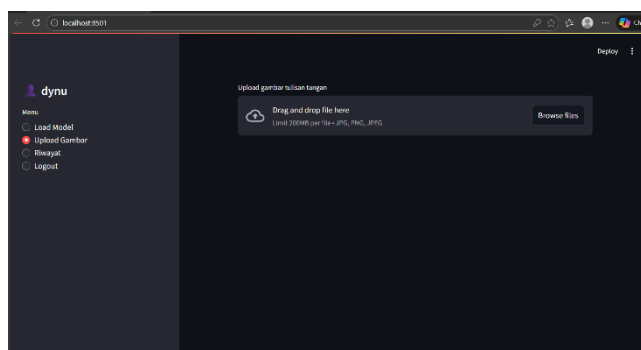
Gambar 5. registrasi

Gambar 5 *Login/Register*, menampilkan pengguna melakukan login atau registrasi untuk mengakses sistem. Fitur ini berfungsi untuk menjaga keamanan data serta memungkinkan setiap aktivitas pengguna tercatat secara personal.



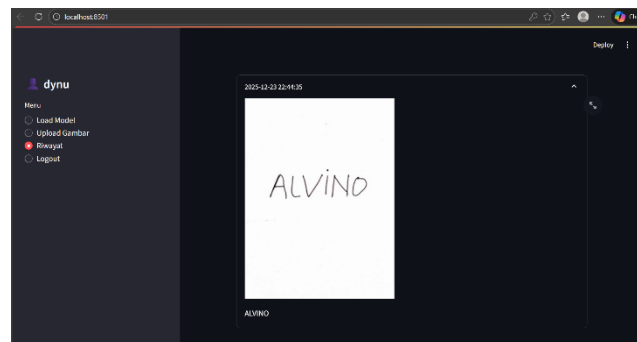
Gambar 6. load model

Gambar 6 *Load Model CNN* , Pada menu Load Model, pengguna dapat memuat model CNN yang telah dilatih sebelumnya dalam format file .h5. Model ini berperan sebagai komponen utama dalam proses pengenalan tulisan tangan.



Gambar 7. Upload gambar

Gambar 7, pengguna mengunggah gambar tulisan tangan dengan format JPG, PNG, atau JPEG. Gambar yang diunggah akan diproses dan disesuaikan dengan input model CNN sebelum dilakukan prediksi.



gambar 8. riwayat prediksi

Gambar 8 riwayat prediksi, Fitur Riwayat berfungsi untuk menyimpan dan menampilkan data hasil prediksi yang telah dilakukan oleh pengguna sebelumnya. Riwayat ini mencakup informasi seperti gambar yang diunggah, hasil prediksi, serta waktu prediksi. Fitur ini membantu pengguna dalam melakukan evaluasi, penelusuran ulang hasil prediksi, dan dokumentasi penggunaan sistem.

2.5 Rumus OCR

(SistemKonversi) memiliki fungsi inti `preprocessing()` dan `klasifikasiCNN()`. Akurasi pengenalan model CNN dapat dievaluasi menggunakan persamaan seperti *Mean Squared Error* (MSE) dalam proses pelatihan [3].

$$dEk = \frac{1}{m} + \sum_{i=1}^m (f(d_i, k) - k_i)^2 \text{ atau } dEk = \frac{1}{m} (f(d_i, k) - k_i)^2 \dots\dots\dots (1)$$

2.6 Skenario Uji Coba Komponen Sistem

Tabel 1. Skenario Uji Coba

Aspek Uji Coba	Parameter Masukan	Tujuan Uji
Input Citra	Kualitas (Buram, Bersih)	Optimalisasi <i>preprocessing</i>
Klasifikasi CNN	Variasi Tulisan Tangan	Evaluasi Akurasi (70 %)
Koreksi Pengguna	Kesalahan Karakter	Validasi Mekanisme Koreksi
Output Suara	Teks Hasil OCR	Kejelasan Audio (TTS)

Skenario uji coba sistem dilakukan untuk mengevaluasi kinerja setiap komponen utama, meliputi input citra, proses klasifikasi menggunakan CNN, mekanisme koreksi oleh pengguna, serta output suara berbasis *Text-to-Speech*. Rincian aspek uji coba, parameter masukan, dan tujuan pengujian ditunjukkan pada Tabel 1.

