



Analisis Perbandingan *Optical Character Recognition* Google Vision dengan Microsoft Computer Vision pada Pembacaan KTP-el

Jonathan Valentino ¹, Yeremia Alfa Susetyo ^{2*}

¹ Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Kristen Satya Wacana, Kota Salatiga, Provinsi Jawa Tengah, Indonesia.

^{2*} Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Kristen Satya Wacana, Kota Salatiga, Provinsi Jawa Tengah, Indonesia.

article info

Article history:

Received 16 March 2023

Received in revised form

16 May 2023

Accepted 20 August 2023

Available online October 2023

DOI:

<https://doi.org/10.35870/jtik.v7i4.1046>

Keywords:

OCR; Identity Card; Google Vision API; Microsoft Computer Vision.

Kata Kunci:

OCR; Kartu Tanda Penduduk Elektronik; Google Vision API; Microsoft Computer Vision.

abstract

In this era, the need of digital data is rapidly increasing. Electronic Residential Identity Card or KTP-el is the official identity card for resident of Indonesia. One fast way to extract information on an image is by using OCR/Optical Character Recognition. Competition between Google Vision API and Microsoft Computer Vision in providing OCR service encourage companies to choose the right provider. Method conducted in this research including literature review on both OCR service provider, identification and KTP-el sample image retrieval, data grouping, code implementation and accuracy testing, result analysis and discussion, and conclusion. The result of this research show that Microsoft Computer Vision have better accuracy in reading characters in KTP-el with an accuracy percentage of 0,81% to 15,8% difference to Google Vision. Google Vision has competitive accuracy, but suffers from deficiencies when reading KTP-el with blur and noise.

abstract

Pada era ini, kebutuhan akan data digital sangat meningkat. Kartu Tanda Penduduk Elektronik atau KTP-el adalah kartu identitas resmi penduduk Indonesia. Salah satu cara mendapatkan informasi secara cepat pada sebuah gambar adalah menggunakan OCR / Optical Character Recognition. Kompetisi antara Google Vision API dan Microsoft Computer Vision pada layanan OCR mendorong perusahaan untuk memilih penyedia layanan yang tepat. Tahapan penelitian yang dilakukan pada penelitian ini adalah studi pustaka terhadap dua penyedia layanan OCR, identifikasi dan pengambilan data gambar KTP-el sebagai sampel, pengelompokan data, implementasi kode dan pengujian akurasi, analisis hasil dan pembahasan, dan kesimpulan. Hasil dari penelitian ini menunjukkan bahwa Microsoft Computer Vision lebih unggul dalam membaca karakter dalam KTP-el dengan selisih persentase akurasi 0,81% hingga 15,8% lebih tinggi dari Google Vision. Google Vision memiliki akurasi yang bersaing, namun memiliki kekurangan saat membaca KTP-el dengan blur dan noise.

Corresponding Author. Email: yeremiaalfasusetyo@uksw.edu ^{2}.

1. Latar Belakang

Pada era digital, teknologi adalah bagian penting dalam kehidupan manusia. Teknologi telah menjadi solusi dalam membantu manusia menjalankan aktivitas dan pekerjaannya. Perkembangan teknologi semakin mendorong pelaku bisnis untuk mendigitalisasikan proses bisnis mereka. Dalam tahap transformasi tersebut, digitalisasi data fisik seperti gambar, dokumen, dan buku adalah hal yang esensial untuk dilakukan. Dokumen digital dapat digunakan oleh perusahaan untuk menyederhanakan proses dan meningkatkan kualitas layanan dari perusahaan [1]. Pada banyak perusahaan, berbagai proses bisnis membutuhkan kartu identitas sebagai alat verifikasi [2].

Pemerintah Indonesia telah memperkenalkan Kartu Tanda Penduduk Elektronik (KTP-el) sebagai kartu identitas resmi penduduk Indonesia. Pertukaran data diri untuk keperluan administrasi di Indonesia masih banyak dilakukan secara manual, salah satu penyebabnya adalah kurangnya sosialisasi akan metode ekstraksi KTP-el [3]. Salah satu metode untuk melakukan ekstraksi pada sebuah gambar adalah *Optical Character Recognition* (OCR).

