

Analisis Kinerja Model CNN-LSTM Berbasis *Optical Character Recognition* untuk Ekstraksi Informasi e-KTP Berdasarkan Kategori Teks

Fauzan Ihza Fajar, Selly Anastassia Amellia Kharis, dan Asmara Iriani Tarigan



Volume 14, Issue 1, Pages 90–102, April 2026

Diterima 28 Januari 2026, Direvisi 10 April 2026, Disetujui 14 April 2026, Diterbitkan 18 April 2026

To Cite this Article : F. I. Fajar, S. A. A. Kharis, dan A. I. Tarigan, "Analisis Kinerja Model CNN-LSTM Berbasis *Optical Character Recognition* untuk Ekstraksi Informasi e-KTP Berdasarkan Kategori Teks", *Euler J. Ilm. Mat. Sains dan Teknol.*, vol. 14, no. 1, pp. 90–102, 2026, <https://doi.org/10.37905/euler.v14i1.37593>

© 2026 by author(s)

JOURNAL INFO • EULER : JURNAL ILMIAH MATEMATIKA, SAINS DAN TEKNOLOGI



	Homepage	:	http://ejournal.ung.ac.id/index.php/euler/index
	Journal Abbreviation	:	Euler J. Ilm. Mat. Sains dan Teknol.
	Frequency	:	Three times a year
	Publication Language	:	English (preferable), Indonesia
	DOI	:	https://doi.org/10.37905/euler
	Online ISSN	:	2776-3706
	License	:	Creative Commons Attribution-NonCommercial 4.0 International License
	Publisher	:	Department of Mathematics, Universitas Negeri Gorontalo
	Country	:	Indonesia
	OAI Address	:	http://ejournal.ung.ac.id/index.php/euler/oai
	Google Scholar ID	:	QF_r-gAAAAJ
	Email	:	euler@ung.ac.id

JAMBURA JOURNAL • FIND OUR OTHER JOURNALS



Jambura Journal of Biomathematics



Jambura Journal of Mathematics



Jambura Journal of Mathematics Education



Jambura Journal of Probability and Statistics

Analisis Kinerja Model CNN-LSTM Berbasis *Optical Character Recognition* untuk Ekstraksi Informasi e-KTP Berdasarkan Kategori Teks

Fauzan Ihza Fajar¹, Selly Anastassia Amellia Kharis^{1,*}, Asmara Iriani Tarigan¹

¹Program Studi Matematika, Universitas Terbuka, Tangerang Selatan 15437, Indonesia

ARTICLE HISTORY

Diterima 28 Januari 2026
Direvisi 10 April 2026
Disetujui 14 April 2026
Diterbitkan 18 April 2026

KATA KUNCI

Connectionist Temporal Classification
Convolutional Neural Network
e-KTP
Long Short-Term Memory
Optical Character Recognition

KEYWORDS

Connectionist Temporal
Classification
Convolutional Neural Network
e-KTP
Long Short-Term Memory
Optical Character Recognition

ABSTRAK. Teknologi *Optical Character Recognition* (OCR) berperan penting dalam otomatisasi ekstraksi informasi dari dokumen identitas seperti Kartu Tanda Penduduk Elektronik (e-KTP). Namun, pengenalan teks yang panjang dan variasi karakter yang kompleks masih menjadi tantangan saat ini. Hal ini dapat menyebabkan tingkat kesalahan yang tinggi. Penelitian ini bertujuan untuk mengatasi keterbatasan tersebut dengan mengeksplorasi model OCR berbasis *deep learning* yang mengombinasikan Convolutional Neural Network (CNN), Long Term Short Memory (LSTM), dan Connectionist Temporal Classification (CTC) untuk mengatasi permasalahan segmentasi karakter dan meningkatkan akurasi pengenalan teks secara end-to-end. CNN berfungsi mengekstraksi fitur visual karakter, LSTM memahami konteks urutan karakter, dan CTC memungkinkan pengenalan teks tanpa segmentasi manual. Kontribusi utama penelitian ini terletak pada analisis kinerja model CNN-LSTM berbasis CTC dalam mengekstraksi informasi e-KTP berdasarkan kategori teks dengan tingkat kompleksitas berbeda, yaitu Tempat dan Tanggal Lahir (TTL), Nama, dan Nomor Induk Kependudukan (NIK). Evaluasi dilakukan menggunakan metrik Character Error Rate (CER) untuk mengukur akurasi pengenalan karakter secara detail. Hasil eksperimen menunjukkan bahwa model memberikan performa terbaik pada kategori TTL dengan CER sebesar 0,84%, NIK sebesar 1,29%, dan Nama sebesar 4,33% yang mengindikasikan tantangan lebih tinggi pada teks dengan variasi karakter yang kompleks. Temuan ini menunjukkan bahwa performa model sangat dipengaruhi oleh karakteristik data teks, khususnya tingkat keragaman dan panjang sekuens. Secara keseluruhan, penelitian ini menunjukkan bahwa pendekatan CNN-LSTM berbasis CTC efektif dalam melakukan ekstraksi informasi pada dokumen e-KTP secara end-to-end, serta memberikan wawasan empiris mengenai perbedaan performa model pada berbagai kategori teks sebagai dasar pengembangan metode OCR yang lebih adaptif terhadap kompleksitas data.

ABSTRACT. *Optical Character Recognition* (OCR) technology plays an important role in automating information extraction from identity documents such as the Electronic Identity Card (e-KTP). However, recognizing long text sequences and handling complex character variations remain significant challenges. These issues can lead to high error rates. This study aims to address these limitations by exploring a deep learning-based OCR model that integrates Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), and Connectionist Temporal Classification (CTC) in an end-to-end framework without explicit character segmentation. CNN is employed to extract visual features, LSTM captures sequential dependencies, and CTC enables flexible alignment between input images and output text. The main contribution of this study lies in analysing the performance of a CNN-LSTM model with CTC in extracting e-KTP information across text categories with different complexity levels, namely Date and Place of Birth (TTL), name, and national identification number (NIK). Performance is evaluated using the Character Error Rate (CER). The results show that the model achieves the best performance on TTL with a CER of 0.84%, followed by NIK at 1.29%, and Name at 4.33% indicating higher difficulty in recognizing more complex text patterns. These findings demonstrate that model performance is influenced by text characteristics, particularly variability and sequence length. Overall, the proposed approach is effective for end-to-end e-KTP information extraction and provides insights for developing more adaptive OCR models.



This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution-NonCommercial 4.0 International License. **Editorial of EULER:** Department of Mathematics, Universitas Negeri Gorontalo, Jln. Prof. Dr. Ing. B. J. Habibie, Bone Bolango 96554, Indonesia.

*Penulis Korespondensi.