3. HASIL DAN PEMBAHASAN

Berdasarkan hasil uji coba, sistem mampu mengenali karakter tulisan tangan dengan tingkat akurasi sebesar 70%. Nilai akurasi tersebut menunjukkan bahwa sebagian besar karakter dapat dikenali dengan benar, namun masih terdapat kesalahan pengenalan, khususnya pada karakter yang memiliki bentuk visual serupa, seperti huruf “O” dengan angka “0” atau huruf “l” dengan angka “1”. Hal ini menunjukkan bahwa model CNN telah mampu menangkap pola dasar tulisan tangan, tetapi masih memiliki keterbatasan dalam membedakan karakter yang mirip secara visual.

Dari sisi kinerja waktu, proses sistem mulai dari unggah citra, *preprocessing*, hingga menghasilkan output teks membutuhkan waktu rata-rata 2 - 3 detik per citra. Waktu tersebut dipengaruhi oleh kualitas gambar input dan ukuran citra. Proses konversi teks ke suara menggunakan modul *Text-to-Speech* (TTS) berlangsung relatif cepat, dengan waktu kurang dari 1 detik, sehingga sistem masih dapat digunakan secara interaktif oleh pengguna.

Tahap *preprocessing* citra, yang meliputi penghilangan *noise* dan normalisasi, terbukti berpengaruh terhadap hasil pengenalan. Pengujian menunjukkan bahwa tanpa *preprocessing*, akurasi pengenalan berada pada kisaran yang lebih rendah. Setelah *preprocessing* diterapkan, kualitas citra input menjadi lebih seragam dan lebih mudah dikenali oleh model CNN, sehingga akurasi meningkat hingga mencapai 70%. Hal ini menegaskan bahwa *preprocessing* merupakan tahap penting dalam sistem pengenalan tulisan tangan.

Output sistem berupa teks hasil pengenalan ditampilkan kepada pengguna dan dapat dikoreksi secara manual apabila terjadi kesalahan. Setelah diverifikasi, teks tersebut dikonversi menjadi suara Bahasa Indonesia. Hasil pengujian menunjukkan bahwa suara yang dihasilkan sesuai dengan teks hasil pengenalan dan dapat dipahami dengan jelas, sehingga sistem tidak hanya berfungsi sebagai alat digitalisasi dokumen, tetapi juga sebagai media penyampaian informasi berbasis audio.

Secara keseluruhan, hasil pengujian menunjukkan bahwa sistem telah berfungsi sesuai dengan tujuan penelitian, meskipun masih memiliki keterbatasan dalam hal akurasi dan kemampuan membedakan karakter yang mirip. Keterbatasan ini terutama disebabkan oleh variasi gaya tulisan tangan dan keterbatasan jumlah serta keragaman data latih yang digunakan.

4. SIMPULAN

Berdasarkan hasil pengujian yang telah dilakukan, dapat disimpulkan bahwa sistem konversi tulisan tangan berbasis *Convolutional Neural Network* (CNN) berhasil diimplementasikan dan mampu mengenali karakter tulisan tangan berupa huruf dan angka. Sistem memperoleh tingkat akurasi pengenalan sebesar **70%**, yang menunjukkan bahwa model telah mampu mengenali sebagian besar karakter tulisan tangan dengan baik, meskipun masih terdapat kesalahan pada karakter yang memiliki bentuk visual serupa.

Penerapan tahapan *preprocessing* citra, khususnya penghilangan *noise* dan normalisasi, terbukti berperan penting dalam meningkatkan kualitas citra input dan stabilitas hasil pengenalan. Hasil pengujian menunjukkan bahwa akurasi sistem meningkat setelah *preprocessing* diterapkan dibandingkan dengan pengujian tanpa prapemrosesan citra.

Dari sisi kinerja, sistem mampu menghasilkan *output* teks dalam waktu rata-rata 2–3 detik per citra, serta menghasilkan *output* suara melalui modul *Text-to-Speech* dalam waktu kurang dari 1 detik. Hal ini menunjukkan bahwa sistem cukup responsif untuk digunakan dalam proses digitalisasi dokumen tulisan tangan secara *non real-time*.

Secara keseluruhan, sistem telah memenuhi tujuan penelitian dalam mengonversi tulisan tangan menjadi teks dan suara Bahasa Indonesia. Namun demikian, sistem masih memiliki keterbatasan dalam hal akurasi pengenalan dan kemampuan membedakan karakter yang memiliki kemiripan bentuk. Keterbatasan ini menunjukkan perlunya pengembangan lebih lanjut, terutama pada peningkatan kualitas *dataset* dan optimasi model, agar performa sistem dapat ditingkatkan di masa mendatang.

