

Implementasi dan Evaluasi Akurasi Optical Character Recognition (OCR) untuk Verifikasi KTP pada Aplikasi Asuransi

Firstio Daniel Anmar^{*1}, Muhamad Miftahudin², F.R.Dwi Febriantoro³

^{1,2}Prodi Teknik Informatika, ³Prodi Teknologi Informasi, Fakultas Informatika dan Komputer,
Universitas Binaniaga Indonesia, Bogor

e-mail: ^{*1}firstioanmar@gmail.com, ²m.miftahudin@unbin.ac.id, ³dwifebriantoro@unbin.ac.id

INFORMASI ARTIKEL

Histori artikel:

Naskah masuk, 20 Juni 2026

Direvisi, 26 Juli 2026

Diiterima, 31 Juli 2026

ABSTRAK

Abstract- Identity verification using the Indonesian identity card (KTP) in insurance registration applications is generally still performed manually, making the process time-consuming, prone to human error, and highly dependent on the quality of uploaded documents. This study aims to implement Optical Character Recognition (OCR) technology to automate KTP verification and to evaluate its accuracy so that the process becomes faster and more reliable. The method used is Research and Development (R&D) with a quantitative approach through requirement analysis, observation, questionnaires, and system testing using Tesseract.js on a limited sample of five KTP images and six respondents. The development was carried out through image preprocessing, text extraction, data parsing, logical validation, and verification score calculation. Four indicators were measured separately to avoid conflating them: OCR confidence, per-field extraction accuracy, expert-tested system feasibility, and user satisfaction. The average OCR confidence ranged from 39%–62% depending on image quality. Despite this moderate confidence, the name and date of birth fields were extracted with high accuracy (85% to 100%), whereas the identity number (NIK) field failed to match the ground truth on every tested sample, making it the primary limitation of the current implementation. The system accelerates data extraction to an average of 2–3 seconds per document, much faster than manual verification. Expert black-box testing showed a system feasibility of 100%, while user testing resulted in an overall satisfaction rate of 98%; both figures reflect usability and functional completeness rather than OCR reading accuracy. Given the small sample size and the unresolved NIK-reading limitation, the findings should be regarded as preliminary evidence that OCR can accelerate KTP verification, while further work with a larger dataset and dedicated numeric-character handling is required before the accuracy improvement can be generalized.

Kata Kunci:

Optical Character Recognition,
Verifikasi Identitas,
Aplikasi Asuransi

Abstrak- Proses verifikasi identitas menggunakan Kartu Tanda Penduduk (KTP) pada aplikasi pendaftaran asuransi umumnya masih dilakukan secara manual, sehingga memakan waktu lama, rawan kesalahan manusia, serta sangat bergantung pada kualitas dokumen yang diunggah. Penelitian ini bertujuan mengimplementasikan teknologi Optical Character Recognition (OCR) untuk mengotomatisasi verifikasi KTP sekaligus mengevaluasi tingkat akurasi agar proses menjadi lebih cepat dan andal. Metode yang digunakan adalah Research and Development (R&D) dengan pendekatan kuantitatif melalui analisis kebutuhan, observasi, penyebaran kuesioner, serta pengujian sistem menggunakan Tesseract.js pada sampel terbatas berupa lima citra KTP dan enam responden. Pengembangan dilakukan melalui tahapan preprocessing citra, ekstraksi teks, parsing data, validasi logika, dan penghitungan skor verifikasi. Empat indikator diukur secara terpisah agar tidak saling bertukar, yaitu confidence OCR, akurasi ekstraksi per bidang data (field), kelayakan sistem hasil uji ahli, dan kepuasan pengguna. Nilai confidence OCR rata-rata berada pada rentang 39%–62% bergantung pada kualitas gambar. Meskipun confidence tergolong menengah, bidang nama dan tanggal lahir terbaca dengan akurasi tinggi (85% hingga 100%), sedangkan bidang Nomor Induk Kependudukan (NIK) tidak cocok dengan data sebenarnya pada seluruh sampel yang diuji sehingga menjadi keterbatasan utama implementasi saat ini. Sistem mampu mempercepat proses ekstraksi data menjadi rata-rata 2–3 detik per dokumen, jauh lebih cepat dibandingkan verifikasi manual. Uji coba ahli dengan pendekatan black box menunjukkan kelayakan sistem sebesar 100%, sedangkan uji coba pengguna menghasilkan tingkat kepuasan 98%; kedua angka tersebut mencerminkan usability dan kelengkapan fungsional sistem, bukan akurasi pembacaan OCR. Mengingat ukuran sampel yang kecil dan keterbatasan pembacaan NIK yang belum teratasi, temuan ini merupakan bukti awal bahwa OCR dapat mempercepat proses verifikasi KTP, sementara peningkatan akurasi masih memerlukan pengujian pada dataset yang lebih besar serta penanganan khusus

untuk karakter numerik sebelum dapat digeneralisasi.

