



# Penerapan Membaca Tulisan di dalam Gambar Menggunakan Metode OCR Berbasis Website pada e-KTP

Muhammad Rizal Toha<sup>1</sup>, Agung Triayudi<sup>2\*</sup> 

<sup>1,2</sup> Sistem Informasi, Universitas Nasional DKI Jakarta, Indonesia

## ARTICLE INFO

### Article history:

Received December 16, 2021

Revised December 20, 2021

Accepted March 14, 2022

Available online April 25, 2022

### Kata Kunci:

OCR, Website, e-KTP

### Keywords:

OCR, Web Based, e-KTP



This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

Copyright © 2022 by Author. Published by Universitas Pendidikan Ganesha.

## ABSTRAK

Penanganan yang membutuhkan data pribadi masih banyak dilakukan secara manual. Tidak jarang masyarakat memfotokopi e-KTP untuk memenuhi berbagai persyaratan administrasi. Dilihat dari tujuan pembuatan e-KTP, penggunaan e-KTP saat ini tentu kurang tepat. *Optical Character Recognition* (OCR) adalah proses yang memungkinkan sistem tanpa campur tangan manusia mengidentifikasi skrip atau abjad yang tertulis dalam komunikasi verbal pengguna. Identifikasi karakter optik telah berkembang pada individu dari aplikasi pengetahuan yang berkembang pesat di bidang deteksi pola dan kecerdasan buatan. Maka dari itu, penelitian ini bertujuan untuk merancang perangkat lunak pembacaan e-KTP dengan metode *optical character recognition* (OCR) berbasis web. Jenis penelitian merupakan jenis kuantitatif dengan pendekatan yang digunakan pengambilan data langsung ke relawan penelitian. Sistem yang akan dibuat untuk membaca tulisan dalam gambar menggunakan metode *Optical Character Recognition*. data yang dibahas dalam penelitian ini yaitu jumlah kartu sebanyak 20 e-KTP, jumlah attribute sebanyak 14 attribute, dan jumlah data sebanyak 280 data. Teknik yang digunakan dalam menganalisis data yaitu analisis deskriptif kualitatif dan kuantitatif. Dari hasil penelitian yang telah dilakukan dalam pendeteksian atribut pada e-KTP menghasilkan akurasi sebesar 100 persen, hasil pengujian 1 dari beberapa atribut yang ditemukan tidak nihil sebesar 98,09 persen dan hasil pengujian 2 yang mendeteksi kategori e-KTP yang masih bagus dan yang kurang bagus sebesar 67,61 persen.

## ABSTRACT

Handling that requires personal data is still mostly done manually. It is not uncommon for people to photocopy their e-KTP to fulfill various administrative requirements. Judging from the purpose of making e-KTP, the current use of e-KTP is certainly not appropriate. *Optical Character Recognition* (OCR) is a process that allows the system, without human intervention, to identify scripts or alphabets written in the user's verbal communication. *Optical character identification* has evolved in individuals from rapidly developing knowledge applications in pattern detection and artificial intelligence. Therefore, this study aims to design e-KTP reading software with web-based optical character recognition (OCR) methods. This type of research is a quantitative type, with the approach used to collect data directly from research volunteers. The system that will be made to read the text in the image uses the *Optical Character Recognition* method. The data discussed in this study are the number of cards, as many as 20 e-KTP, the number of attributes, as many as 14 attributes, and the amount of data, as much as 280 data. The technique used in analyzing the data is descriptive qualitative, and quantitative analysis. From the results of research that have been carried out in detecting attributes, one-KTP produces an accuracy of 100 percent, the results of testing 1 of several attributes found are not nil at 98.09 percent, and the results of testing two which detect categories of e-KTP that are still good and those that are not. Good at 67.61 percent.

## 1. Introduction

Perkembangan teknologi saat ini telah banyak memberi pengaruh terhadap perkembangan ilmu pengetahuan, salah satunya adalah dalam hal pengenalan karakter. Pengenalan karakter adalah teknik untuk mendeteksi segmentasi dan mengidentifikasi karakter dalam gambar. Lebih tepatnya, pengenalan karakter adalah proses mendeteksi dan mengenali karakter dalam gambar input dan di konversikan ke kode standar AS untuk pertukaran informasi atau setara yang dapat diedit mesin lainnya (Sharma et al., 2012). Di Indonesia, pengelolaan yang membutuhkan data pribadi sebagian besar masih manual. Seringkali orang menyalin e-KTP untuk memenuhi berbagai kebutuhan manajemen. Bergantung pada tujuan e-KTP, penggunaan e-KTP saat ini tentu tidak praktis (Abdullah & Muhammad, 2021; Mursari & Wibowo, 2021; Sulstiyono & Saian, 2019). Pemerintah Republik Indonesia telah memperkenalkan Kartu Tanda Penduduk Elektronik Indonesia (e-KTP) sebagai kartu identitas bagi penduduk Indonesia. Dalam e-KTP tersebut terdapat 5 informasi penting antara lain NIK, biodata, foto, tanda tangan, dan sidik jari (Afifah et al., 2020;

\*Corresponding author.