1. Pendahuluan

Optical Character Recognition (OCR) merupakan teknologi yang memberikan kemampuan kepada komputer untuk mengenali dan mengekstraksi teks dari gambar atau dokumen yang dipindai, baik dari tulisan cetak maupun tulisan tangan [1, 2]. Menurut Smith [3], *OCR* adalah proses otomatis untuk diubah ke dalam format digital yang bisa disunting dan dianalisis oleh komputer. Teknologi ini berperan secara signifikan dalam proses digitalisasi dokumen, terutama dalam konteks otomasi identifikasi termasuk proses pengecekan data personal [4]. Kartu Tanda Penduduk Elektronik (e-KTP) di Indonesia menjadi salah satu dokumen utama yang menyimpan data kependudukan secara elektronik serta berperan penting sebagai identitas resmi [5]. Karena hal tersebut, diperlukan sistem *OCR* yang andal dan adaptif dalam mengekstraksi informasi dari e-KTP menjadi semakin penting, walaupun penerapan *OCR* pada e-KTP menghadapi berbagai tantangan. Beragamnya tata letak teks, perbedaan posisi dan ukuran *font*, serta kondisi pencahayaan saat proses pemindaian dapat mengganggu akurasi sistem. Selain itu, struktur informasi pada e-KTP bersifat semi-terstruktur dengan kombinasi teks pendek dan panjang sehingga meningkatkan kompleksitas dalam proses ekstraksi informasi secara otomatis. Shi, Bai, dan Yao [6] menyebutkan bahwa model *OCR* yang dirancang untuk mengenali urutan teks yang kompleks dapat mengatasi sebagian dari tantangan ini, terutama jika sistem dilatih pada data dengan variasi visual yang tinggi.

Sejumlah penelitian telah mengembangkan pendekatan *OCR* berbasis *deep learning* untuk mengatasi permasalahan tersebut. Pendekatan berbasis *Convolutional Neural Network (CNN)* terbukti efektif dalam mengekstraksi fitur visual dari citra dokumen, terutama untuk teks dengan format yang relatif konsisten [7]. Namun, untuk dokumen seperti e-KTP yang memuat informasi dalam bentuk sekuensial, *CNN* saja tidak cukup. Pendekatan solusi yang lebih efisien dilakukan dengan mengombinasikan *CNN* dengan *Long Short-Term Memory (LSTM)*, di mana *CNN* bertugas mengekstraksi fitur visual dan *LSTM* menangani urutan karakter dalam konteks waktu [8, 9]. *LSTM*, sebagai salah satu varian *Recurrent Neural Network (RNN)*, memiliki kemampuan untuk mengingat informasi jangka panjang dan mengatasi permasalahan *vanishing gradient*, sehingga sangat cocok untuk memproses data sekuensial. Dalam arsitektur gabungan ini, fitur hasil ekstraksi *CNN* diubah menjadi representasi sekuensial yang kemudian diproses oleh *LSTM* untuk memprediksi urutan karakter secara lebih akurat. Kombinasi ini tidak hanya meningkatkan performa pengenalan teks, tetapi juga mampu menangani variasi panjang teks dan kompleksitas struktur dokumen [10]. Dengan demikian, integrasi *CNN* dan *LSTM* menjadi solusi yang efektif dan banyak diadopsi dalam pengembangan sistem *OCR* modern, khususnya untuk dokumen identitas seperti e-KTP [11, 12].

Fungsi *Connectionist Temporal Classification (CTC) Loss* digunakan untuk mengatasi masalah pelabelan data yang tidak selalu sejajar secara posisi atau panjang teks yang bervariasi. Metode ini memungkinkan sistem mengenali teks tanpa harus mengetahui posisi pasti setiap karakter, memberikan fleksibilitas lebih tinggi dalam menghadapi tata letak yang tidak seragam [13]. Evaluasi model *CNN-LSTM* yang dilatih dengan *CTC Loss* menunjukkan performa yang menjanjikan dalam konteks dokumen kompleks [14]. Meskipun demikian, sebagian besar penelitian sebelumnya masih berfokus pada peningkatan akurasi secara umum tanpa mempertimbangkan variasi karakteristik teks dalam satu dokumen, khususnya pada dokumen identitas seperti e-KTP yang memiliki beberapa kategori teks dengan tingkat kompleksitas berbeda (misalnya NIK, nama, dan TTL). Selain itu, evaluasi kinerja model sering kali dilakukan secara agregat, sehingga belum memberikan pemahaman mendalam mengenai performa model pada masing-masing kategori teks. Oleh karena itu, diperlukan kajian yang tidak hanya menerapkan arsitektur *CNN-LSTM-CTC*, tetapi juga menganalisis secara spesifik bagai-

mana model tersebut bekerja pada berbagai jenis informasi dalam e-KTP. Analisis ini penting untuk mengidentifikasi kelemahan model serta menentukan strategi peningkatan performa yang lebih terarah.

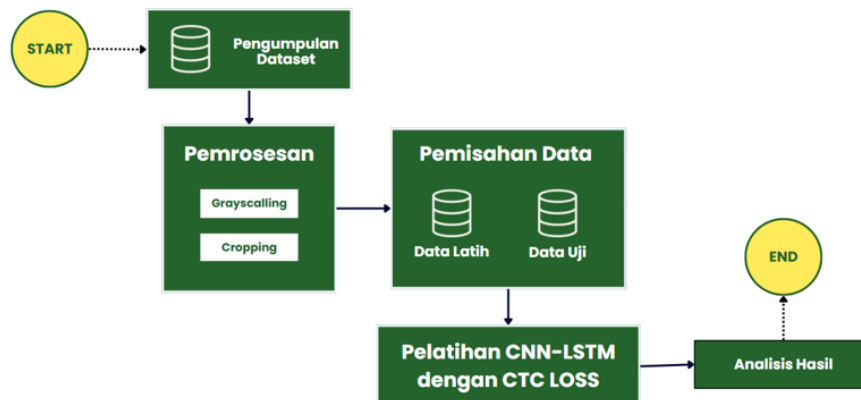
Studi mengenai OCR juga berkembang dalam berbagai bidang. Harahap *et al.* mengembangkan OCR untuk pembuatan aplikasi kalkulator tulisan tangan sederhana [15], sementara Banu *et al.* menekankan peran OCR dalam pendidikan interaktif [16]. OCR dapat meningkatkan aksesibilitas informasi, kemandirian pengguna, literasi *Braille*, serta membuka peluang pendidikan dan pekerjaan yang lebih luas [17]. Perkembangan OCR di luar ranah pendidikan juga terlihat dalam aplikasi administratif, seperti pada pengolahan data identitas [11]. Sugiarta *et al.* mengimplementasikan OCR berbasis CNN untuk mengekstraksi informasi dari e-KTP, dan menunjukkan bahwa arsitektur konvolusional efektif dalam mengenali teks yang memiliki format standar [16]. Selanjutnya, Panuntun dan Sejati mengintegrasikan CNN dengan OCR untuk sistem deteksi otomatis data KTP, menghasilkan akurasi tinggi sebesar 92% serta presisi penuh dalam pengenalan informasi [18]. Bahkan, pendekatan yang mengombinasikan CNN dan LSTM sebagaimana diteliti oleh Wijaya dan Lubis semakin menunjukkan potensi OCR modern dalam menangani variasi bentuk dan posisi teks melalui pelatihan berbasis dataset multibahasa [19]. Ketiga studi ini menunjukkan bahwa pendekatan OCR berbasis *deep learning* tidak hanya unggul dalam akurasi, tetapi juga adaptif terhadap konteks teks kompleks dan dokumen resmi—sebuah peluang yang relevan pula bila diterapkan pada sektor pendidikan yang mengandalkan dokumen fisik atau tulisan tangan sebagai media utama.