Copyright © 2026 LPPM - STMIK IKMI Cirebon
This is an open access article under the CC-BY license

Penulis Korespondensi:

Muhamad Miftahudin

Program Studi Teknik Informatika,
Fakultas Informatika dan Komputer
Universitas Binaniaga Indonesia
Jl. Raya Pajajaran No.100, RT.02/RW.03, Bantarjati,
Kec. Bogor Utara, Kota Bogor, Jawa Barat 16153
Email: miftahudin@unbin.ac.id

1. Pendahuluan

Kemajuan teknologi informasi telah membawa perubahan signifikan dalam pengelolaan data dan dokumen identitas penduduk. Salah satu dokumen penting di Indonesia adalah Kartu Tanda Penduduk (KTP) yang dilengkapi chip elektronik untuk menyimpan data penting seperti nama, Nomor Induk Kependudukan (NIK), alamat, serta data biometrik pemilik kartu. Inovasi ini diharapkan mampu meningkatkan efisiensi pengelolaan data kependudukan sekaligus mengurangi potensi penyalahgunaan data [1]. Namun pada praktiknya, pemanfaatan chip tersebut masih terbatas karena alat pembaca KTP relatif mahal, sehingga banyak organisasi, khususnya organisasi kecil, tetap melakukan verifikasi identitas secara manual yang memakan waktu dan rentan terhadap kesalahan manusia [1].

Kondisi serupa terjadi pada perusahaan pialang (broker) asuransi tempat penelitian ini dilakukan. Dalam mekanisme yang berjalan, calon nasabah mengisi formulir digital berisi informasi pribadi seperti nama lengkap, NIK, tempat dan tanggal lahir, jenis kelamin, alamat email, serta nomor telepon, kemudian mengunggah foto atau hasil pemindaian KTP sebagai dokumen pendukung. Tim administrasi selanjutnya memeriksa satu per satu kesesuaian antara data yang diketik pada formulir dengan informasi yang tertera pada KTP. Permasalahan yang paling sering muncul adalah kualitas file KTP yang tidak memenuhi standar, seperti resolusi rendah, pencahayaan tidak merata, posisi miring atau terpotong, bahkan dokumen yang diunggah bukan KTP. Selain itu, sering terjadi ketidaksesuaian antara data input dengan data pada KTP akibat kesalahan pengetikan, sehingga banyak pendaftaran ditolak oleh pihak asuransi meskipun kesalahan tersebut bersifat teknis. Ketergantungan pada pemeriksaan manual juga membuat proses tidak skalabel ketika volume pendaftaran meningkat.

Hasil survei awal terhadap pengguna sistem dan staf administrasi menunjukkan bahwa 46,9% responden sering mengalami kesalahan atau

ketidakakuratan dalam proses verifikasi manual KTP, 31,3% menyatakan kadang-kadang mengalami hal serupa, dan hanya 21,9% yang jarang mengalami kesalahan. Selain itu, 59,4% responden merasa bahwa proses verifikasi manual sangat menghambat efisiensi kerja mereka. Temuan ini menegaskan bahwa tingkat akurasi dan efisiensi proses verifikasi manual masih rendah dan menjadi permasalahan mendasar dalam sistem yang berjalan.

Teknologi Optical Character Recognition (OCR) hadir sebagai solusi yang mampu mengekstraksi data secara otomatis dari gambar dokumen. OCR bekerja melalui tahapan pre-processing, segmentasi karakter, ekstraksi fitur, dan pengenalan karakter untuk mengonversi teks pada citra menjadi teks digital yang dapat diproses lebih lanjut. Penelitian terdahulu menunjukkan bahwa otomatisasi ekstraksi data KTP berbasis OCR mampu mengurangi kesalahan manusia dan menurunkan waktu verifikasi secara signifikan dibandingkan metode manual [2]. Meskipun demikian, keberhasilan OCR sangat dipengaruhi oleh kualitas citra; gambar yang buram, terlalu gelap, tidak sejajar, atau mengandung elemen non-teks seperti logo dan watermark dapat menurunkan akurasi pengenalan [3].

Sejumlah penelitian telah menerapkan OCR pada berbagai domain. Tesseract OCR berhasil diterapkan untuk pengenalan aksara Jawa pada aplikasi Android dan menunjukkan fleksibilitas teknologi ini dalam menangani berbagai jenis teks [4]. Pengenalan teks pada kartu identitas Vietnam berbasis deep features network juga memberikan hasil yang menjanjikan [5], sementara kajian teknologi untuk pengembangan aplikasi verifikasi identitas mengidentifikasi OCR sebagai salah satu komponen kunci [6]. Tinjauan komprehensif OCR untuk teks non-Latin memberikan perspektif tambahan mengenai tantangan pengenalan karakter [7]. Pada domain kesehatan, OCR terbukti meningkatkan efisiensi penelusuran rekam medis [8] dan ekstraksi informasi resep menggunakan OCR-Tesseract [9], sedangkan pada domain transportasi OCR dikombinasikan dengan REST API untuk