5. SARAN

1. Pengembangan Dataset dan *Robustness* Model: Diperlukan pengembangan dataset tulisan tangan yang lebih masif dan beragam, termasuk variasi gaya penulisan non-standar atau kondisi pencahayaan ekstrem, untuk melatih model CNN agar lebih tangguh (*robust*) terhadap segala jenis input.
2. Eksplorasi Arsitektur Sekuensial: Menguji kombinasi CNN dengan *Recurrent Neural Network* (RNN) atau arsitektur Transformer untuk memungkinkan pengenalan tulisan tangan berkesinambungan (*cursive*), sehingga sistem dapat memproses kata atau kalimat secara utuh, tidak hanya per karakter.
3. Peningkatan Fungsionalitas *Real-Time*: Penelitian selanjutnya harus berfokus pada implementasi sistem pada platform mobile dengan kemampuan deteksi dan konversi tulisan tangan *real-time* dari kamera, sekaligus mengoptimalkan kecepatan komputasi.
4. Penyempurnaan Modul Suara TTS: Menambahkan opsi kustomisasi suara (gender, aksen) dan mengintegrasikan model TTS yang lebih canggih untuk menghasilkan output audio dengan intonasi dan emosi yang lebih natural dan dinamis.

DAFTAR PUSTAKA

- [1] Aksenta, A., Irmawati, I., Ridwan, A., Hayati, N., Sepriano, S., Herlinah, H., ... & Ginting, T. W. (2023). *Literasi Digital: Pengetahuan & Transformasi Terkini Teknologi Digital Era Industri 4.0 dan Society 5.0*. PT. Sonpedia Publishing Indonesia.
- [2] Soemantoro, Achsinifina H. *Mengenal potensi anak melalui tulisan tangan*. Grasindo, 2009.
- [3] Haisar, F. (2018). Klasifikasi Analisis Sentimen Meme Dengan Metode Optical Character Recognition (OCR) dan Algoritma Naive Bayes. *Pengemb. Teknol. Inf. dan Ilmu Komput*, 2(8).
- [4] Prasetya, M. I., Yadi, I. Z., Kunang, Y. N., & Permatasari, S. D. (2025). Prapemrosesan untuk Klasifikasi Gambar Aksara OKU Timur. *Jurnal Teknologi Dan Sistem Informasi Bisnis*, 7(1), 208–215.
- [5] Yaqin, M. A. (2025). *PENGARUH PEMANFAATAN ARTIFICIAL INTELLIGENCE TERHADAP AKSESIBILITAS LITERASI DIGITAL PENYANDANG DISABILITAS TUNANETRA BERBASIS APLIKASI SCREEN READER PADA SMARTPHONE* (Doctoral dissertation, Universitas Islam Sultan Agung Semarang).

-
- [6] Abdiansah, L., Sumarno, S., Eviyanti, A., & Azizah, N. L. (2025). Penerapan Algoritma Convolutional Neural Networks untuk Pengenalan Tulisan Tangan Aksara Jawa: Implementation of Convolutional Neural Networks Algorithm for Javanese Handwriting Recognition. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 5(2), 496-504.
 - [7] Amartama, Salis Nilam, Alvi Nurul Hidayah, Putri Kartika Sari, and Risky Aswi Ramadhani. 2024. “Implementasi *Convolutional Neural Network* (CNN) dalam Pengenalan Pola Penulisan Tangan. (2024). *Seminar Nasional Teknologi & Sains*, 3(1), 133-138.
 - [8] Pujoseno, J., & Nugroho, R. (2021). Pengenalan Karakter Tulisan Tangan Menggunakan Convolutional Neural Network. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 8(4), 743–750.
 - [9] Yaqin, M. A., & Hidayat, R. (2023). Integrasi Text-to-Speech untuk Meningkatkan Aksesibilitas Sistem Informasi Digital. *Jurnal Sistem Informasi*, 19(1), 55–63.
 - [10] Permatasari, S. D., & Kunang, Y. N. (2024). Pengembangan Sistem OCR Terintegrasi Berbasis Web untuk Digitalisasi Dokumen Tidak Terstruktur. *Journal of Computer Science and Informatics*, 5(1), 12–21.