Berdasarkan latar belakang tersebut, penelitian ini berfokus pada pengembangan sistem OCR berbasis kombinasi CNN, LSTM, dan CTC Loss yang diadaptasi secara khusus untuk mengekstraksi informasi dari e-KTP. Kombinasi ini dipilih karena CNN efektif dalam mengekstraksi fitur spasial dari citra dokumen [20], sementara LSTM mampu menangkap dependensi urutan karakter dalam teks, sehingga sangat efektif dalam menangani teks yang tidak tersegmentasi secara eksplisit [21]. Penggunaan CTC Loss memungkinkan sistem untuk melakukan pelatihan tanpa memerlukan *alignmen* yang presisi antara input gambar dan label teks, menjadikannya ideal untuk data dengan struktur variatif seperti e-KTP. Pendekatan serupa telah diterapkan dalam penelitian oleh Panjalu dan Wisudawati [23] yang mengembangkan sistem pembacaan gerak bibir menggunakan CNN, Bi-LSTM, dan CTC Loss Function, mencapai akurasi sebesar 96,9%, *Word Error Rate* (WER) sebesar 0,66%, dan *Character Error Rate* (CER) sebesar 0,16%.

Penelitian ini bertujuan untuk menganalisis kinerja model OCR berbasis CNN-LSTM dengan CTC dalam mengekstraksi informasi e-KTP dan sejauh mana karakteristik teks memengaruhi tingkat akurasi pengenalan. Kontribusi utama penelitian ini adalah memberikan analisis empiris mengenai perbedaan kinerja model OCR pada berbagai kategori teks dalam e-KTP, yang belum banyak dibahas pada penelitian sebelumnya. Dengan pendekatan ini, diharapkan dapat dihasilkan sistem yang tidak hanya akurat, tetapi juga fleksibel terhadap variasi tampilan dokumen, serta dapat mendukung otomasi administrasi kependudukan secara lebih efisien di masa mendatang.

2. Metode

Pengembangan sistem OCR untuk pengenalan teks pada gambar memerlukan pemahaman mendalam terhadap beberapa teori dasar, khususnya pada teknologi pemrosesan gambar dan jaringan saraf tiruan. Penelitian ini dilakukan dengan tahapan pengumpulan dataset, pemrosesan data (*grayscale* dan *cropping*), pemisahan data menjadi data latih dan data uji, pelatihan model CNN-LSTM menggunakan CTC Loss, serta analisis hasil akhir, seperti yang ditunjukkan pada Gambar 1.



Gambar 1. Tahapan Penelitian

2.1. Pengumpulan Dataset

Pada tahap pertama, dilakukan pengumpulan data berupa gambar e-KTP yang berasal dari nasabah sebuah bank swasta di Indonesia. Dataset yang digunakan dalam penelitian ini berjumlah 2500 citra e-KTP. Jumlah tersebut telah ditentukan sejak awal penelitian untuk memastikan kecukupan data dalam mengevaluasi performa sistem OCR. Proses pengumpulan data dilakukan melalui kerja sama resmi dengan pihak bank, disertai perizinan yang sah dan jaminan perlindungan privasi nasabah, di mana semua data telah dianonimkan sebelum digunakan dalam penelitian. Data kemudian dikelompokkan ke dalam tiga kategori teks utama, yaitu Nomor Induk Kependudukan (NIK), Nama, dan Tempat dan Tanggal Lahir (TTL). Distribusi data pada masing-masing kategori dianalisis untuk memastikan representasi yang seimbang dalam proses pelatihan dan evaluasi model. Fokus pengumpulan data adalah untuk mengevaluasi karakteristik visual seperti ketepatan posisi gambar dan pencahayaan e-KTP, yang berpotensi memengaruhi akurasi ekstraksi teks. Setelah data berhasil dikumpulkan, tahap selanjutnya adalah pemrosesan gambar untuk meningkatkan kualitas input bagi sistem OCR.

2.2. Pemrosesan Gambar

Pemrosesan gambar (*preprocessing*) dilakukan untuk meningkatkan kualitas gambar KTP yang diambil dari kamera atau *scanner*, sehingga memudahkan proses ekstraksi teks dengan hasil yang lebih akurat dan konsisten. Pemrosesan gambar dilakukan dengan melakukan langkah-langkah sebagai berikut.

1. Konversi gambar ke skala abu-abu (*grayscale*) merupakan langkah awal dalam *preprocessing*, yaitu mengonversi gambar KTP menjadi skala abu-abu. Proses ini menyederhanakan data visual dengan menghilangkan elemen warna yang tidak diperlukan tanpa mengurangi informasi penting. Dengan mengurangi kompleksitas gambar, model dapat lebih mudah mengenali pola karakter yang relevan untuk tahap pengenalan teks.
2. *Cropping* area spesifik pada gambar dilakukan setelah dikonversi ke skala abu-abu. Gambar diproses lebih lanjut dengan melakukan *cropping* untuk mengekstraksi area tertentu, seperti NIK, Nama, dan Tempat Tanggal Lahir. Proses *cropping* dilakukan secara manual untuk memastikan hasil yang presisi dan menghindari kesalahan manusia. Pendekatan ini memastikan informasi penting dapat diisolasi dengan baik, sehingga meminimalkan elemen yang tidak relevan dan memaksimalkan kualitas data untuk tahapan ekstraksi berikutnya.

2.3. Pemisahan Data

Dataset yang digunakan dalam sistem OCR perlu dipersiapkan dan dibagi ke dalam beberapa subset untuk memastikan model dapat belajar secara optimal serta dievaluasi secara adil. Gambar e-KTP yang telah dikumpulkan diseleksi untuk memastikan kualitas dan kelayakan data. Setelah itu, dataset dari masing-masing kategori dibagi menjadi dua bagian utama yaitu data pelatihan (*training*) dan data pengujian (*testing*) dengan proporsi 80% untuk pelatihan dan 20% untuk pengujian. Untuk meningkatkan validitas hasil evaluasi, pembagian data dilakukan menggunakan teknik *k-Fold Cross Validation* dengan nilai $k = 5$. Pendekatan ini memungkinkan setiap data memperoleh peran sebagai data pelatihan dan pengujian secara bergantian dalam lima iterasi, sehingga mengurangi risiko bias dan memberikan gambaran performa model yang lebih stabil dan menyeluruh.

2.4. Pelatihan Data

Pelatihan model OCR pada penelitian ini menggunakan pendekatan berbasis *deep learning* menggunakan kombinasi CNN, LSTM, dan fungsi *loss CTC*. Pendekatan ini untuk mengenali teks dari gambar secara *end-to-end* tanpa memerlukan segmentasi karakter secara eksplisit. Model ini dibangun dengan menggabungkan beberapa komponen utama, yaitu CNN untuk mengekstraksi fitur spasial dari citra, LSTM untuk menangani dependensi sekuensial dalam data, dan *CTC loss* untuk menyelaraskan output prediksi dengan label tanpa segmentasi manual yang dijelaskan seperti berikut.