E-mail addresses: [agungtriyudi@civitas.unas.ac.id](mailto:agungtriyudi@civitas.unas.ac.id) (Agung Triayudi)

Basiroh & Lestari, 2020; Soeseno & Liliana, 2017). e-KTP diharapkan dapat mempermudah pertukaran data bagi masyarakat karena chip dipasang pada kartu. Chip adalah kartu pintar berbasis mikroprosesor dengan memori 8KB untuk menyimpan data biometrik, tanda tangan, foto, dan sidik jari dengan kualitas perekaman tertinggi (Sulstiyono & Saian, 2019; Widya Sari et al., 2014).

Namun, pertukaran data diri untuk keperluan administrasi di Indonesia masih banyak yang dilakukan dengan cara manual, yaitu dengan mengisi formulir kertas ataupun fotokopi e-KTP. Penggunaan e-KTP seperti ini tentu tidak sesuai dengan tujuan diciptakannya e-KTP itu sendiri yang diharapkan dapat memudahkan pertukaran data dengan embedded chip yang terpasang pada kartu. Chip ini berisi data diri pemegang e-KTP, lengkap dengan data biometric dan foto wajah (Abdullah & Muhammad, 2021; Purba et al., 2019). e-KTP memiliki chip berkapasitas 4-8 KB yang memuat data NIK, nama, tempat dan tanggal lahir, jenis kelamin, agama, status perkawinan, golongan darah, alamat, pekerjaan, kewarganegaraan, foto, masa berlaku, tempat dan tanggal dikeluarkan, tanda tangan, serta nama dan nomor induk pegawai yang menandatangani. Salah satu penyebab pertukaran data manual ini masih terjadi adalah karena kurangnya sosialisasi terkait metode ekstraksi data pada e-KTP. Masih banyak masyarakat yang tidak tahu bahwa data pada e-KTP bisa dibaca oleh alat khusus (e-KTPreader) (Hudaya et al., 2021). Penyebab lainnya yang juga turut menghalangi penerapan otomatisasi pertukaran data e-KTP adalah mahalnya harga e-KTP reader yang beredar di pasaran terlepas dari fungsinya yang sederhana. Ini menyebabkan beberapa organisasi kecil nonpemerintah (seperti UMKM atau koperasi kecil) memutuskan untuk tidak menggunakan alat pembaca e-KTP karena pengadaan alat ini sering kali tidak bisa dijustifikasi dengan kebutuhan organisasi dan anggaran yang tersedia. Pertimbangan harga dan kegunaan yang mungkin terlihat tidak signifikan bagi organisasi kecil non-pemerintahan membuat mereka tidak menggunakan alat ini. Padahal penggunaan alat ini akan membantu integrasi data kependudukan di Indonesia.

Teknologi optical character recognition (OCR) memiliki banyak aplikasi dalam pemrosesan dokumen. Banyak dokumen memiliki gambar latar belakang dan gambar latar belakang meningkatkan keamanan atau dampak visual dokumen (Basiroh & Lestari, 2020; K Ibnutama & Suryanata, 2020; Mursari & Wibowo, 2021). Proses ekstraksi karakter tersebut dapat dilakukan dengan metode *Optical Character Recognition* menggunakan pustaka Tesseract (Effendi et al., 2019; Hudaya et al., 2021; K Ibnutama & Suryanata, 2020). Perangkat lunak pembaca e-KTP dengan metode optical character recognition (OCR) berbasis web. Dengan metode ini, e-KTP akan di ekstraksi dengan mengenali pola huruf pada gambar. Teks dalam foto e-KTP akan diekstrak dan kemudian disusun. Metode Optical Character Recognition (OCR) digunakan karena relatif lebih mudah dan hasil ekstraksi akan dapat lebih akurat (Khairi Ibnutama et al., 2019). Optical Character Recognition merupakan sebuah sistem pada komputer yang mampu membaca karakter (huruf, angka maupun simbol tertentu) yang terdapat pada citra atau foto (Patel et al., 2012; Purba et al., 2019). Foto dapat berupa tulisan cetak maupuntulisan tangan. Hasil ekstraksi foto tersebut menghasilkan teks dengan menyesuaikan pola karakter per baris dengan pola yang terdapat pada database OCR. Tingkat akurasi pada OCR sangat bergantung pada kualitas citra dan metode yang digunakan pada pengolahan citra. Kelebihan dari sistem OCR yaitu pengguna dapat memasukkan data tanpa harus menggunakan papan ketik, melainkan dapat menggunakan pena elektronik untuk menulis layaknya menulis di sebuah kertas. Perangkat lunak yang dibangun menggunakan metode *Optical Character Recognition* (OCR) berbasis web, dengan ruang lingkup ekstraksi teks dari foto e-KTP dan pengujian dengan mengukur presentase keakuratan dengan mengenali teks dari kondisi e-KTP yang berbeda-beda (baik atau rusak karena pudar atau pun plastik yang terkelupas). Tujuan penelitian ini adalah merancang dan membangun sistem informasi pembaca teks dari gambar foto e-KTP dengan metode OCR dan melakukan pengujian dan analisis pada hasil sistem informasi yang dibangun. Dengan adanya sistem informasi ini dapat memudahkan pihak yang berkepentingan untuk memindahkan data jika menggunakan e-KTP.