deteksi kendaraan [10]. Penelitian yang secara khusus membahas ekstraksi data KTP, seperti implementasi sistem ekstraksi dan validasi e-KTP untuk organisasi non-pemerintah [1], pengembangan Indonesian ID Card Extractor berbasis OCR dan post-processing bahasa alami [2], optimasi Tesseract OCR melalui peningkatan kualitas citra e-KTP [12], serta pemeriksaan KTP berbasis OCR dan pengenalan komponen kartu [13], terbukti mampu mengurangi kesalahan manusia dan mempercepat proses dibandingkan verifikasi manual. Pendekatan berbasis layanan cloud, seperti ekstraksi informasi KTP menggunakan Google Cloud Vision [14], juga menunjukkan akurasi yang baik namun bergantung pada koneksi internet dan biaya penggunaan API, sehingga kurang sesuai untuk kebutuhan pemrosesan lokal yang ringan seperti pada penelitian ini. Namun, keempat penelitian tersebut berfokus pada tahap ekstraksi dan validasi data KTP secara mandiri (standalone) dan belum mengintegrasikan hasil ekstraksi dengan alur bisnis suatu aplikasi transaksional, belum menghitung skor kepercayaan gabungan antar field, dan belum menguji kelayakannya melalui pengujian ahli maupun pengguna. Berbeda dengan penelitian-penelitian tersebut, kebaruan penelitian ini terletak pada penerapan pustaka Tesseract.js secara terintegrasi pada proses verifikasi KTP di dalam aplikasi pendaftaran asuransi berbasis web yang sudah berjalan, dengan menambahkan tiga kontribusi utama: (1) pasca-pemrosesan berbasis aturan (rule-based parsing) dan kamus wilayah untuk menormalkan hasil OCR, (2) pencocokan otomatis antara hasil ekstraksi dan data formulir yang diinput calon nasabah, serta (3) penghitungan skor verifikasi gabungan sebagai dasar keputusan lolos, gagal, atau tinjauan manual, yang sekaligus dievaluasi kelayakannya melalui pengujian ahli dan pengguna.

Berdasarkan uraian tersebut, penelitian ini bertujuan mengembangkan dan menguji sistem verifikasi KTP otomatis berbasis OCR pada aplikasi pendaftaran asuransi untuk meningkatkan akurasi identifikasi informasi pada KTP yang diunggah nasabah serta mengurangi tingkat penolakan akibat kesalahan verifikasi data. Hipotesis yang diajukan adalah penerapan OCR berpengaruh positif terhadap peningkatan akurasi hasil verifikasi identitas nasabah (H1) dan terhadap peningkatan efisiensi waktu verifikasi dibandingkan metode manual (H2).

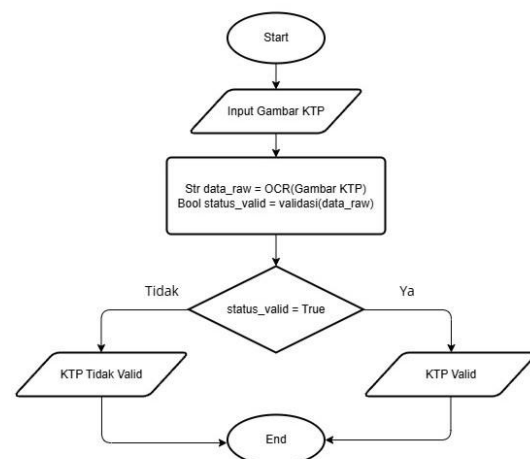
2. Metode Penelitian

Jenis penelitian ini adalah penelitian pengembangan (Research and Development/R&D) dengan pendekatan kuantitatif melalui eksperimen. Pendekatan ini dipilih karena memungkinkan peneliti mengukur variabel penelitian berupa

akurasi hasil ekstraksi data KTP dan waktu yang dibutuhkan untuk proses verifikasi, menganalisisnya secara statistik, serta menguji hipotesis yang telah dirumuskan. Rancangan penelitian meliputi pengumpulan citra KTP dalam format digital, ekstraksi data menggunakan OCR, dan analisis tingkat akurasi serta waktu proses verifikasi. Model prosedural yang digunakan adalah prototyping, yaitu sistem dikembangkan mulai dari versi awal yang sederhana kemudian diuji, dievaluasi, dan disempurnakan secara berulang (iteratif) hingga mencapai kinerja yang stabil dan sesuai kebutuhan pengguna.

2.1 Metode yang Diusulkan

Model teoritis algoritma OCR yang digunakan menjelaskan bagaimana sistem melakukan ekstraksi data dari KTP melalui serangkaian langkah terstruktur, sebagaimana ditunjukkan pada Gambar 1. Proses dimulai dari penerimaan input berupa citra KTP hasil unggahan pengguna. Citra tersebut kemudian diproses oleh mesin OCR yang mendeteksi area-area teks pada gambar dan melakukan transkripsi karakter sehingga menghasilkan data mentah berupa string. Data mentah tidak serta-merta dapat digunakan, sehingga sistem menjalankan fungsi validasi yang menguji kesesuaian isi data dengan aturan atau pola yang berlaku, misalnya panjang NIK harus 16 digit serta kelengkapan field seperti nama. Apabila hasil validasi bernilai benar, sistem mengeluarkan keluaran KTP valid; sebaliknya, KTP dinyatakan tidak valid dan pengguna diminta mengunggah ulang citra dengan kualitas yang lebih baik.



