

BAB I

PENDAHULUAN

1.1 Latar Belakang

Transformasi digital dalam dua dekade terakhir telah mendorong percepatan digitalisasi dokumen pada berbagai sektor, mulai dari administrasi pemerintahan, arsip pendidikan, hingga layanan kesehatan. Menurut data (BPS 2025), pasar global untuk teknologi *Optical Character Recognition* (OCR) diproyeksikan mencapai USD 33,9 miliar pada tahun 2027, dengan pertumbuhan tahunan rata-rata 15,2%. (UNESCO 2024) juga melaporkan bahwa lebih dari 60% arsip nasional di negara berkembang masih berbentuk fisik. Kondisi ini menunjukkan bahwa kebutuhan digitalisasi dokumen berbasis OCR semakin mendesak, terutama untuk meningkatkan aksesibilitas dan pengelolaan informasi.

Namun, tantangan besar muncul ketika dokumen fisik tersebut berisi tulisan tangan, terutama dalam bentuk sambung atau *cursive handwriting*. Bentuk tulisan sambung sangat lazim digunakan, baik pada bahasa Latin maupun non-Latin, tetapi justru lebih sulit dikenali secara otomatis dibanding teks cetak. *International Conference on Document Analysis and Recognition* (Mondal and Jawahar 2023) melaporkan bahwa tingkat kesalahan pengenalan karakter pada tulisan sambung bisa mencapai 20% lebih tinggi dibanding teks cetak. Hal ini dipengaruhi oleh variasi gaya individu, keterhubungan antarhuruf, dan kualitas pemindaian dokumen. Akibatnya, banyak arsip penting yang belum bisa diubah ke bentuk digital secara akurat, sehingga menimbulkan kesenjangan antara kebutuhan digitalisasi dan performa teknologi yang tersedia.

Perkembangan *deep learning* memberikan solusi melalui kombinasi *Convolutional Neural Network* (CNN) dan *Recurrent Neural Network* (RNN) dengan fungsi *Connectionist Temporal Classification*

(CTC) loss. CNN berperan dalam mengekstraksi fitur spasial dari citra tulisan, sementara RNN efektif dalam mempelajari urutan temporal dari karakter atau kata. Dengan tambahan CTC loss, model dapat melakukan pelabelan sekuensial tanpa segmentasi eksplisit, memungkinkan proses pelatihan secara *end-to-end*. Sejumlah penelitian mendukung efektivitas pendekatan ini. (Ma 2024) menunjukkan bahwa kombinasi CNN–RNN mampu menurunkan *Character Error Rate* (CER) dan *Word Error Rate* (WER) dibanding CNN murni pada dataset IAM dan RIMES. (Kizilirmak and Yanikoglu 2023) membuktikan bahwa CNN–BiLSTM dengan CTC loss dapat mengurangi CER secara signifikan pada tulisan tangan sambung bahasa Inggris melalui teknik augmentasi data. Sementara itu, (Retsinas et al. 2022) dan (Ferro et al. 2024) menegaskan bahwa modifikasi arsitektur CNN–RNN mampu meningkatkan performa pengenalan teks sekuensial dibandingkan model konvensional.

Meski demikian, hasil penelitian tersebut masih menyisakan sejumlah keterbatasan. Sebagian besar studi terfokus pada dataset asing seperti Arab, Bangla, dan Inggris (Imane et al. 2025); (Elsayed et al. 2024); (Akter et al. 2022); (Waly et al., 2024). Selain itu, banyak penelitian hanya berfokus pada pengenalan kata atau baris tertentu, belum mengoptimalkan pengenalan teks pada dokumen panjang secara menyeluruh. Beberapa studi bahkan masih berbentuk tinjauan literatur, misalnya oleh (Alhamad et al. 2024) dan (AlKendi et al. 2024), sehingga kontribusinya terbatas pada pemetaan tren penelitian, bukan implementasi empiris. Hal ini menunjukkan adanya *research gap*, yaitu perlunya penelitian yang secara komprehensif menerapkan arsitektur CNN–RNN dengan berbagai variasi RNN dan CTC loss untuk pengenalan tulisan tangan sambung pada dokumen digital.