## 2. Method

Jenis penelitian merupakan jenis kuantitatif, dengan pendekatan yang digunakan pengambilan data langsung ke relawan penelitian. Sistem yang akan dibuat untuk membaca tulisan dalam gambar menggunakan metode *Optical Character Recognition*. Data yang digunakan dalam penelitian ini adalah 20 e-KTP. Setiap e-KTP memiliki 14 atribut yang tersimpan dalam sistem yang dihasilkan dengan data ATRT1= NIK, ATRT2= Nama, ATRT3 = Tempat, Tanggal Lahir, ATRT4 = Jenis Kelamin, ATRT5 = Golongan Darah, ATRT6= Alamat, ATRT7= RT, ATRT8 = RW, ATRT9= Kelurahan/Desa, ATRT10 = Kecamatan, ATRT11 = Agama, ATRT12 = Status Perkawinan, ATRT13 = Pekerjaan, ATRT14 = Kewarganegaraan. Sesuai dengan data yang akan dibahas dalam penelitian ini yaitu jumlah kartu sebanyak 20 e-KTP, jumlah attribute sebanyak 14 attribute, dan jumlah data sebanyak 280 data. Penelitian ini menggunakan metode *Optical Character Recognition* (OCR).

Saat menggunakan OCR, proses awal mengubah gambar dengan hanya satu nilai saluran per piksel menjadi skala abu-abu. Artinya, bagian MERAH = HIJAU = BIRU. Setelah berhasil mengonversi gambar ke skala abu-abu, ubah menjadi gambar biner. Langkah ini merupakan langkah awal untuk pengenalan karakter. Dengan melakukan ambang batas (Afifah et al., 2020; Aprillian et al., 2019). Tujuannya adalah untuk menemukan area yang berisi subjek dan latar belakang. Prosesnya adalah mengubah skala keabuan citra menjadi citra biner. Pertama kita mendefinisikan nilai sebagai nilai pembatas agar sesuai dengan setiap piksel. Metode ini digunakan untuk mengisolasi objek dari latar belakang yang ada (Bezerra & de Oliveira, 2013; Memon et al., 2020). Menentukan lokasi area di mana baris meteran penggunaan daya akan ditampilkan setelah gambar biner. Langkah ini menggunakan teknik blur untuk mengekstrak posisi teks pada gambar. Sebelum pelumasan, batas bawah dan atas ditentukan terlebih dahulu. Perhitungan berurutan kemudian dilakukan pada piksel hitam garis. Jika jumlah piksel hitam berurutan kurang dari batas ini, semua piksel hitam berurutan dikonversi menjadi putih. Langkah ini diikuti dengan langkah segmentasi. Langkah ini bertujuan untuk mendapatkan setiap nomor. Setiap piksel dari suatu objek diberi label yang sama, dan piksel latar belakang tidak diberi label. (Sandhika & Erik, 2014). Langkah selanjutnya adalah mendekati pemilihan template. Ini adalah saat proses ini membandingkan gambar input yang berisi template tertentu dengan template database. Nilai kesesuaian tertinggi menunjuk ke gambar template yang paling cocok dengan gambar input. Tesseract adalah mesin OCR open source. Dikembangkan oleh Hewlett-Packard antara tahun 1984 dan 1994 (Mursari & Wibowo, 2021). Hal ini sangat portabel, lebih berfokus pada pengurangan bias daripada akurasi. Saat ini, tesseract versi 3.01 dan tersedia berbagai bahasa (Mursari & Wibowo, 2021; Patel et al., 2012). OCR adalah metode mendigitalkan tulisan tangan atau mengekstraksi teks ke dalam format digital yang dapat dimainkan. Metode OCR melibatkan tiga proses utama *pre-processing*, *pengenalan (recognition)* dan *post-procesing* (Ibnutama & Suryanata, 2020). Pengolahan citra adalah suatu teknik yang digunakan untuk mengolah atau memanipulasi suatu citra dalam bentuk pengolahan citra dua dimensi (Effendi et al., 2019; Sulstiyono & Saian, 2019). Greyscale image merupakan ukuran dari derajat keabuan tiap piksel (Aisyah et al., 2015; Firmanto et al., 2019).