Gambar 1 Flowchart algoritma OCR

Secara teknis, penelitian ini menggunakan pendekatan hibrida yang mengombinasikan ekstraksi teks menggunakan Tesseract.js versi 6.0.1

sebagai mesin OCR, pemrosesan pasca-OCR berbasis aturan, serta rekonsiliasi data dengan basis data relasional. Pelaksanaan eksperimen dilakukan melalui lima langkah berurutan sebagai berikut:

- a. Pengumpulan dataset, yaitu mengumpulkan lima citra e-KTP dengan anotasi ground truth untuk field utama (NIK, nama, tempat/tanggal lahir, alamat) dengan memperhatikan variasi kondisi gambar seperti pencahayaan, kualitas, dan noise. Ground truth ditetapkan dengan mencocokkan data hasil ekstraksi terhadap data aktual yang tertera pada masing-masing KTP secara manual oleh peneliti. Jumlah sampel ini bersifat terbatas dan digunakan sebagai studi kasus awal (pilot study) untuk menguji kelayakan pendekatan yang diusulkan, sehingga hasil akurasi yang diperoleh belum dapat digeneralisasi pada populasi KTP yang lebih luas.
- b. Ekstraksi OCR menggunakan Tesseract.js dengan konfigurasi bahasa Indonesia (`language="ind"`) serta page segmentation mode (`psm`) dan OCR engine mode (`oem`) yang dioptimalkan untuk dokumen KTP. Proses mencakup segmentasi citra menjadi blok teks, baris, kata, dan karakter; ekstraksi fitur oleh jaringan LSTM; serta klasifikasi karakter beserta tingkat keyakinan (`confidence score`) dan waktu pemrosesan.
- c. Pasca-pemrosesan dan ekstraksi field, yaitu normalisasi hasil OCR, penghapusan karakter non-alfanumerik, serta pemetaan teks ke field KTP menggunakan regex dan rule-based parsing. Kamus provinsi/kabupaten/kota digunakan untuk koreksi ejaan.
- d. Validasi logika dan cross check, yaitu penerapan aturan seperti panjang NIK 16 digit, validasi format tanggal, pemeriksaan format nama, dan validasi minimal token alamat. Data yang tidak sesuai ditandai (flag) untuk ditinjau secara manual.
- e. Perhitungan skor verifikasi untuk menentukan kelayakan data hasil OCR, dihitung dengan menjumlahkan hasil kali bobot setiap field dengan nilai `confidence-nya`, seperti pada persamaan (1).

$$\text{Score} = \sum (B_i \times C_i) \quad (1)$$

Dengan B_i adalah bobot untuk field ke- i , C_i adalah nilai `confidence` hasil OCR pada field ke- i , dan `Score` adalah nilai akhir verifikasi dokumen. Keluaran yang diukur meliputi distribusi skor per dokumen serta persentase data yang otomatis lolos, gagal, atau perlu tinjauan manual.

2.2 Prosedur Pengembangan

Prosedur pengembangan sistem ekstraksi data KTP berbasis OCR dilakukan melalui beberapa tahapan sistematis sebagaimana ditunjukkan pada Gambar 2, yaitu: (1) pengumpulan data berupa citra KTP digital berformat .jpg dan .png yang diperoleh dari unggahan pengguna; (2) penetapan algoritma OCR, yaitu Tesseract.js, pustaka OCR berbasis JavaScript yang mendukung bahasa Indonesia dan dapat berjalan pada sisi client maupun server; (3) analisis dan desain sistem dengan mengidentifikasi kebutuhan fungsional serta menyusun desain dalam bentuk model UML (use case diagram, activity diagram, sequence diagram); (4) pembangunan prototipe menggunakan bahasa pemrograman PHP dengan framework Laravel dan basis data MySQL, di mana modul Tesseract.js diintegrasikan untuk melakukan ekstraksi data langsung dari citra yang diunggah; serta (5) uji coba produk menggunakan dataset KTP yang terkumpul. Jika hasil pengujian belum sesuai, sistem dikembalikan ke tahap analisis dan desain untuk diperbaiki secara iteratif hingga diperoleh tingkat akurasi dan keandalan yang memadai.

2.3 Uji Coba Produk

Uji coba produk dilakukan melalui dua tahapan, yaitu uji coba ahli dan uji coba pengguna, dengan pemilihan subjek secara purposive sampling. Uji coba ahli melibatkan dua orang ahli di bidang teknologi informasi dan menggunakan pendekatan black box testing, sehingga ahli hanya berinteraksi dengan tampilan sistem tanpa perlu mengetahui kode di dalamnya. Pengujian difokuskan pada empat skenario utama, yaitu kemampuan sistem menerima input data KTP, kemampuan menampilkan hasil ekstraksi data, kecepatan proses verifikasi, dan kemudahan penggunaan sistem.

Uji coba pengguna melibatkan enam orang pegawai asuransi yang memiliki pengalaman langsung dalam proses verifikasi identitas nasabah. Setiap responden menggunakan aplikasi sesuai alur kerja yang ditentukan, mulai dari mengunggah citra KTP, meninjau hasil ekstraksi, hingga melakukan verifikasi data, kemudian mengisi kuesioner berskala Likert 5 poin yang mencakup aspek kemudahan penggunaan, kejelasan informasi, kecepatan serta responsivitas sistem, dan kepuasan secara keseluruhan. Data primer diperoleh langsung dari hasil ekstraksi sistem, waktu pemrosesan, dan umpan balik responden, sedangkan data sekunder diperoleh dari literatur dan dokumentasi teknis yang relevan. Jumlah responden ahli (dua orang) dan responden pengguna (enam orang) tergolong kecil sehingga hasil pengujian ini bersifat indikatif terhadap kelayakan awal sistem, bukan generalisasi penuh terhadap seluruh pengguna potensial.