2.4.1. Convolutional Neural Network (CNN)

CNN merupakan komponen utama dalam pengembangan sistem OCR yang mempunyai kemampuan untuk mengekstraksi fitur visual dari gambar secara hierarkis. CNN bekerja dengan menggunakan operasi konvolusi untuk mendeteksi pola pada area lokal gambar, seperti tepi dan bentuk, yang penting dalam mengenali karakter pada teks. Operasi konvolusi dinyatakan secara matematis pada pers. (1).

$$O(i, j) = \sum_{m=-k}^k \sum_{n=-k}^k K(m, n) \times I(i + m, j + n), \quad (1)$$

di mana O adalah peta fitur keluaran, I adalah gambar input, dan K adalah filter konvolusi. pers. (1) menggambarkan prinsip dasar konvolusi dua dimensi yang banyak digunakan dalam CNN untuk mengenali fitur spasial, seperti yang dijelaskan oleh Dumoulin dan Visin (2016) dalam panduan aritmetika konvolusi mereka. Dengan menumpuk lapisan konvolusi dan *pooling*, CNN mampu belajar dari berbagai fitur abstrak, mulai dari pola sederhana hingga kompleks. CNN sering digunakan bersama arsitektur seperti CRNN (*Convolutional Recurrent Neural Network*) untuk menangani variasi teks, seperti rotasi atau perubahan gaya huruf. Menurut Yadav *et al.* [24], CNN memiliki keunggulan dalam belajar dari pola visual sederhana hingga kompleks dalam lapisan-lapisan konvolusi.

2.4.2. Long Short-Term Memory (LSTM)

Integrasi LSTM digunakan dalam OCR untuk memproses data teks secara berurutan. LSTM adalah tipe RNN (*Recurrent Neural Network*) yang dapat mengingat atau melupakan informasi melalui mekanisme *gates* (*input*, *forget*, dan *output*), sehingga dapat menangani data sekuensial tanpa masalah *vanishing gradient*. Dalam OCR, LSTM digunakan untuk menganalisis keluaran sekuensial dari CNN dan memprediksi probabilitas karakter pada setiap waktu. Persamaan

matematis yang digunakan untuk menggambarkan pembaruan sel memori pada *LSTM* dirujuk dari Sherstinsky [25], yaitu

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t, \quad (2)$$

di mana f_t adalah *forget gate*, i_t adalah *input gate*, dan \tilde{C}_t adalah nilai kandidat. Persamaan ini menjelaskan bagaimana *LSTM* menggabungkan informasi sebelumnya dengan input saat ini dalam satu kerangka yang memungkinkan pemrosesan informasi jangka panjang secara efisien. Sherstinsky [25] menegaskan bahwa struktur ini menjadikan *LSTM* unggul dalam menangani data sekuensial karena mampu menjaga informasi penting selama rentang waktu yang panjang. Yadav *et al.* [24] menyatakan bahwa mekanisme *gates* pada *LSTM* (*input*, *forget*, dan *output*) sangat efektif dalam mengatasi masalah *vanishing gradient* pada data sekuensial.

2.4.3. Connectionist Temporal Classification (CTC) Loss

CTC loss merupakan metode penting dalam pelatihan model urutan yang tidak memerlukan penyesuaian eksplisit antara input dan label target. Teknik ini banyak digunakan pada sistem pengenalan karakter otomatis (*OCR*) maupun pengenalan ucapan, di mana panjang input dan output tidak selalu sama. *CTC* memungkinkan model mempertimbangkan semua jalur penyesuaian yang mungkin dengan menyisipkan token *blank* dan mengizinkan pengulangan simbol. Persamaan matematis yang digunakan untuk mendefinisikan fungsi *loss CTC* dirujuk dari Liu, Zhou, dan Yu (2018) dinyatakan pada pers. (3).

$$L_{CTC} = -\log p(l | X) = -\log \sum_{\pi \in B^{-1}(l)} p(\pi | X), \quad (3)$$

di mana X adalah urutan input, l adalah label target, dan $B^{-1}(l)$ merupakan himpunan semua jalur penyesuaian yang valid yang akan dikompres menjadi urutan target. Fungsi ini memungkinkan model untuk dilatih secara *end-to-end* tanpa perlu segmentasi karakter manual. Pendekatan ini sangat penting untuk mengenali teks pada sistem *OCR*, terutama jika data tidak tersegmentasi atau memiliki *noise*. Pendekatan *CTC loss* memungkinkan model *OCR* untuk menangani data teks tanpa segmentasi eksplisit, sesuai dengan yang dijelaskan oleh Campiotti dan Lotufo [26].

2.5. Evaluasi

Tahapan evaluasi dilakukan pada model yang telah dilatih untuk pengenalan teks dengan menggunakan metrik utama yaitu *Character Error Rate (CER)*. Metrik ini digunakan untuk mengukur akurasi sistem *OCR* dalam mengenali teks pada level karakter dan kata. Rumus perhitungan *CER* merujuk pada pendekatan yang dijelaskan oleh Campiotti dan Lotufo [26], yang menghitung jumlah kesalahan karakter menggunakan jarak *Levenshtein*, kemudian membaginya dengan panjang teks referensi. Perhitungan *CER* dinyatakan pada pers. (4).

$$CER = \left(\frac{\text{Levenshtein Distance}}{\text{Panjang Teks Referensi}} \right) \times 100. \quad (4)$$

Persamaan ini dapat mengukur jumlah operasi minimum (substitusi, penyisipan, dan penghapusan) yang diperlukan untuk mengubah satu string menjadi string lainnya, sedangkan Panjang Teks Referensi mengacu pada total karakter dalam *ground truth*. Penggunaan *CER* dalam penelitian Campiotti dan Lotufo bertujuan untuk menilai sejauh mana sistem *OCR* dapat mendekati hasil pengenalan karakter yang akurat dalam lingkungan teks yang tidak tersegmentasi dengan jelas. Dengan menggunakan *CER* sebagai metrik evaluasi, model dapat dinilai berdasarkan kemampuannya mengenali teks secara akurat pada level karakter. Nilai

CER yang rendah menunjukkan performa pengenalan yang baik, sedangkan nilai yang tinggi mengindikasikan perlunya peningkatan arsitektur atau data pelatihan agar sistem OCR lebih andal dalam aplikasi dunia nyata.

3. Hasil dan Pembahasan

3.1. Pengumpulan Dataset

Sebelum memasuki tahap pelatihan dan evaluasi model, dilakukan serangkaian proses awal yang sangat krusial dalam membentuk kualitas data masukan bagi sistem OCR. Tahapan awal dimulai dengan pengumpulan dataset berupa gambar KTP milik nasabah dari sebuah bank swasta di Indonesia. Sebanyak 2500 gambar dikumpulkan untuk mewakili variasi kondisi nyata, termasuk perbedaan posisi, pencahayaan, dan kualitas visual lainnya yang secara signifikan dapat mempengaruhi akurasi hasil ekstraksi teks. Setelah itu, gambar-gambar tersebut melewati proses seleksi awal agar dapat memastikan bahwa hanya data yang layak dan representatif yang diterapkan pada tahap pelatihan.

3.2. Pemrosesan Gambar

Pemrosesan gambar (*preprocessing*) yang dilakukan untuk meningkatkan kualitas visual sehingga informasi teks pada gambar dapat dikenali lebih akurat. Gambar dikonversi ke skala abu-abu untuk mengurangi kompleksitas tanpa kehilangan informasi penting, kemudian dilakukan *cropping* manual terhadap area spesifik seperti NIK, Nama, dan Tempat/Tanggal Lahir guna fokus pada bagian penting yang akan diekstraksi.