### 3. Result and Discussion

#### Results

Lingkungan implementasi adalah spesifikasi perangkat lunak yang digunakan untuk mengimplementasikan sistem. Karakteristik teknis dari peralatan yaitu *Processor; Intel Core i5 dual-core 1.6 GHz (Macbook Air 2018), Random Access Memory; 8 GB, Storage; SSD 128 GB*. Adapun spesifikasi teknis perangkat lunak yang digunakan untuk menguji sistem ini yaitu sistem Operasi: MacOS, *Web Server: XAMPP, Database Server: MySQL, Library Tesseract OCR (PHP), Text Editor: Sublime Text, dan Web Browser: Google Chrome*. Pengujian sistem pertama akurasi deteksi e-KTP menggunakan 20 data e-KTP menunjukkan jumlah e-KTP yang dapat dibaca dalam teks. Tes kedua adalah untuk memverifikasi bahwa properti e-KTP dibaca dengan benar. Berdasarkan 16 atribut yang ada, e-KTP memiliki banyak atribut nihil. Tes terakhir adalah untuk kesalahan baca e-KTP, Anda dapat melihat bahwa ada banyak kesalahan baca untuk setiap e-KTP dari 14 properti yang ada. Pengujian sistem dimulai dengan pendeteksian e-KTP. Jika e-KTP berhasil, itu akan menampilkan beberapa baris/array dan menyimpannya di database yang ada. Berdasarkan hasil penelitian yang dilakukan, akurasi dan pendeteksian e-KTP pada sistem dihitung dengan rumus (2). Semua tabel dan gambar harus jelas/tidak kabur/buram. Ukuran huruf pada tabel dan gambar harus dapat dibaca oleh mata normal dengan mudah. Posisi tabel atau gambar disuatu halaman, sebaiknya terletak dibagian atas atau bawah halaman pada tiap kolom. Meletakkan tabel atau gambar ditengah halaman atau paragraf supaya dihindari. Tabel dan gambar diletakkan pada posisi tengah setiap kolom (*center alignment*). Akurasi e-KTP diperoleh sebesar 75 %. Pengujian pertama, nilai yang dicari dalam pengujian ini adalah keakuratan sistem pendeteksian kata dalam e-KTP beberapa atribut yang ditemukan tidak terbaca. Berdasarkan data yang dimasukkan, hasil pengujian I untuk menguji keakuratan sistem dalam mendeteksi diperoleh sebesar 98,09 %. Hasil pengujian pertama terhadap sistem disajikan pada Tabel 1.

Tabel 1. Pengujian I terhadap Sistem

No	Citra KTP	Hasil Output	Keterangan	Presentase	
1		ATRT1	3204110207830009	Seluruh karakter dapat dikenali oleh sistem.	100%
		ATRT2	Wahyu Andhiko Nuswantoro		
		ATRT3	Madiun, 02-07-1983		
		ATRT4	Laki-Laki		
		ATRT5	AB		
		ATRT6	Gading Junti Asri Blok H4 No1		
		ATRT7	009		

No	Citra KTP	Hasil Output	Keterangan	Presentase	
2		ATRT8	005	Seluruh karakter dapat dikenali oleh sistem.	100%
		ATRT9	Sangkanhurip		
		ATRT10	Katapang		
		ATRT11	Katholik		
		ATRT12	Belum Kawin		
		ATRT13	Karyawan Swasta		
		ATRT14	WNI		
		ATRT1	3312050610960004		
		ATRT2	Abdurrahim Budiman Wicaksono		
		ATRT3	Wonogiri, 06-10-1996		
		ATRT4	Laki-Laki		
		ATRT5	-		
		ATRT6	Dusun Tegalombo		
		ATRT7	002		
ATRT8	004				
ATRT9	Banyakprodo				
ATRT10	Tirtomoyo				
ATRT11	Islam				
ATRT12	Kawin				
ATRT13	Karyawan Swasta				
ATRT14	WNI				
7		ATRT1	3172024512871002	Seluruh karakter dapat dikenali oleh sistem.	100%
		ATRT2	Fitriyah		
		ATRT3	Banyumas, 05-12-1998		
		ATRT4	Perempuan		
		ATRT5	-		
		ATRT6	Jl. Budi Mulia No.20		
		ATRT7	009		
		ATRT8	008		
		ATRT9	Pademangan Barat		
		ATRT10	Pademangan		
		ATRT11	Islam		
		ATRT12	Belum Kawin		
		ATRT13	Pelajar/Mahasiswa		
		ATRT14	WNI		
9		ATRT1	7404271108790001	Seluruh karakter dapat dikenali oleh sistem.	100%
		ATRT2	Abdul Wahid		
		ATRT3	Kanawa, 11-08-1979		
		ATRT4	Laki-Laki		
		ATRT5	-		
		ATRT6	Dusun Sampuabalo		
		ATRT7	000		
		ATRT8	000		
		ATRT9	Sampuabalo		
		ATRT10	Siotapina		
		ATRT11	Islam		
		ATRT12	Kawin		
		ATRT13	Petani/pekebun		
		ATRT14	WNI		
13		ATRT1	2104010701630001	Seluruh karakter dapat dikenali oleh sistem.	100%
		ATRT2	Alias Wello		
		ATRT3	Dabo Singkep, 07-01-1963		
		ATRT4	Laki-Laki		
		ATRT5	B		
		ATRT6	Jl. Imam Bonjol Kampung Mentigi Laut		
		ATRT7	002		