2.4 Teknik Analisis Data

Data hasil uji ahli dianalisis dengan membandingkan hasil aktual terhadap hasil yang diharapkan pada setiap skenario. Persentase keberhasilan fungsi dihitung menggunakan persamaan (2), sedangkan data kuesioner pengguna dianalisis dengan menghitung persentase kelayakan menggunakan persamaan (3).

$$P = \frac{(\text{Jumlah Skenario Berhasil} / \text{Jumlah Skenario}) \times 100\%}{(2)}$$

$$P = \frac{(\text{Skor Perolehan} / \text{Skor Maksimal}) \times 100\%}{(3)}$$

Hasil perhitungan persentase kemudian diinterpretasikan berdasarkan kategori kelayakan, yaitu 81–100% sangat layak, 61–80% layak, 41–60% cukup layak, 21–40% kurang layak, dan 0–20% tidak layak. Selanjutnya, kinerja sistem diukur dari sisi akurasi hasil ekstraksi dan efisiensi waktu verifikasi. Akurasi dihitung dengan membandingkan hasil ekstraksi terhadap data sebenarnya pada dokumen KTP menggunakan persamaan (4), sedangkan efisiensi waktu diukur menggunakan waktu tugas (task time) pada persamaan (5), yaitu lamanya sistem memproses satu kali verifikasi mulai dari input gambar hingga hasil ditampilkan, yang kemudian dirata-ratakan dari seluruh uji coba [11].

$$\text{Akurasi} = \frac{(\text{Jumlah Data Benar} / \text{Jumlah Data Diuji}) \times 100\%}{(4)}$$

$$tij = tend - tstart \quad (5)$$

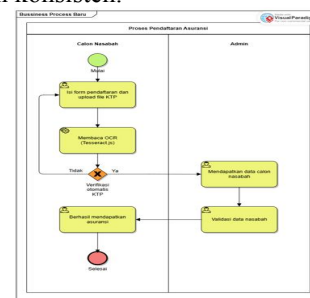
3. Hasil Dan Pembahasan

3.1 Analisis Kebutuhan

Analisis kebutuhan dilakukan melalui penyebaran kuesioner, observasi langsung, dan analisis proses bisnis pada perusahaan pialang asuransi. Hasil kuesioner terhadap staf administrasi menunjukkan seluruh responden (100%) menyatakan aplikasi pendaftaran asuransi digital sangat mudah digunakan dan sangat puas terhadap kinerja serta respons aplikasi. Pada aspek kejelasan informasi, 66,7% responden menilai sangat jelas dan 33,3% menilai jelas; pada aspek kesesuaian informasi, 83,3% menilai sangat sesuai dan 16,7% menilai sesuai; sedangkan seluruh responden menilai tampilan antarmuka sangat bagus. Hasil observasi menunjukkan kualitas citra KTP yang diunggah calon nasabah sangat bervariasi, mulai dari resolusi tinggi, resolusi rendah, hingga pencahayaan yang kurang optimal, sehingga menimbulkan kendala pada pembacaan data secara manual dan menegaskan kebutuhan akan sistem verifikasi otomatis.

3.2 Perancangan Proses Bisnis

Pada proses bisnis lama, seluruh tahapan verifikasi KTP dan penentuan kelayakan pendaftaran bergantung pada pemeriksaan manual oleh admin, yaitu mencocokkan satu per satu data yang diinput calon nasabah dengan file KTP yang diunggah. Pada proses bisnis baru yang dirancang (Gambar 3), setelah file KTP diunggah, sistem secara otomatis menjalankan modul OCR Tesseract.js untuk mengekstraksi informasi pada KTP seperti NIK, nama, tanggal lahir, dan alamat. Hasil ekstraksi dibandingkan dengan data yang dimasukkan calon nasabah dan dievaluasi berdasarkan aturan format serta ambang kepercayaan (confidence threshold) yang telah ditetapkan. Jika verifikasi otomatis memenuhi kriteria, data diteruskan ke tahap validasi akhir oleh admin; jika tidak, sistem memberikan umpan balik langsung kepada calon nasabah untuk memperbaiki data atau mengunggah ulang file KTP. Dengan alur ini, beban kerja admin berkurang, alur verifikasi menjadi lebih cepat, dan penilaian pada tahap awal menjadi lebih konsisten.



Gambar 3 Proses bisnis baru

3.3 Implementasi Sistem

Prototipe diimplementasikan sebagai aplikasi berbasis web menggunakan PHP 8 dengan framework Laravel, basis data MySQL, build tool Vite, serta JavaScript dengan pustaka jQuery dan Tesseract.js. Implementasi mencakup empat fungsi utama. Pertama, fungsi upload file dengan preview dan validasi yang memastikan file yang diunggah sesuai format gambar yang diizinkan dan berada dalam batas ukuran maksimal. Kedua, fungsi ekstraksi data KTP dengan OCR Tesseract.js yang memproses gambar secara asinkron dan menampilkan indikator progres secara real-time. Ketiga, fungsi parsing dan validasi yang mengolah teks mentah hasil OCR menjadi data terstruktur dengan mengidentifikasi pola dan kata kunci untuk mengekstraksi NIK, nama lengkap, tanggal lahir, dan alamat, disertai validasi awal seperti pemeriksaan jumlah digit NIK. Keempat, fungsi pencocokan data KTP dengan data formulir menggunakan algoritma perbandingan string yang hasilnya disajikan dalam bentuk tabel confidence berisi persentase kesamaan pada setiap bidang data.