Computer Vision digunakan sebagai alat pemrosesan dokumen digital, khususnya dalam melakukan ekstraksi data pada teks digital. *Optical Character Recognition* (OCR) adalah teknik yang dipakai untuk mengekstrak karakter dalam dokumen digital. Dewasa ini semakin banyak pilihan alat OCR, berbayar maupun gratis dan mendukung banyak bahasa. Sehubungan dengan itu, semakin banyak penyedia layanan OCR yang bersaing dalam mengembangkan API dengan kualitas pemindaian yang baik seperti Google Vision API, Tesseract [4], hingga Microsoft Computer Vision.

Setiap penyedia layanan memiliki tawaran biaya, limitasi, dukungan bahasa pemrograman, dan fitur yang berbeda di dalamnya. Layanan yang paling menguntungkan dan sejalan dengan proses bisnis perusahaan menjadi nilai penting bagi perusahaan dalam memilih penyedia layanan yang tepat. Untuk dapat mengetahui layanan terbaik yang menyediakan OCR, perusahaan membutuhkan perbandingan akan dasar fitur, pembiayaan, integrasi teknologi, dan hasil dari deteksi OCR pada pembacaan KTP-el.

Google Cloud Vision adalah layanan untuk mengekstraksi informasi yang melekat pada sebuah citra melalui REST API. Google Cloud Vision digunakan pada bidang komputer vision seperti pemindaian teks (*text recognition*) [6]. Google Cloud Vision API menggunakan model pembelajaran pra-terlatih untuk mendeteksi object, teks, dan metadata pada gambar.

Microsoft Cognitive Service adalah seperangkat API yang memungkinkan pengembang untuk menambahkan fitur cerdas seperti deteksi video, pengenalan wajah, bicara, penglihatan dan teks ke dalam aplikasi. Terdapat enam kategori API dalam Azure Cognitive Service diantaranya, Computer Vision, Speech, Language, Knowledge, Search and Labs [7].

Kompetisi antar layanan yang menyediakan alat OCR mendorong banyak perusahaan untuk memilih layanan yang tepat dan sesuai harapan perusahaan. Adanya teknologi Google Cloud Vision dan Microsoft Azure Computer Vision memberikan pilihan solusi untuk melakukan ekstraksi data dalam gambar khususnya KTP el. Penting bagi perusahaan untuk mengetahui hasil komparasi dari kedua layanan tersebut.

Pada penelitian ini akan dilakukan analisis perbandingan OCR Google Vision dengan Microsoft Computer Vision dengan memperhatikan perbedaan fitur, dan hasil akurasi pada pembacaan KTP-el pada berbagai kondisi. Dilakukan pengujian terhadap sampel KTP-el pada berbagai resolusi, dengan peningkatan kualitas citra, penurunan kualitas citra, dan orientasi citra. Tahapan metode yang dilakukan penulis meliputi studi literatur, pengumpulan data sampel KTP-el, pengelompokkan sampel KTP-el, pengambilan data pengujian OCR, analisis hasil pengujian dan pembahasan, dan kesimpulan.

Berdasarkan permasalahan di atas, maka rumusan masalah dari penelitian ini adalah bagaimana hasil analisis perbandingan antara Google Vision dan Microsoft Computer Vision dalam mendeteksi karakter dalam KTP-el, dan bagaimana perbedaan fitur dari layanan OCR model pra terlatih yang ditawarkan Google Vision dan Microsoft Computer Vision. Tujuan dari penelitian ini adalah menganalisis perbandingan antara layanan Google Vision dengan

Microsoft Computer Vision dalam melakukan OCR pemindaian gambar pada citra KTP-el. Kontribusi dari artikel penelitian ini mencakup: 1) memberikan acuan dasar pertimbangan yang objektif pada kedua layanan bagi perusahaan dan pembaca, dan 2) memberikan informasi layanan OCR yang lebih unggul dari kedua platform kepada pembaca. iii) penelitian ini menampilkan kemampuan OCR dengan model pra-terlatih menggunakan API dalam mendeteksi kartu identitas kepada pembaca.

Pada penelitian yang berjudul “Ekstraksi Informasi/Data e-KTP Menggunakan *Optical Character Recognition* Convolutional Neural Network” yang dilakukan oleh Gunawan Sugiarta, Dianthika Puteri Andini, dan Syarief Hidayatullah menyimpulkan bahwa penggunaan metode OCR dan empat layer CNN (Convolutional Neural Network) berhasil mengekstraksi data yang terdapat pada 35 sampel e-KTP dengan rata-rata kecepatan 30 detik dan *error rate* 5%. Disimpulkan juga bahwa citra e-KTP berpengaruh terhadap hasil ekstraksi. Penelitian tersebut membahas tentang ekstraksi informasi / data dengan metode OCR Convolutional Neural Network. Metode tersebut diimplementasikan ke dalam bentuk aplikasi untuk secara otomatis menginputkan data e-KTP yang sudah dibaca oleh OCR [7].