No	Citra KTP	Hasil Output	Keterangan	Presentase	
14		ATRT8	001	Terdapat 1 variabel yang terbaca oleh sistem namun tidak tepat sesuai dengan citra KTP.	92,85%
		ATRT9	Tanjung Uban Kota		
		ATRT10	Bintan Utara		
		ATRT11	Islam		
		ATRT12	Kawin		
		ATRT13	Wiraswasta		
		ATRT14	WNI		
		ATRT1	9215061403960006		
		ATRT2	Samsudin		
		ATRT3	Karawang, 14-7-1996		
		ATRT4	Laki-Laki		
		ATRT5	-		
		ATRT6	Dusun Jatimulya		
		ATRT7	016		
15		ATRT8	004	Terdapat 1 variabel yang terbaca oleh sistem namun tidak tepat sesuai dengan citra KTP.	92,85%
		ATRT9	Karyamulya		
		ATRT10	Batujaya		
		ATRT11	Islam		
		ATRT12	Belum Kawin		
		ATRT13	Pelajar/Mahasiswa		
		ATRT14	WNI		
		ATRT1	34750411106020005		
		ATRT2	Yufka Yudha Pratama		
		ATRT3	Jakarta, 11-10-2002		
		ATRT4	Laki-Laki		
		ATRT5	-		
		ATRT6	Cililitan Besar		
		ATRT7	008		
ATRT8	009				
ATRT9	Cililitan				
ATRT10	Kramatjati				
ATRT11	Islam				
ATRT12	Belum Kawin				
ATRT13	Pelajar/Mahasiswa				
ATRT14	WNI				
<b>Average Pembacaan Citra KTP</b>				<b>98,09%</b>	

Pengujian kedua, berdasarkan hasil pendeteksian properti e-KTP yang masih baik/bagus pada pengujian 1. Sistem mendeteksi mana yang e-KTP masih bagus/ baik dan mana yang e-KTP sudah mulai rusak/tidak baik. Berdasarkan pengujian kedua yang dilakukan, diperoleh hasil 51, 42 %. Hasil pengujian disajikan pada [Tabel 2](#).

**Tabel 2. Pengujian II**

No	Citra KTP	Hasil Output	Keterangan	Presentase	
1		ATRT1	6302190403730002	Terdapat 11 variabel yang tidak terbaca oleh sistem, 1 diantaranya terbaca namun tidak sama dengan citra KTP.	21,42%
		ATRT2	Sulok		
		ATRT3	Nganjuk, 04-03-1973		
		ATRT4	Laki-Laki		
		ATRT5	Tidak Terbaca Sistem		
		ATRT6	Tidak Terbaca Sistem		
		ATRT7	Tidak Terbaca Sistem		
		ATRT8	Tidak Terbaca Sistem		
		ATRT9	Tidak Terbaca Sistem		
		ATRT10	Tidak Terbaca Sistem		
		ATRT11	Tidak Terbaca Sistem		
		ATRT12	Tidak Terbaca Sistem		
		ATRT13	Tidak Terbaca Sistem		