3.3. Pemisahan Data

Setelah *preprocessing* selesai, dataset kemudian dipisahkan menjadi data pelatihan dan data pengujian dengan proporsi 80:20. Untuk menjaga akurasi dan menghindari bias dalam evaluasi, pembagian data dilakukan menggunakan teknik *k-Fold Cross Validation* dengan nilai $k = 5$. Pendekatan ini memastikan seluruh data mengalami peran sebagai data latih dan uji secara bergantian, memberikan hasil evaluasi model yang lebih konsisten dan reliabel.

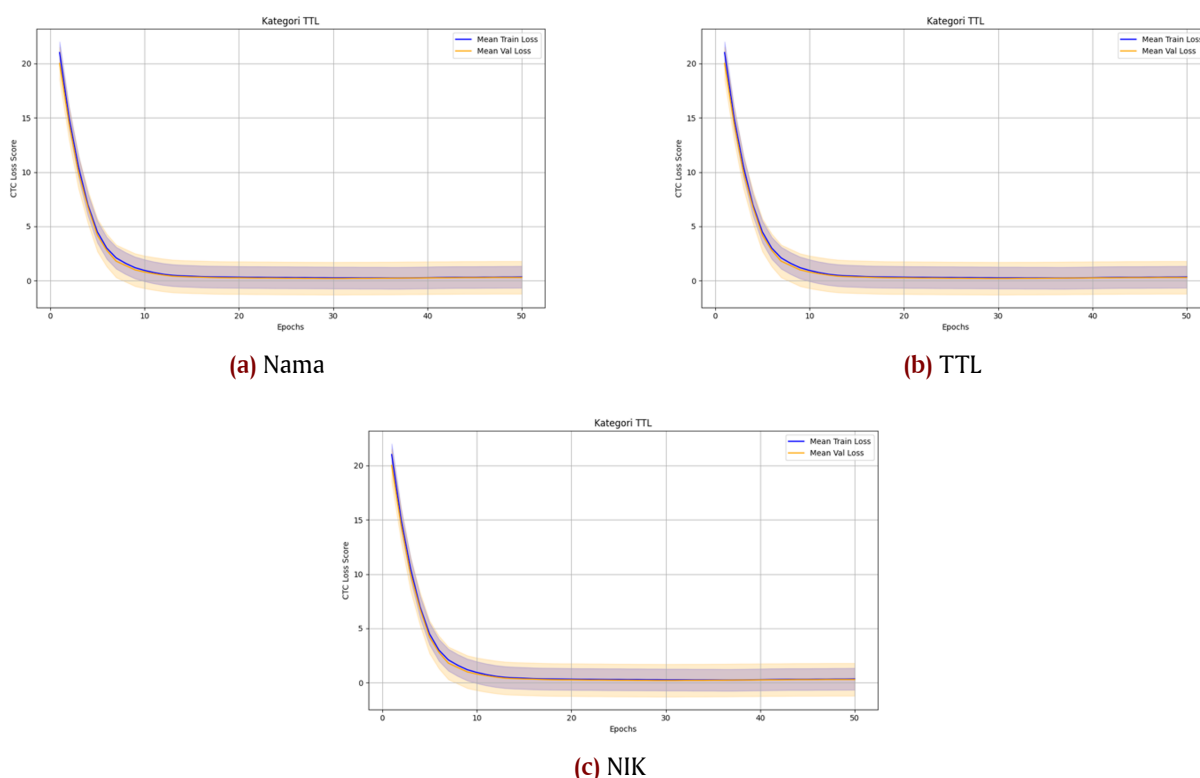
3.4. Pelatihan Data

Setelah dilakukan pemrosesan, model OCR dikembangkan menggunakan kombinasi CNN dan Long Short-Term Memory (LSTM). CNN bekerja dengan menggunakan operasi konvolusi pada pers. (1) untuk mendeteksi pola pada area lokal gambar. CNN digunakan untuk mengekstraksi fitur spasial dari gambar input, sementara LSTM menangani data urutan untuk mengenali pola teks. Pembaruan sel memori pada LSTM menggunakan pers. (2). Model ini diintegrasikan dengan CTC sebagai fungsi *loss* untuk menangani masalah teks dengan panjang variabel. Perhitungan CTC menggunakan pers. (4). Parameter utama seperti jumlah layer CNN, jumlah unit LSTM, *learning rate*, dan jumlah *epoch* dioptimalkan melalui beberapa eksperimen guna mengoptimalkan performa model. Model dibangun dengan sejumlah lapisan yang dirancang untuk mengidentifikasi pola dalam data citra berukuran 200×50 piksel dengan satu kanal warna (*grayscale*). Dengan kombinasi lapisan tersebut, model dapat mengenali dan mengelompokkan data secara efektif sesuai dengan kategori yang diinginkan tanpa memperkecil ukuran data sambil mempertahankan informasi penting.

Fitur yang sudah diekstrak diproses melalui beberapa lapisan konvolusi dan *max pooling* tambahan untuk mendapatkan pola yang lebih kompleks. Data kemudian diubah bentuknya agar dapat diproses oleh lapisan *bidirectional LSTM*, yang mampu memahami konteks urutan data dari kedua arah secara bersamaan. Lapisan *dropout* juga diterapkan untuk mencegah

model terlalu menghafal data latihan. Pada tahap akhir, terdapat perbedaan arsitektur pada tiga model yang dilatih, disesuaikan dengan karakteristik masing-masing kategori. Untuk kategori TTL, model menggunakan lapisan *dense* dengan 11 output yang merepresentasikan klasifikasi ke dalam 11 kelas, yaitu angka 0 hingga 9 serta simbol '-'. Untuk kategori NIK, digunakan lapisan *dense* dengan 10 output yang merepresentasikan angka 0 hingga 9. Sementara itu, pada kategori Nama, model menggunakan lapisan *dense* dengan 28 output yang merepresentasikan seluruh huruf alfabet serta simbol '.' dan ','. Kombinasi lapisan-lapisan tersebut memungkinkan model untuk mengenali dan mengklasifikasikan data secara efektif sesuai dengan karakteristik masing-masing kategori.

Arsitektur yang telah dibuat diimplementasikan ke dalam model dan dilatih dengan dataset yang telah disiapkan. Model akan dilatih ke dalam masing-masing kategori yaitu, NIK, Nama, dan Tempat Tanggal Lahir (TTL), dan diberikan pada **Gambar 2**.



Gambar 2. Hasil pelatihan kategori Nama, TTL, NIK

Gambar 2 menunjukkan hasil dari masing-masing pelatihan pada setiap kategori. **Gambar 2a** menunjukkan pelatihan pada kategori TTL, **Gambar 2b** menunjukkan pelatihan pada kategori Nama, dan **Gambar 2c** menunjukkan pelatihan pada kategori NIK. Garis biru pada **Gambar 2** merepresentasikan *training loss*, yaitu kesalahan model saat belajar dari data pelatihan pada setiap *epoch*. Garis ini menunjukkan seberapa baik model mempelajari pola data pelatihan. Garis oranye merepresentasikan *validation loss*, yaitu kesalahan model saat diuji pada data validasi yang tidak terlihat selama pelatihan, yang menunjukkan kemampuan generalisasi model. Pada kategori TTL, terlihat bahwa kedua garis, baik *training loss* (biru) maupun *validation loss* (oranye), menurun secara konsisten dan mencapai nilai yang rendah serta stabil setelah *epoch* ke-15. Hal ini mengindikasikan bahwa model berhasil belajar pola data TTL dengan baik sekaligus mempertahankan kemampuan generalisasi terhadap data baru. Pola konvergensi yang stabil ini menunjukkan bahwa kompleksitas model seimbang dengan kom-

pleksitas data pada kategori TTL, sehingga tidak terjadi indikasi *overfitting* maupun *underfitting* yang signifikan.