Pada penelitian yang berjudul “Comparative analysis of Tesseract and Google Cloud Vision for Thai vehicle registration certificate” oleh Karanrat Thammarak, Prateep Kongkla, Yaowarat Sirisathitkul, dan Sarun Intakosum menyimpulkan bahwa Google Cloud Vision lebih akurat dengan tingkat akurasi 84.43% daripada Tesseract dengan tingkat akurasi 47.02%. Penelitian tersebut menganalisis perbandingan antara Google Cloud Vision dan Tesseract untuk mendeteksi sertifikat tanda nomor kendaraan negara Thailand. Dilakukan tinjauan pustaka yang lengkap meliputi perbandingan fitur dasar pada kedua alat OCR, serta tabel artikel yang mengimplementasikan OCR pada bahasa Thailand. Metode yang digunakan pada penelitian tersebut adalah studi literatur terhadap alat dan teknik untuk meningkatkan kualitas proses pemindaian, serta pengujian terhadap 84 sampel sertifikat nomor kendaraan dengan berbagai kondisi. Hasil eksperimen dari penelitian tersebut sangat lengkap, meliputi akurasi pada ukuran gambar berbeda, pada

gambar yang mengalami kecacatan, gambar yang di tingkatkan, dan aksara Thailand yang salah diinterpretasikan oleh OCR [4].

Pada penelitian berjudul “Examining Robustness of Google Vision API Based on the Performance on Noisy Image,” oleh Akshat Pathak, Aviral Ruhela, Anshul K. Saroha, dan Anant Bhardwaj menghasilkan kesimpulan bahwa Google Cloud Vision tidak *robust* terhadap gambar dengan *noise*. Dengan ditambahkannya gaussian noise dan impulsive noise pada gambar yang dideteksi, Google Cloud Vision gagal melakukan fungsi deteksi teks dan deteksi label secara maksimal, hingga ke titik dimana teknologi ini tidak dapat mendeteksi teks secara efektif lagi. Penelitian tersebut melakukan robustness testing terhadap Google Cloud Vision dalam mendeteksi gambar yang memiliki *noise*. Pengujian dilakukan dalam 5 kategori gambar, dengan total 75 sampel gambar. Penelitian ini menghasilkan kesimpulan bahwa Google Cloud Vision tidak robust terhadap gambar ber-noise [8].

Google Vision API adalah layanan berbasis cloud yang memiliki salah satu fungsi utama yaitu OCR (*Optical Character Recognition*). OCR berfungsi untuk mengekstraksi teks dalam sebuah citra. Google Vision API menyediakan alat analisis berupa deteksi label, *landmark detection*, *web detection*, *face detection*, *content moderation*, *ML kit integration*, *handwriting detection*, *integrated REST API*, dan OCR. Pada OCR, Google menyediakan fitur ekstraksi teks yang dibedakan menjadi dua yaitu *Text Detection* dan *Document Text Detection*. Perbedaan dari kedua fitur OCR tersebut adalah *Document Text Detection* memiliki *response* yang lebih optimal untuk mendeteksi dense text, halaman, blok, paragraf, kata, dan jeda yang dikembalikan sebagai output. Kedua fitur diatas akan mendeteksi teks dalam sebuah gambar, dan menghasilkan output dalam bentuk JSON berisi hasil ekstraksi teks baik dalam bentuk kalimat lengkap, hingga kata-kata individual didalamnya [9].