No	Citra KTP	Hasil Output	Keterangan	Presentase	
2		ATRT14	Tidak Terbaca Sistem	Terdapat 2 variabel yang terbaca oleh sistem namun tidak sama dengan citra KTP.	85,71%
		ATRT1	3307159706906003		
		ATRT2	1b4j Hajar		
		ATRT3	Wonosobo, 07-06-1990		
		ATRT4	Laki-Laki		
		ATRT5	-		
		ATRT6	Dukuh Sembir		
		ATRT7	007		
		ATRT8	006		
		ATRT9	Dempel		
		ATRT10	Kalibawang		
		ATRT11	Islam		
		ATRT12	Kawin		
		ATRT13	Wiraswasta		
ATRT14	WNI				
3		ATRT1	Tidak Terbaca Sistem	Terdapat 13 variabel yang tidak terbaca oleh sistem, 1 diantaranya terbaca namun tidak sama dengan citra KTP.	7,14%
		ATRT2	Imade Astika Parwata		
		ATRT3	Pasas. 16-04-1974		
		ATRT4	Tidak Terbaca Sistem		
		ATRT5	Tidak Terbaca Sistem		
		ATRT6	Tidak Terbaca Sistem		
		ATRT7	Tidak Terbaca Sistem		
		ATRT8	Tidak Terbaca Sistem		
		ATRT9	Tidak Terbaca Sistem		
		ATRT10	Tidak Terbaca Sistem		
		ATRT11	Tidak Terbaca Sistem		
		ATRT12	Tidak Terbaca Sistem		
		ATRT13	Tidak Terbaca Sistem		
		ATRT14	Tidak Terbaca Sistem		
<b>Average Pembacaan Citra KTP</b>				<b>51,42%</b>	

Hasil pengujian sistem secara keseluruhan pengenalan kata e-ktp menggunakan metode optical character recognition disajikan pada Tabel 3.

Tabel 3. Pengujian

No	Detail	Jumlah Valid	Jumlah Nihil	Presentase
1	Deteksi e-KTP	15	5	75%
2	Pengujian 1	206	210	98,09%
3	Pengujian 2	36	70	51,42%
4	Keseluruhan	206	210	98,09%

Hasil pengujian yang diperoleh dievaluasi sesuai dengan metode matriks konfusi. Hasil evaluasi uji validasi e-KTP ditunjukkan pada Tabel 5, 6, 7, dan 8.

Tabel 4. Hasil Evaluasi dengan Confusion Matrix Pengujian Deteksi e-KTP

Prediksi	Valid	Nihil
<b>Aktual Valid</b>	Benar Positif (TP=15)	Salah Negatif (FN=0)
<b>Aktual Nihil</b>	Salah Positif (FP=5)	Benar Negatif (TN=0)

Tabel 5. Hasil Evaluasi dengan Confusion Matrix Pengujian I

Prediksi	Valid	Nihil
<b>Aktual Valid</b>	Benar Positif (TP=206)	Salah Negatif (FN=0)
<b>Aktual Nihil</b>	Salah Positif (FP=4)	Benar Negatif (TN=0)

**Tabel 6.** Hasil Evaluasi dengan *Confusion Matrix* Pengujian II

Prediksi Aktual	Valid	Nihil
Valid	Benar Positif (TP=36)	Salah Negatif (FN=30)
Nihil	Salah Positif (FP=4)	Benar Negatif (TN=0)

**Tabel 7.** Hasil Evaluasi dengan *Confusion Matrix* Keseluruhan Pengujian

Prediksi Aktual	Valid	Nihil
Valid	Benar Positif (TP=206)	Salah Negatif (FN=0)
Nihil	Salah Positif (FP=4)	Benar Negatif (TN=0)

Nilai akurasi, memori, akurasi, dan perhitungan FScore dapat diperoleh dari tabel survey sebelumnya pada [Tabel 8](#).

**Tabel 8.** Hasil Akurasi, *Recall*, *Precision*, dan *F-Score*

No	Detail	Presentase	Recall	Presisi	F-Score
1	Deteksi e-KTP	75%	100%	100%	100%
2	Pengujian 1	98,09%	100%	98,09%	98,72%
3	Pengujian 2	51,42%	100%	51,42%	67,61%
4	Keseluruhan	98,09%	100%	98,09%	98,72%

Deteksi e-KTP dengan tingkat akurasi sebesar 75%, *recall* sebesar 100%, *precision* dan *f-score* sebesar 75%. Pengujian 1 dengan tingkat akurasi sebesar 98%, *recall* sebesar 100%, *precision* sebesar 98,09% dan *f-score* sebesar 98,72%. Pengujian 2 dengan tingkat akurasi sebesar 51%, *recall* sebesar 100%, *precision* sebesar 51,42% dan *f-score* sebesar 67,61%.