Pada kategori *Nama*, hasil pelatihan menunjukkan penurunan *training loss* yang signifikan, terutama pada sepuluh *epoch* pertama, yang menandakan bahwa model mampu mempelajari pola dalam data pelatihan dengan cukup baik. Namun demikian, *validation loss* justru mulai menunjukkan tren kenaikan setelah melewati *epoch* ke-10. Pola ini mengindikasikan terjadinya *overfitting*, yaitu kondisi ketika model terlalu menyesuaikan diri terhadap data pelatihan dan kehilangan kemampuan *generalisasi* terhadap data yang belum pernah dilihat sebelumnya. Akibatnya, meskipun akurasi terhadap data pelatihan meningkat, performa terhadap data validasi mulai menurun. Fenomena *overfitting* ini kemungkinan disebabkan oleh kompleksitas karakter pada kategori *Nama* yang lebih tinggi, termasuk variasi panjang teks, kombinasi huruf yang lebih beragam, serta distribusi karakter yang tidak merata. Berbeda dengan asumsi umum pada OCR tulisan tangan, kesulitan pada kategori ini bukan disebabkan oleh variasi tulisan, melainkan oleh kompleksitas sekuens karakter.

Sementara itu, untuk kategori *NIK*, *training loss* juga menunjukkan penurunan secara konsisten, mengindikasikan bahwa model terus belajar dari data pelatihan. Berbeda dengan kategori *Nama*, *validation loss* pada kategori ini mengalami fluktuasi yang lebih moderat dan baru menunjukkan sedikit kenaikan setelah sekitar *epoch* ke-20. Hal ini mengisyaratkan adanya potensi *overfitting* ringan, namun dalam tingkat yang lebih terkendali. Model pada kategori *NIK* tampaknya masih mampu mempertahankan performa terhadap data validasi dalam rentang waktu pelatihan yang lebih panjang dibandingkan kategori *Nama*. Hal ini menunjukkan bahwa data numerik seperti *NIK* memiliki pola yang lebih terstruktur dan konsisten, sehingga lebih mudah dipelajari oleh model dibandingkan data alfabetik yang memiliki variasi lebih tinggi.

Secara keseluruhan, perbedaan pola antara *training loss* dan *validation loss* pada ketiga kategori ini mencerminkan proses pelatihan model dan kemampuan *generalisasi* model terhadap variasi data yang berbeda.

3.5. Evaluasi

Setelah melakukan pelatihan pada model, selanjutnya peneliti akan melakukan evaluasi model dengan *Character Error Rate (CER)*. CER merupakan ukuran yang digunakan untuk mengevaluasi kesalahan prediksi model pada level karakter sesuai dengan pers. (4). Hasil dari evaluasi CER setiap kategori ditunjukkan pada Tabel 1. Nilai CER yang diperoleh merepresentasikan performa rata-rata dari beberapa iterasi pengujian dengan penggunaan teknik *5-fold cross validation*. Secara umum, variasi antar *fold* relatif kecil yang menunjukkan bahwa model memiliki stabilitas performa yang baik terhadap perbedaan subset data.

Tabel 1. Hasil evaluasi CER setiap kategori

Kategori	CER
NIK	1.29%
Nama	4.33%
TTL	0.84%

Hasil evaluasi model menggunakan metrik CER menunjukkan bahwa kategori *NIK* memiliki CER sebesar 1.29%, kategori *Nama* 4.33%, dan kategori *TTL* 0.84%. Hasil tersebut menunjukkan bahwa model menunjukkan performa terbaik pada kategori *TTL*, dengan tingkat kesalahan karakter yang paling rendah. Faktor penyebabnya berasal dari karakteristik gambar

pada kategori TTL yang cenderung lebih mudah dan mengandung lebih sedikit variasi dibandingkan kategori *Nama*, yang cenderung lebih kompleks. Perbedaan ini menunjukkan bahwa karakteristik data, khususnya panjang sekuens dan variasi kombinasi karakter, memiliki pengaruh signifikan terhadap performa model OCR. Untuk kategori *Nama*, nilai CER yang lebih besar mengindikasikan bahwa model mungkin kesulitan dalam mengidentifikasi karakter secara akurat, yang mungkin disebabkan oleh kompleksitas kombinasi karakter dan panjang teks. Peneliti melanjutkan dengan analisis lebih lanjut dan melakukan perbaikan pada model untuk mengurangi CER pada kategori yang lebih sulit, seperti *Nama*.

Hasil evaluasi menunjukkan bahwa kategori Tempat dan Tanggal Lahir (TTL) memiliki tingkat kesalahan terendah (CER 0,84%), diikuti oleh kategori NIK (1,29%), sedangkan kategori *Nama* mencatat tingkat kesalahan tertinggi (4,33%). Secara konseptual, performa model ini juga menunjukkan peningkatan dibanding pendekatan berbasis CNN saja, karena integrasi LSTM memungkinkan pemodelan dependensi sekuensial yang lebih baik, khususnya pada teks dengan panjang variabel. Dengan strategi lanjutan seperti *data augmentation* dan *model ensemble*, akurasi di kategori *Nama* juga diharapkan bisa ditingkatkan untuk mendukung implementasi menyeluruh. Misalnya, dalam proses verifikasi KTP secara otomatis di instansi pemerintah atau platform digital (e.g. Dukcapil, bank, atau aplikasi pinjaman online), informasi seperti tanggal lahir dan NIK merupakan kunci validasi utama. Dengan nilai CER yang rendah pada kedua kategori tersebut, sistem ini dinilai sudah memadai untuk diimplementasikan dalam proses otomatisasi input data kependudukan yang sebelumnya dilakukan secara manual. Implementasi pada skala industri masih memerlukan pengujian lebih lanjut, termasuk validasi pada *dataset* yang lebih besar, variasi kondisi citra yang lebih luas, serta evaluasi *robustness* terhadap *noise* dan distorsi.

Tingginya CER pada kategori *Nama* menunjukkan bahwa model masih perlu diperbaiki sebelum dapat diandalkan penuh dalam konteks yang menuntut akurasi tinggi pada elemen identitas seperti nama lengkap, seperti dalam pencetakan dokumen resmi atau sistem legal. Oleh karena itu, meskipun model menunjukkan potensi kuat, penggunaannya di dunia nyata perlu disesuaikan berdasarkan sensitivitas data yang diproses. Dengan strategi lanjutan seperti *data augmentation* dan *model ensemble*, akurasi di kategori *Nama* juga diharapkan bisa ditingkatkan untuk mendukung implementasi menyeluruh. Temuan ini selaras dengan tujuan awal penelitian, yaitu mengembangkan sistem OCR berbasis CNN, LSTM, dan CTC Loss yang mampu menangani data teks pada dokumen e-KTP dengan fleksibilitas tinggi terhadap variasi struktur. Secara khusus, penelitian ini menunjukkan bahwa evaluasi berbasis kategori teks memberikan wawasan yang lebih mendalam dibanding evaluasi agregat, sehingga dapat menjadi dasar dalam pengembangan sistem OCR yang lebih adaptif terhadap karakteristik data. Hasil evaluasi pada kategori NIK dan *Nama* menunjukkan bahwa pendekatan arsitektural yang digunakan telah berhasil mengoptimalkan proses ekstraksi data dengan tingkat kesalahan yang relatif rendah pada data numerik, serta mengidentifikasi tantangan pada data teks kompleks seperti nama lengkap. Dengan demikian, penelitian ini memberikan kontribusi nyata terhadap upaya otomatisasi administrasi kependudukan melalui sistem OCR yang dapat terus dikembangkan untuk mencapai akurasi tinggi di seluruh komponen data identitas.