3.4 Hasil Ekstraksi OCR

Gambar 4 menunjukkan contoh keluaran OCR Tesseract.js pada citra KTP, di mana sistem berhasil mendeteksi area teks utama yang ditandai dengan bounding box. Melalui proses ini terlihat bahwa OCR mampu mendeteksi teks utama pada KTP seperti NIK, nama lengkap, alamat, serta tempat dan tanggal lahir. Selain ditampilkan secara visual, hasil ekstraksi juga dianalisis secara kuantitatif melalui nilai confidence rata-rata dan durasi pemrosesan, sebagaimana contoh pengukuran pada salah satu dokumen yang disajikan pada Tabel 1.

Gambar 4 Keluaran OCR Tesseract.js pada KTP

Tabel 1 Nilai Confidence dan Durasi OCR

Confidence OCR	Durasi OCR
39,00%	7,50 detik

Nilai confidence pada Tabel 1 masih berada pada kisaran menengah, sehingga tahap pasca-pemrosesan berupa normalisasi teks, parsing berbasis aturan, dan validasi logika diperlukan untuk meningkatkan keandalan hasil ekstraksi sebelum data digunakan pada proses verifikasi.

3.5 Hasil Uji Coba Ahli

Pengujian prototipe dilakukan oleh dua orang ahli sistem di bidang teknik informatika menggunakan pengujian black box terhadap empat skenario. Hasil pengujian disajikan pada Tabel 2, di mana skor 1 menunjukkan hasil pengujian sesuai dengan yang diharapkan.

Tabel 2 Hasil Uji Coba Ahli

No	Skenario Pengujian	Hasil yang Diharapkan	Ahli 1	Ahli 2
1	Pengujian fungsi input data KTP	Sistem dapat menerima input data KTP	1	1
2	Pengujian fungsi output hasil ekstraksi	Sistem dapat menampilkan hasil ekstraksi data KTP	1	1
3	Pengujian kecepatan proses verifikasi	Sistem dapat memproses verifikasi KTP dalam waktu < 1 menit	1	1
4	Pengujian kemudahan penggunaan	Sistem dapat digunakan dengan mudah	1	1

Seluruh skenario dinyatakan berhasil oleh kedua ahli, sehingga jumlah skor yang diperoleh adalah 8 dari skor maksimal 8. Berdasarkan persamaan (2), persentase kelayakan sistem adalah $8/8 \times 100\% = 100\%$, sehingga sistem dikategorikan sangat layak digunakan dari sisi fungsionalitas.

3.6 Hasil Uji Coba Pengguna

Pengujian pengguna dilakukan terhadap enam responden pegawai asuransi dengan lima pernyataan kuesioner. Total skor observasi dari seluruh responden sebesar 147 dari skor maksimal 150, sehingga berdasarkan persamaan (3) diperoleh persentase kelayakan keseluruhan sebesar 98%. Rincian skor per kategori disajikan pada Tabel 3.

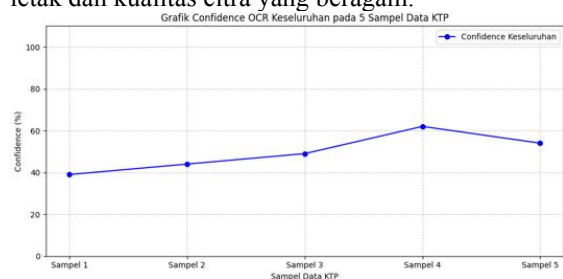
Tabel 3 Rincian Hasil Uji Coba Pengguna per Kategori

Kategori	Jumlah Pertanyaan	Skor Maksimal	Skor Observasi	Persentase Kelayakan (%)
Overall	5	150	147	98,00
System Usability	2	60	58	96,67
Performance & Responsiveness	1	60	56	93,33
Interface Quality	2	30	29	96,67

Setiap kategori menunjukkan persentase kelayakan di atas 90%, dengan nilai tertinggi pada kategori keseluruhan (overall) sebesar 98,00%. Berdasarkan kategori kelayakan, aplikasi verifikasi KTP ini termasuk sangat layak digunakan serta telah memenuhi ekspektasi pengguna dari segi kemudahan penggunaan, kecepatan respons, dan kualitas tampilan antarmuka.