Berdasarkan celah penelitian tersebut, penelitian ini berfokus pada pengembangan sistem *Handwritten Text Recognition* (HTR) berbasis kombinasi CNN–RNN dengan CTC loss, dengan menguji dua variasi

RNN, yaitu BiLSTM dan GRU. Setiap variasi dibandingkan untuk mengetahui arsitektur mana yang memiliki performa terbaik dalam mengenali tulisan tangan sambung. Evaluasi dilakukan menggunakan beberapa metrik kuantitatif, yakni *Character Error Rate (CER)*, *Word Error Rate (WER)*, *Sentence Error Rate (SER)*, *Character Recognition Rate (CRR)*, dan *Word Recognition Rate (WRR)*. Pendekatan kuantitatif dipilih karena memungkinkan pengukuran empiris berbasis data numerik yang objektif, sehingga hasilnya dapat menjadi dasar valid untuk menentukan performa model terbaik.

Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi empiris terhadap pengembangan sistem HTR berbasis *deep learning*, serta mendukung percepatan digitalisasi dokumen secara lebih akurat, inklusif, dan dapat diandalkan.

1.2 Identifikasi Masalah

Berdasarkan uraian pada latar belakang, dapat diidentifikasi sejumlah permasalahan yang melatarbelakangi penelitian ini, yaitu sebagai berikut:

1. Keterbatasan Implementasi Arsitektur CNN–RNN dengan CTC Loss pada Tulisan Tangan Sambung

Meskipun penelitian mengenai pengenalan tulisan tangan sudah cukup banyak, penerapan metode end-to-end yang menggabungkan CNN, RNN, dan CTC Loss khususnya pada tulisan tangan sambung masih jarang dilakukan, sehingga efektivitas arsitektur tersebut belum banyak dievaluasi secara komprehensif.

2. Keterbatasan Dataset dan Representasi Bahasa

Sebagian besar penelitian masih berfokus pada tulisan tangan non-sambung, baik berbahasa Inggris maupun bahasa asing lainnya. Akibatnya, kajian serta ketersediaan dataset khusus tulisan tangan sambung masih terbatas untuk mendukung peningkatan akurasi model.

3. Keterbatasan Pendekatan dan Evaluasi Model

Pendekatan kombinasi CNN–RNN dengan CTC loss telah menunjukkan hasil baik, namun umumnya hanya menggunakan satu arsitektur RNN tanpa membandingkan variasi lain seperti BiLSTM dan GRU. Selain itu, evaluasi model masih terbatas pada metrik CER dan WER, sehingga belum mencerminkan performa model secara menyeluruh.

1.3 Rumusan Masalah

Berdasarkan identifikasi masalah yang telah dijelaskan sebelumnya, maka permasalahan yang akan dikaji dalam penelitian ini dapat dirumuskan sebagai berikut:

1. Bagaimana merancang dan membangun sistem pengenalan tulisan tangan sambung berbasis kombinasi *Convolutional Neural Network* (CNN), *Recurrent Neural Network* (RNN), dan *Connectionist Temporal Classification* (CTC) loss agar mampu melakukan pengenalan teks secara otomatis?
2. Bagaimana perbandingan performa dua varian arsitektur RNN, yaitu *Bidirectional LSTM* (BiLSTM), dan *Gated Recurrent Unit* (GRU), dalam mendeteksi serta mengenali pola tulisan tangan sambung pada dokumen digital?
3. Seberapa baik tingkat akurasi masing-masing model CNN–RNN dengan CTC loss berdasarkan hasil evaluasi menggunakan metrik *Character Error Rate* (CER), *Word Error Rate* (WER), *Sentence Error Rate* (SER), *Character Recognition Rate* (CRR), dan *Word Recognition Rate* (WRR)?