4. Kesimpulan

Hasil pelatihan dan evaluasi model OCR yang menggunakan kombinasi CNN dan LSTM menunjukkan bahwa performa model bervariasi pada setiap kategori teks. Model menunjukkan bahwa performa terbaik pada kategori Tempat Tanggal Lahir (TTL), dengan nilai CER terendah sebesar 0,84%. Hal ini mengindikasikan bahwa model lebih efektif dalam mengenali

teks dengan pola yang relatif sederhana dan terstruktur. Sementara itu, pada kategori *Nama* model menghadapi tantangan yang lebih besar, dengan indikasi *overfitting* yang terlihat dari meningkatnya *validation loss* dan tingginya nilai *CER* sebesar 4,33%. Tantangan ini terutama disebabkan oleh kompleksitas data, seperti variasi panjang teks dan kombinasi karakter yang lebih beragam, bukan karena gaya tulisan. Di sisi lain, kategori *NIK* menunjukkan performa cukup baik, dengan *CER* sebesar 1,29% yang menandakan bahwa model mampu mengenali pola numerik yang relatif konsisten secara efektif.

Secara keseluruhan, hasil penelitian ini menunjukkan bahwa pendekatan *CNN-LSTM* berbasis *CTC* efektif untuk ekstraksi informasi pada e-KTP, namun performanya sangat dipengaruhi oleh karakteristik teks yang diproses. Oleh karena itu, beberapa strategi disarankan untuk mengoptimalkan performa model. Salah satunya adalah penerapan teknik *data augmentation*, seperti rotasi, pemotongan, atau penambahan *noise* pada gambar, untuk meningkatkan keragaman data pelatihan dan mengurangi *overfitting*. Selain itu, teknik *regularization* seperti *dropout* atau *L2 regularization* juga dapat membantu model belajar secara lebih general. Optimalisasi *hyperparameter*, termasuk jumlah lapisan *CNN*, unit *LSTM*, dan *learning rate*, juga perlu dieksplorasi lebih lanjut guna memperoleh konfigurasi terbaik. Terakhir, penerapan pendekatan *ensemble model* dapat menjadi alternatif strategis untuk meningkatkan akurasi dan stabilitas prediksi, terutama pada kategori yang memiliki variasi data tinggi seperti *Nama*.

Penelitian ini masih memiliki keterbatasan karena jumlah dataset yang terbatas serta belum adanya pengujian pada kondisi citra yang lebih beragam dan lingkungan nyata. Dengan demikian, penelitian ini memberikan kontribusi dalam menunjukkan pentingnya analisis performa *OCR* berdasarkan kategori teks, sekaligus menjadi dasar untuk pengembangan metode yang lebih adaptif terhadap kompleksitas data pada dokumen identitas.

Kontribusi Penulis. Fauzan Ihza Fajar: Konseptualisasi, metodologi, perangkat lunak, analisis formal, penulisan–persiapan draf asli. Selly Anastassia Amellia Kharis: Penulisan–tinjauan dan penyuntingan, supervisi, validasi. Asmara Iriani Tarigan: Penulisan–tinjauan dan penyuntingan, supervisi, validasi.

Ucapan Terima Kasih. Penulis mengucapkan terima kasih kepada semua pihak yang telah berkontribusi dalam penelitian ini khususnya untuk Lembaga Penelitian dan Pengabdian kepada Masyarakat (LPPM) Universitas Terbuka. Kami sangat menghargai editor dan reviewer atas masukan serta dukungannya dalam menyempurnakan karya ini.

Pembiayaan. Penelitian ini didanai oleh Lembaga Penelitian dan Pengabdian kepada Masyarakat (LPPM) Universitas Terbuka.

Konflik Kepentingan. Para penulis menyatakan tidak ada konflik kepentingan yang terkait dengan artikel ini.

Ketersediaan Data. Tidak tersedia.

Referensi

- [1] O. Rahmdani, “Evaluasi Kinerja Tesseract-OCR dalam Pengenalan Teks Tulisan Tangan Menggunakan Dataset Kustom,” *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 3, Jul. 2025, doi: [10.23960/jitet.v13i3.7162](https://doi.org/10.23960/jitet.v13i3.7162).
- [2] M. Riyandi Fauzi, N. D. Agus, and A. Z. Ajulian, “Mengubah Tulisan Tangan Menjadi Text Digital OCR (Optical Character Recognition) dengan Menggunakan Metode Segmentasi dan Korelasi,” *Transient: Jurnal Ilmiah Teknik Elektro*, vol. 2, no. 4, pp. 1013–1017, 2014, doi: [10.14710/transient.v2i4.1013-1017](https://doi.org/10.14710/transient.v2i4.1013-1017).