3.7 Uji Hasil dan Pembahasan

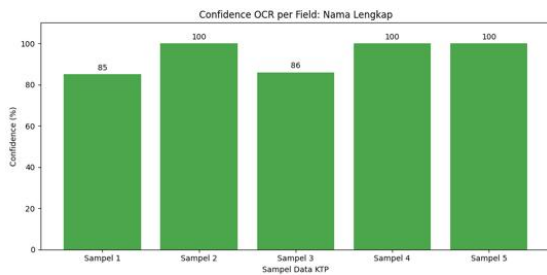
Uji hasil dilakukan dengan mengukur akurasi dan waktu proses ekstraksi serta pencocokan teks pada lima sampel data KTP. Hasil pengujian menunjukkan tingkat kepercayaan (confidence) OCR dalam mengenali teks masih bervariasi pada rentang 39% hingga 62%, sebagaimana ditunjukkan pada Gambar 5. Pada tingkat terendah (39%) masih banyak ketidakakuratan dalam pengenalan karakter, sedangkan nilai tertinggi (62%) menunjukkan performa lebih baik namun tetap menyisakan ruang perbaikan. Temuan ini mengindikasikan bahwa kinerja OCR dipengaruhi oleh kualitas gambar, jenis font, dan kondisi fisik KTP, sejalan dengan temuan penelitian sebelumnya bahwa kualitas citra merupakan faktor penentu keberhasilan OCR [3]. Variasi confidence yang ditemukan juga konsisten dengan evaluasi Tesseract-OCR pada jenis dokumen lain, seperti ekstraksi data pada dokumen C1 [15] dan pengujian pada dataset dokumen berbahasa Thai [16], yang sama-sama melaporkan bahwa tanpa penanganan khusus, akurasi Tesseract cenderung tidak stabil pada dokumen dengan tata letak dan kualitas citra yang beragam.



Gambar 5 Grafik confidence OCR per dokumen KTP

Analisis per bidang data menunjukkan variasi performa yang cukup mencolok. Pada bidang nama lengkap, tingkat confidence berada pada kisaran 85% hingga 100% dengan status cocok pada sebagian besar sampel, sebagaimana ditunjukkan pada Gambar 6, meskipun masih terdapat kesalahan minor seperti spasi berlebih atau pengulangan karakter. Bidang tanggal lahir

menunjukkan performa sangat baik dengan confidence 100% pada seluruh sampel, dan bidang jenis kelamin sebagian besar terbaca dengan confidence 100%. Sebaliknya, Nomor KTP (NIK) menjadi tantangan terbesar karena seluruh sampel terbaca tidak cocok, yang kemungkinan disebabkan oleh karakter angka yang relatif kecil serta adanya kombinasi simbol seperti titik dan strip yang menyulitkan sistem mengenali digit dengan tepat.



Gambar 6 Grafik confidence bidang nama lengkap

Dari sisi efisiensi, proses ekstraksi dan pencocokan data KTP menggunakan OCR memerlukan waktu rata-rata 2–3 detik per dokumen, jauh lebih cepat dibandingkan metode manual yang membutuhkan sekitar 15–30 detik per dokumen, sehingga hipotesis H2 (penerapan OCR berpengaruh positif terhadap peningkatan efisiensi waktu verifikasi) didukung oleh data yang diperoleh. Perlu ditegaskan bahwa angka akurasi lebih dari 90% pada kondisi citra ideal dihitung khusus dari bidang nama dan tanggal lahir yang seluruh sampelnya terbaca sesuai ground truth, dan tidak mencakup bidang NIK yang justru tidak cocok pada seluruh sampel yang diuji. Dengan demikian, hipotesis H1 (penerapan OCR berpengaruh positif terhadap peningkatan akurasi hasil verifikasi identitas) hanya didukung sebagian (partial support): OCR terbukti akurat untuk field nama, jenis kelamin, dan tanggal lahir, tetapi belum dapat diandalkan untuk field NIK yang merupakan identitas kunci dalam proses verifikasi. Pengujian hipotesis pada penelitian ini juga masih bersifat deskriptif berdasarkan persentase akurasi dan belum menggunakan uji statistik formal (misalnya uji beda rata-rata sebelum dan sesudah penerapan OCR), sehingga kesimpulan di atas perlu dikonfirmasi lebih lanjut melalui pengujian statistik pada sampel yang lebih besar. Fokus perbaikan selanjutnya terletak pada pembacaan NIK yang bersifat krusial, antara lain melalui peningkatan kualitas preprocessing citra dan penerapan teknik pemrosesan khusus untuk karakter numerik.

4. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, dapat ditarik kesimpulan sebagai berikut:

1. Sistem verifikasi identitas berbasis Optical Character Recognition (OCR) berhasil diimplementasikan pada aplikasi asuransi. Akurasi pengenalan pada kondisi citra ideal mencapai lebih dari 90% untuk bidang nama, jenis kelamin, dan tanggal lahir, sehingga OCR terbukti efektif membantu proses verifikasi pada bidang-bidang tersebut. Sebaliknya, bidang Nomor Induk Kependudukan (NIK) belum dapat diandalkan karena seluruh sampel yang diuji terbaca tidak cocok dengan data sebenarnya, sehingga OCR belum sepenuhnya meningkatkan ketepatan verifikasi identitas nasabah secara menyeluruh.
2. Proses ekstraksi dan pencocokan data KTP menggunakan OCR memerlukan waktu rata-rata 2–3 detik per dokumen, jauh lebih cepat dibandingkan metode manual (15–30 detik), sehingga teknologi OCR terbukti meningkatkan efisiensi waktu verifikasi secara signifikan.
3. Penerapan teknik preprocessing gambar seperti binarization dan peningkatan kontras terbukti membantu meningkatkan kualitas citra masukan serta akurasi hasil pembacaan OCR.
4. Penambahan mekanisme validasi hasil ekstraksi mampu mengurangi kesalahan input dan meningkatkan keandalan sistem dalam proses verifikasi identitas. Dengan hasil ini, hipotesis H2 mengenai peningkatan efisiensi waktu verifikasi didukung penuh oleh data, sedangkan hipotesis H1 mengenai peningkatan akurasi verifikasi didukung sebagian, yaitu pada bidang nama dan tanggal lahir, namun belum pada bidang NIK.
5. Penelitian ini memiliki keterbatasan berupa jumlah sampel pengujian yang kecil (lima citra KTP untuk pengujian akurasi dan enam responden untuk pengujian pengguna) serta belum digunakannya pengujian statistik formal, sehingga hasil yang diperoleh bersifat awal (preliminary) dan belum dapat digeneralisasi. Penelitian selanjutnya disarankan memperluas dataset dan jumlah responden, menerapkan pengujian statistik yang lebih ketat, serta mengembangkan teknik preprocessing dan pengenalan karakter khusus untuk meningkatkan akurasi pembacaan NIK.