1.4 Tujuan Penelitian

Penelitian ini memiliki tujuan sebagai berikut:

1. Mengimplementasikan kombinasi arsitektur *Convolutional Neural Network* (CNN) dan *Recurrent Neural Network* (RNN) dengan fungsi *Connectionist Temporal Classification* (CTC) Loss dalam sistem pengenalan tulisan tangan sambung (*Handwritten Text Recognition*).
2. Menganalisis dan membandingkan kinerja dua varian model RNN, yaitu BiLSTM dan GRU, guna memperoleh model yang paling optimal dalam mengenali teks tulisan tangan sambung.
3. Mengevaluasi performa model menggunakan beberapa metrik pengukuran, yaitu *Character Error Rate* (CER), *Word Error Rate* (WER), *Sentence Error Rate* (SER), *Character Recognition Rate* (CRR), dan *Word Recognition Rate* (WRR), dengan memanfaatkan IAM Handwriting Dataset sebagai sumber data tulisan tangan sambung berbahasa Inggris.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan beberapa manfaat sebagai berikut:

1. Memberikan kontribusi dalam pengembangan ilmu pengetahuan di bidang *Deep Learning*, khususnya pada penerapan kombinasi CNN–RNN dengan CTC Loss untuk sistem pengenalan tulisan tangan sambung (*Handwritten Text Recognition*).
2. Menjadi acuan bagi penelitian selanjutnya dalam membandingkan performa berbagai variasi RNN, yaitu LSTM, BiLSTM, dan GRU, dalam mengenali teks tulisan tangan.
3. Memberikan dasar bagi pengembangan sistem digitalisasi dokumen berbasis OCR yang mampu mengenali tulisan tangan sambung secara lebih akurat.

4. Mendorong pemanfaatan teknologi kecerdasan buatan dalam proses otomatisasi pengenalan tulisan tangan untuk mendukung digitalisasi arsip dan dokumen penting.
5. Mendukung upaya transformasi digital di berbagai sektor dengan menghadirkan model pengenalan tulisan tangan yang lebih efisien, akurat, dan dapat diandalkan.

1.6 Batasan Masalah

Agar penelitian ini tetap terarah dan sesuai dengan tujuan yang telah ditetapkan, maka ruang lingkup penelitian dibatasi pada hal-hal berikut:

1. Data yang digunakan berasal dari IAM Handwriting Dataset, yang berisi citra tulisan tangan sambung (*cursive handwriting*) berhuruf Latin. Data ini bersifat offline, sehingga penelitian tidak mencakup proses pengenalan tulisan tangan secara real-time.
2. Model yang dikembangkan menggunakan kombinasi *Convolutional Neural Network (CNN)* dan *Recurrent Neural Network (RNN)* dengan penerapan *Connectionist Temporal Classification (CTC) loss* sebagai metode pelabelan sekuensial untuk pengenalan teks.
3. Variasi arsitektur RNN yang dianalisis meliputi dua jenis, yaitu *Bidirectional LSTM (BiLSTM)* dan *Gated Recurrent Unit (GRU)*, dengan tujuan membandingkan performa masing-masing dalam mengenali pola tulisan tangan sambung.
4. Evaluasi performa model dilakukan menggunakan lima metrik kuantitatif, yakni *Character Error Rate (CER)*, *Word Error Rate (WER)*, *Sentence Error Rate (SER)*, *Character Recognition Rate (CRR)*, dan *Word Recognition Rate (WRR)*, sebagai indikator tingkat akurasi hasil pengenalan teks.

1.7 Mata Kuliah yang mendasari penelitian

Penelitian ini didukung oleh beberapa mata kuliah yang berkaitan dengan topik dan metode yang digunakan, yaitu:

1. Pengolahan Citra Digital

Mata kuliah ini memberikan dasar tentang teknik pengolahan gambar seperti konversi, normalisasi, dan peningkatan kualitas citra yang berguna dalam tahap preprocessing data tulisan tangan.

2. Algoritma dan Struktur Data

Pembelajaran ini membantu dalam memahami logika pemrograman, efisiensi proses, serta pengelolaan data yang diterapkan pada perancangan dan pelatihan model.

3. Machine Learning

Mata kuliah ini menjadi dasar dalam memahami konsep pembelajaran mesin, pelatihan model, serta evaluasi akurasi yang digunakan dalam penelitian ini.

4. Deep Learning

Ilmu dari mata kuliah ini mendasari penerapan arsitektur jaringan saraf tiruan seperti CNN, RNN, BiLSTM dan GRU yang digunakan dalam pengembangan model pengenalan tulisan tangan.

5. Kecerdasan Buatan (Artificial Intelligence)

Mata kuliah ini memberikan pemahaman umum tentang sistem cerdas yang dapat meniru kemampuan manusia dalam mengenali pola dan membuat keputusan otomatis.