### Discussion

Pemerintah Republik Indonesia telah memperkenalkan Kartu Tanda Penduduk Elektronik Indonesia (e-KTP) sebagai kartu identitas bagi penduduk (Suriadi & Hasibuan, 2013; Wahyono & Zakiyah, 2020; Yani et al., 2017). Informasi yang disajikan pada e-KTP mencakup lima jenis yaitu NIK, biodata, foto, sidik jari dan tanda tangan (Amrizal et al., 2016; Apandi & Sugianto, 2019). E-KTP dapat mempermudah pertukaran data karena chip yang dipasangka pada kartu. Perangkat lunak pembaca e-KTP dengan metode Optical Character Recognition (OCR) berbasis web. Dengan metode ini, e-KTP akan di ekstraksi dengan mengenali pola huruf pada gambar. Teknologi OCR memiliki banyak aplikasi dalam pemrosesan dokumen. Proses ekstraksi karakter tersebut dapat dilakukan dengan metode *Optical Character Recognition* menggunakan pustaka Tesseract (Effendi et al., 2019; Hudaya et al., 2021; K Ibnutama & Suryanata, 2020). Perangkat lunak yang dirancang berhasil mengekstrak semua data diri yang ada pada e-KTP. Teks dalam foto e-KTP akan diekstrak dan kemudian disusun. Metode OCR digunakan karena relatif lebih mudah dan hasil ekstaksi akan dapat lebih akurat (Khairi Ibnutama et al., 2019). Optical Character Recognition merupakan sebuah sistem pada komputer yang mampu membaca karakter (huruf, angka maupun simbol tertentu) yang terdapat pada citra atau foto (Patel et al., 2012; Purba et al., 2019). Hasil ekstraksi foto tersebut menghasilkan teks dengan menyesuaikan pola karakter per baris dengan pola yang terdapat pada database OCR. Kelebihan dari sistem OCR yaitu pengguna dapat memasukkan data tanpa harus menggunakan papan ketik, melainkan dapat menggunakan pena elektronik untuk menulis layaknya menulis di sebuah kertas.

### 4. Conclusions and Suggestions

Perangkat lunak yang dirancang berhasil mengekstrak semua data diri yang ada pada e-KTP. Meski nilai akurasi OCR yang dicapai pada penelitian ini dianggap belum mencerminkan kondisi nyata mengingat data yang digunakan sangat minim. Dengan adanya penelitian ini dapat memudahkan pihak yang memerlukan data menggunakan e-KTP tanpa harus *meninput* data kembali dan hanya tinggal *upload* di website dan data telah siap untuk digunakan. Maka, saran kepada peneliti yang akan datang diharapkan mampu mengoptimalisasi proses pendeteksian OCR lebih baik lagi. Data penelitian yang akan datang dapat

terhubung dengan data pusat pemerintah agar dapat terintegrasi dan data tersebut menjadi akurat, di samping itu, diharapkan juga pada penelitian yang akan datang dapat mengembangkan aplikasi berbasis website yang lebih luas lagi.