- [3] Y. Darmi, M. F. Sepriansyah, and Y. Darnita, "Penerapan Metode Optical Character Recognition (OCR) untuk Mengidentifikasi Teks pada Identitas Dokumen Surat Izin Mengemudi (SIM)," *Jurnal Mahasiswa Teknik Informatika*, vol. 9, no. 4, 2025, doi: [10.36040/jati.v9i4.13987](https://doi.org/10.36040/jati.v9i4.13987).
- [4] M. Tampang, I. Sartika, and F. Ruhana, "Kualitas Pelayanan Publik dalam Pembuatan Kartu Tanda Penduduk Elektronik (E-KTP) di Suku Dinas Kependudukan dan Catatan Sipil Kota Jakarta Selatan," *Jurnal Kajian Pemerintah (JKP)*, vol. 10, pp. 73–85, 2024, doi: [10.25299/jkp.2024.vol10\(1\).16958](https://doi.org/10.25299/jkp.2024.vol10(1).16958).
- [5] B. Shi, X. Bai, and C. Yao, "An End-to-End Trainable Neural Network for Image-based Sequence Recognition and Its Application to Scene Text Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 11, Jul. 2015, doi: [10.1109/TPAMI.2016.2646371](https://doi.org/10.1109/TPAMI.2016.2646371).
- [6] B. Baso, Risald, and N. Huda, "Hybrid Algoritma Convolutional Neural Network dengan Support Vector Machine untuk Klasifikasi Jenis Tenun Timor Hybrid," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, vol. 12, no. 6, pp. 1233–1242, 2025, doi: [10.25126/jtiik.2025126](https://doi.org/10.25126/jtiik.2025126).
- [7] N. Chakraborty, S. Kundu, S. Paul, A. F. Mollah, S. Basu, and R. Sarkar, "Language identification from multi-lingual scene text images: a CNN based classifier ensemble approach," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 7, pp. 7997–8008, Jul. 2021, doi: [10.1007/s12652-020-02528-4](https://doi.org/10.1007/s12652-020-02528-4).
- [8] W. Lu, J. Li, Y. Li, A. Sun, and J. Wang, "A CNN-LSTM-based model to forecast stock prices," *Complexity*, vol. 2020, 2020, doi: [10.1155/2020/6622927](https://doi.org/10.1155/2020/6622927).
- [9] M. Hafizh Fattah, M. Alfian Rosid, S. Aji, P. Studi Informatika, and F. Sains dan Teknologi, "Hybrid CNN-LSTM for Indonesian Cyberbullying Detection on Social Media X," *JITE (Journal of Informatics and Telecommunication Engineering)*, vol. 9, no. 2, pp. 548–563, 2026, doi: [10.31289/jite.v9i2.16938](https://doi.org/10.31289/jite.v9i2.16938).
- [10] B. Rizki Hanafi, P. Stiaji, and W. Agus Triyanto, "Implementasi Image Processing dalam Pemindaian Data KTP menggunakan Optical Character Recognition (OCR)," *Sistemasi: Jurnal Sistem Informasi*, vol. 15, 2026, doi: [10.32520/stmsi.v15i1.5856](https://doi.org/10.32520/stmsi.v15i1.5856).
- [11] N. Syafrie Rahardian, E. Maryanto, and D. A. Nawangnugraeni, "An Integrated Pipeline with Hierarchical Segmentation and CNN for Automated KTP-el Data Extraction on the e-Magang Platform," *Jurnal Teknik Informatika (Jutif)*, vol. 6, no. 5, pp. 3093–3110, Oct. 2025, doi: [10.52436/1.jutif.2025.6.5.5279](https://doi.org/10.52436/1.jutif.2025.6.5.5279).
- [12] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, doi: [10.1109/CVPR.2016.90](https://doi.org/10.1109/CVPR.2016.90).
- [13] M. Geetha, R. C. Suganthe, S. K. Nivetha, S. Hariprasath, S. Gowtham, and C. S. Deepak, "A Hybrid Deep Learning Based Character Identification Model Using CNN, LSTM, And CTC To Recognize Handwritten English Characters And Numerals," in *2022 International Conference on Computer Communication and Informatics (ICCCI)*, 2022, doi: [10.1109/ICCCI54379.2022.9740746](https://doi.org/10.1109/ICCCI54379.2022.9740746).
- [14] R. Ramadhan Harahap, B. Fachri, and R. Prayudi, "Pemanfaatan Teknologi OCR (Optical Character Recognition) dalam Pembuatan Aplikasi Kalkulator Tulisan Tangan Sederhana," *Journal of Science and Social Research*, no. 2, pp. 272–278, 2022, doi: [10.54314/jssr.v5i2.916](https://doi.org/10.54314/jssr.v5i2.916).
- [15] K. Banu, D. Andreas, W. Anggoro, and A. Setiawan, "OCR: Masa Depan Pengenalan Karakter Optik dan Dampaknya pada Kehidupan Modern," *Jurnal Teknologi Informasi*, vol. 9, 2023, doi: [10.52643/jti.v9i2.3798](https://doi.org/10.52643/jti.v9i2.3798).
- [16] Lady Angel, D. Kaura Amelia, N. Nur Syavina, S. Syahirah, M. Fasha Akbar Aulady, and A. Puteri Amelia, "Pemanfaatan Artificial Intelligence (AI) Untuk Meningkatkan Efisiensi dan Akurasi Data pada Sistem Informasi di Lembaga Kearsipan," 2025, doi: [10.56799/peshum.v5i1.13751](https://doi.org/10.56799/peshum.v5i1.13751).
- [17] A. Tri Putra Darti Akhsa, M. Ikhwan Burhan, and A. Munandar, "Integrasi OCR dan TF-IDF untuk Metadata Otomatis pada Pencarian Dokumen Digital," *Jurnal Fasilkom*, vol. 15, pp. 304–311, 2025, doi: [10.37859/jf.v15i2.9918](https://doi.org/10.37859/jf.v15i2.9918).
- [18] S. Kumari, A. Akole, P. Angnani, Y. Bhamare, and Z. Naikwadi, "Enhanced braille display use of OCR and solenoid to improve text to braille conversion," in *2020 International Conference for Emerging Technology (INCET)*, Jun. 2020, doi: [10.1109/INCET49848.2020.9153996](https://doi.org/10.1109/INCET49848.2020.9153996).
- [19] F. Gesang Panuntun and R. Hajar Puji Sejati, "Sistem Otomatisasi Deteksi dan Ekstraksi Data KTP Berbasis Convolutional Neural Network dan Optical Character Recognition," *JSAI: Journal Scientific and Applied Informatics*, vol. 7, no. 3, 2024, doi: [10.36085/j sai.v7i3.7269](https://doi.org/10.36085/j sai.v7i3.7269).
- [20] I. Wijaya, C. Lubis, and K. Kunci, "Pengimplementasian OCR menggunakan CNN untuk Ekstraksi Teks pada Gambar," *Jurnal Ilmu Komputer dan Sistem Informasi*, pp. 1–6, doi: [10.24912/jiksi.v10i1.17836](https://doi.org/10.24912/jiksi.v10i1.17836).
- [21] L. Krismona, A. Ashari, A. Setiawan, and R. Rosnelly, "Deteksi Ketidakkonsistenan Font Sebagai Indikator Pemalsuan dan Penyuntingan Dokumen Digital Menggunakan Convolutional Neural Network (CNN)," *Jurnal Sains Komputer*, vol. 8, no. 2, p. 214, 2025, doi: [10.53513/jsk.v8i2.11741](https://doi.org/10.53513/jsk.v8i2.11741).
- [22] I. Dwijayanti et al., "Ekstraksi Aspek Aksesibilitas untuk Peningkatan Pengalaman Pengguna Menggunakan NER dengan CNN dan LSTM," *Jurnal Fasilkom*, vol. 14, no. 3, pp. 556–563, 2024, doi: [10.37859/jf.v14i3.9918](https://doi.org/10.37859/jf.v14i3.9918).

- 10.37859/jf.v14i3.8032.
- [23] M. T. Panjalu and L. M. Wisudawati, "Pembacaan Gerak Bibir Menggunakan CNN, Bi-LSTM Dan CTC Loss Function Pada Dataset Bahasa Inggris," *Jurnal Ilmiah Komputasi*, vol. 24, no. 1, pp. 51–60, Mar. 2025, doi: [10.32409/jikstik.24.1.3658](https://doi.org/10.32409/jikstik.24.1.3658).
- [24] A. Yadav, S. Singh, M. Siddique, N. Mehta, and A. Kotangale, "OCR using CRNN: A Deep Learning Approach for Text Recognition," in *2023 4th International Conference for Emerging Technology (INCET)*, 2023, doi: [10.1109/INCET57972.2023.10170436](https://doi.org/10.1109/INCET57972.2023.10170436).
- [25] A. Sherstinsky, "Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network," *Physica D: Nonlinear Phenomena*, vol. 404, p. 132306, Mar. 2020, doi: [10.1016/j.physd.2019.132306](https://doi.org/10.1016/j.physd.2019.132306).
- [26] I. Campiotti and R. Lotufo, "Optical character recognition with transformers and CTC," in *DocEng 2022 - Proceedings of the 2022 ACM Symposium on Document Engineering*, Association for Computing Machinery, Inc., Sep. 2022, doi: [10.1145/3558100.3563845](https://doi.org/10.1145/3558100.3563845).