Ucapan Terima kasih

Penulis mengucapkan terima kasih kepada Universitas Binaniaga Indonesia serta Divisi IT PT Tugu Insurance Brokers yang telah memberikan

izin, fasilitas, serta dukungan teknis dan informasi selama penelitian ini berlangsung.

Daftar Pustaka

- [1] Hais, Y. R., 2021, Implementasi Sistem Ekstraksi dan Validasi Data E-KTP Sebagai Solusi Alternatif Otomatisasi Sistem Administrasi Data untuk Organisasi Kecil Non-Pemerintah, *Prosiding SNAST*, hal 46–54.
- [2] Rusli, F. M., Adhiguna, K. A., dan Irawan, H., 2021, Indonesian ID Card Extractor Using Optical Character Recognition and Natural Language Post-Processing, *Proceeding of 9th International Conference on Information and Communication Technology (ICoICT)*, hal 621–626.
- [3] Zhang, X., Bai, W., dan Cui, H., 2023, Review of Optical Character Recognition for Power System Image Based on Artificial Intelligence Algorithm, *Energy Engineering*, No. 3, Vol. 120, hal 665–679.
- [4] Robby, G. A., Tandra, A., Susanto, I., Harefa, J., dan Chowanda, A., 2019, Implementation of Optical Character Recognition Using Tesseract with the Javanese Script Target in Android Application, *Procedia Computer Science*, Vol. 157, hal 499–505.
- [5] Van Hoai, D. P., Duong, H.-T., dan Hoang, V. T., 2021, Text Recognition for Vietnamese Identity Card Based on Deep Features Network, *International Journal on Document Analysis and Recognition (IJDAR)*, No. 1, Vol. 24, hal 123–131.
- [6] Fekete, D., dan Bárkányi, P., 2023, Examination of Technologies That Can Be Used for the Development of an Identity Verification Application, *International Journal of Advanced Natural Sciences and Engineering Researches*, No. 5, Vol. 7, hal 25–32.
- [7] Alghyaline, S., 2022, Arabic Optical Character Recognition: A Review, *CMES – Computer Modeling in Engineering and Sciences*, No. 3, Vol. 135, hal 1825–1861.
- [8] James, J. K., Maran, T., Rice, M. P., Hunt, T. S., Peterson, K. J., Hogan, W. J., Damani, S., dan Ryu, A. J., 2024, Experience With an Optical Character Recognition Search Application for Review of Outside Medical Records, *Mayo Clinic Proceedings: Digital Health*, No. 4, Vol. 2, hal 511–514.
- [9] Ponnuru, M., P, S. P., A, L., B, T. S., dan G, G. C., 2024, Image-Based Extraction of Prescription Information Using OCR-Tesseract, *Procedia Computer Science*, Vol. 235, hal 1077–1086.
- [10] Pradeep, J., Muthukumaran, A., Nareshraman, G., Krishnan, R., dan SakthyRosseouPrasanna, B., 2020, Efficient Vehicle Detection System Using OCR and REST API, *International Journal of Engineering Research & Technology (IJERT)*, No. 7, Vol. 9, hal 821–824.
- [11] Wahyuningrum, T., 2023, *Buku Referensi Kuesioner dalam Pengukuran Usability*, Deepublish, Yogyakarta.
- [12] Ramadhan, G., Mulyana, D. I., dan Adrianto, S., 2025, Optimization of Tesseract OCR for Automatic Text Extraction on Indonesian ID Cards (KTP) Through Image Quality Enhancement Using Preprocessing Techniques, *International Journal Software Engineering and Computer Science (IJSECS)*, No. 3, Vol. 5, hal 938–948.
- [13] Nisha, K., Wahyuni, T., dan Hayat, M. A. M., 2024, Pemeriksaan KTP Menggunakan Optical Character Recognition (OCR) dan Pengenalan Background serta Komponen KTP, *Arus Jurnal Sains dan Teknologi*, No. 2, Vol. 2, hal 490–495.
- [14] Putra, N. R., Rachmadi, R. F., dan Nugroho, S. M. S., 2023, Web Service untuk Ekstraksi Informasi KTP Menggunakan Google Cloud Vision, *Jurnal Teknik ITS*, No. 3, Vol. 12, hal A218–A223.
- [15] Nugroho, I. A., dkk., 2024, The Design of a C1 Document Data Extraction Application Using Tesseract-OCR Engine, *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, No. 1, Vol. 8, hal 152–159.
- [16] Anakpluek, T., dkk., 2025, Improved Tesseract OCR Performance on Thai Document Datasets, *Big Data Research*, Vol. 39, artikel 100508.