## 5. References

- Abdullah, S. S., & Muhammad, F. D. (2021). Penggunaan e-KTP untuk Registrasi Otomatis Memanfaatkan Sistem OCR Dengan Metode Template Matching Correlation. *Media Jurnal Informatika*, 12(2), 57. <https://doi.org/10.35194/mji.v12i2.1224>.
- Afifah, Y., Sujono, A., & Apribowo, C. H. B. (2020). The Line Segmentation Algorithm of Indonesian Alectronic Identity Card (e-KTP) for Data Digitization. *AIP Conference Proceedings*, 2217(April). <https://doi.org/10.1063/5.0000670>.
- Aisyah, E. S. N., Hayat, A. W., Widanti, P., Prasetya, S. Y., & Iskandar, H. (2015). Analisis Kemiripan Pola Citra Digital Menggunakan Metode Euclidean. *Seminar Nasional Teknologi Informasi Dan Multimedia*, 6–8.
- Amrizal, Tukino, & Latif, A. (2016). Rekayasa Perangkat Lunak Sistem e-KTP Terintegrasi Birokrasi Umum di Kota Batam. *Jurnal Teknologi Dan Sistem Informasi*, 2(2). <https://doi.org/10.25077/TEKNOSI.v2i2.2016.97-108>.
- Apandi, T. H., & Sugianto, C. A. (2019). Algoritma Naive Bayes untuk Prediksi Kepuasan Pelayanan Perekaman e-KTP. *Jurnal Informatika*, 7(2). <https://doi.org/10.30595/juita.v7i2.3608>.
- Aprillian, H. D., Dwi Purnomo, H., & Purwanto, H. (2019). Utilization of Optical Character Recognition Technology in Reading Identity Cards. *Journal of Information Technology and Business*, 2(1), 38–46.
- Basiroh, B., & Lestari, W. (2020). Analysis of Plant *Fragaria Xananassa* Disease Diagnoses Using Production Rules Base on Expert System. *Jurnal Pilar Nusa Mandiri*, 16(1), 25–32. <https://doi.org/10.33480/pilar.v16i1.1174>.
- Bezerra, M. E. R., & de Oliveira, A. L. I. (2013). Optical Character Recognition. *Digital Document Analysis and Processing*, 1(1), 207–230. <https://doi.org/10.3810/pgm.1998.03.410>.
- Effendi, M. M., Permana, A. Y., & Nawangsih, I. (2019). Penerapan Ekstraksi Image ke TXT dengan Optical Character Recognition untuk Otomatis Data Kependudukan. *Simposium Nasional Ilmiah Simponi, November*, 496–502. <https://doi.org/10.30998/simponi.v0i0.484>.
- Firmanto, B., Rikasanti, E., Bramanto, A., & Putra, W. (2019). *Optimasi Hasil Akuisisi Obyek Wajah Menggunakan*. 2, 826–840.
- Hudaya, M. M., Siti Saadah, & Hendy Irawan. (2021). Implementation of Verification and Matching E-KTP with Faster R-CNN and ORB. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(4), 783–793. <https://doi.org/10.29207/resti.v5i4.3175>.
- Ibnutama, K., & Suryanata, M. G. (2020). Ekstraksi Karakter Citra Menggunakan Optical Character Recognition Untuk Pencetakan Nomor Kendaraan Pada Struk Parkir. *Jurnal Media ...*, 4, 1119–1125. <https://doi.org/10.30865/mib.v4i4.2432>.
- Ibnutama, Khairi, Panjaitan, Z., & Ginting, E. F. (2019). Modifikasi Metode Template Matching pada OCR Untuk Meningkatkan Akurasi Deteksi Plat Nomor Kendaraan. *Jurnal Teknologi Sistem Informasi Dan Sistem Komputer TGD*, 2(2), 21–29.
- Memon, J., Sami, M., Khan, R. A., & Uddin, M. (2020). Handwritten Optical Character Recognition (OCR): A Comprehensive Systematic Literature Review (SLR). *IEEE Access*, 8, 142642–142668. <https://doi.org/10.1109/ACCESS.2020.3012542>.
- Mursari, L. R., & Wibowo, A. (2021). The Effectiveness of Image Preprocessing on Digital Handwritten Scripts Recognition with The Implementation of OCR Tesseract. *Computer Engineering and Application*, 10(3), 177–186. <https://doi.org/10.18495/comengapp.v10i3.386>.
- Patel, C., Patel, A., & Patel, D. (2012). Optical Character Recognition by Open source OCR Tool Tesseract: A Case Study. *International Journal of Computer Applications*, 55(10), 50–56. <https://doi.org/10.5120/8794-2784>.
- Purba, A. M., Harjoko, A., & Wibowo, M. E. (2019). Text Detection In Indonesian Identity Card Based On Maximally Stable Extremal Regions. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 13(2), 177. <https://doi.org/10.22146/ijccs.41259>.
- R. Sandhika Galih A., Erik, M. L. H. (2014). Penerapan Teknik Ocr ( Optical Character Recognition ) pada Aplikasi Terjemahan Kitab Fiqih Safinah an-Naja Menggunakan Readiris. *Seminar Nasional Informatika, 2014(semnasIF)*, 60–69.
- Sharma, O. P., Ghose, M. K., & Shah, K. B. (2012). An Improved Zone Based Hybrid Feature Extraction Model for Handwritten Alphabets Recognition Using Euler Number. *International Journal of Soft Computing and Engineering (IJSCE)*, 2(2), 504–508.

- Soeseno, J. H., & Liliana. (2017). Segmentasi Area KTP dari Image untuk Otomatisasi Pembacaan Data. *Jurnal Infra Petra*, 5(1), 1-5.
- Sulstiyo, A., & Saian, P. O. N. (2019). Perancangan Sistem Text Extraction Menggunakan Library Tesseract OCR untuk Mengambil Nomor Induk Kependudukan pada Foto Kartu Tanda Penduduk ( Studi Kasus : PT. Bank ABC , Tbk ). April.
- Suriadi, S., & Hasibuan, M. A. (2013). Teknis Perekaman Data dan Penyaluran Elektronik Kartu Tanda Penduduk di Lubuk Pakam. *JPPUMA: Jurnal Ilmu Pemerintahan Dan Sosial Politik UMA (Journal of Governance and Political Social UMA)*, 1(2). <https://doi.org/10.31289/jppuma.v1i2.560>.
- Wahyono, W., & Zakiyah, N. (2020). Model Exploration for Community Satisfaction with E-KTP Service Quality as Mediation. *Dinamika Pendidikan*, 15(1). <https://doi.org/10.15294/dp.v15i1.24892>.
- Widya Sari, E., Rianto, A., & Diatinari Andarawasih, S. (2014). Itechs 201 4. *The 1st International Conference on Information Technology and Security*, 1(1), 111-117.
- Yani, A. A., Yunus, A. Y., & Latief, M. I. (2017). Participation Dynamics of Voters Using ID Card in Local Elections: A Case Study of the 2015 Local Election in South Sulawesi. *JSP: Jurnal Ilmu Sosial Dan Ilmu Politik*, 21(1). <https://doi.org/10.22146/jsp.28700>.