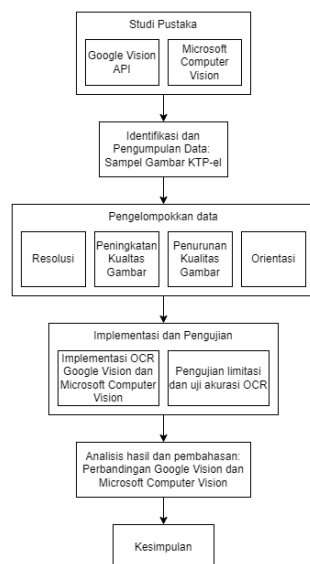
Microsoft Cognitive Service adalah seperangkat APIs, SDKs, dan layanan lainnya yang tersedia untuk membantu developer membangun aplikasi cerdas tanpa membutuhkan pengetahuan mendalam mengenai AI (Artificial Intelligence) dan ML (Machine Learning). Tujuan dari Azure Cognitive Service adalah meningkatkan komputasi dan

produktivitas dengan bantuan sistem cerdas yang ditawarkan [10]. Microsoft Computer Vision menawarkan layanan *image classification*, *content moderation*, OCR, *celebrity recognition*, dan *landmark recognition*. Layanan OCR API pada Microsoft Computer Vision dapat membaca teks dalam gambar dan mengembalikan karakter dalam gambar kepada pengguna, serta dapat digunakan untuk tulisan tangan [11]. Penelitian ini memanfaatkan layanan *text extraction* (OCR) yang merupakan salah satu fungsi utama dari kelompok Computer Vision pada Microsoft Cognitive Service.

Kartu Tanda Penduduk Elektronik (KTP-el) adalah dokumen kependudukan berdasarkan basis data nasional yang memuat sistem keamanan dan pengendalian dari sisi administrasi dan teknologi informasi. Struktur KTP-el terdiri dari sembilan lapisan yang membedakannya dengan KTP konvensional. Pada KTP-el terdapat informasi penting diantaranya NIK (Nomor Induk Kependudukan), nama lengkap, tempat/tanggal lahir, jenis kelamin, agama, status perkawinan, golongan darah, alamat lengkap, pekerjaan, kewarganegaraan, pas foto, tempat dan tanggal dikeluarkannya KTP, dan tanda tangan pemegang KTP [12].

2. Metode Penelitian

Tahapan penelitian yang dilakukan adalah sebagai berikut:



Gambar 1. Tahapan Penelitian

Tahap pertama pada penelitian ini adalah studi pustaka, data yang membantu penelitian diambil dari studi kepustakaan meliputi referensi jurnal nasional, jurnal internasional, artikel, dan situs internet yang berhubungan dengan permasalahan penelitian. Tahap studi pustaka meliputi membaca, mencatat, dan mengolah data yang dibutuhkan untuk menyajikan perbedaan fitur dari Google Vision API dan Microsoft Computer Vision.

Tahap kedua adalah identifikasi dan pengumpulan data. Identifikasi berkaitan dengan penentuan lingkup pembahasan yang akan dikaji pada penelitian ini. Lingkup pembahasan difokuskan pada analisis perbandingan fitur dan akurasi dari kedua layanan. Pengumpulan data dilakukan untuk mendapatkan sampel gambar KTP-el sebanyak 10, yang kemudian dilakukan *image processing* pada tiap sampel hingga didapatkan 120 sampel yang dibagi menjadi beberapa kategori pada tahap selanjutnya.

Tahap ketiga adalah pengelompokan data sampel KTP-el. Pada penelitian ini digunakan 120 sampel gambar KTP-el yang dikategorisasikan berdasarkan Tabel 1. Pada penelitian ini sampel gambar KTP-el dibagi menjadi 4 kategori, yaitu: (i) resolusi, (ii) peningkatan kualitas citra, (iii) penurunan kualitas citra, dan (iv) orientasi citra. Pada masing-masing kategori terdapat sub-kategori yang disajikan dalam tabel 1.

Tabel 1. Kategori pengujian sampel KTP-el

Kategori	Sub-kategori	Jumlah sampel
Resolusi	Resolusi Tinggi (1080 x 720) 300 dpi	10
	Resolusi Sedang (720 x 480) 150 dpi	10
	Resolusi Rendah (360 x 240) 72 dpi	10
Peningkatan Kualitas Citra	Peningkatan Kecerahan (+50 brightness)	10
	Peningkatan Kontras(+50 contrast)	10
	Penajaman (+100 sharpness)	10
	Grayscale	10

Penurunan Kualitas Citra	Blur (Gaussian 3 radius)	10
	Pengurangan Kecerahan (-50 brightness)	10
	Noise (Gaussian 100 intensity)	10
Orientasi Citra	Citra dirotasi	10
	Citra dengan perspektif tidak persegi	10

Tahap keempat adalah implementasi dan pengujian terhadap sampel KTP-el. Dari hasil pengujian OCR menggunakan sampel KTP-el, dihitung nilai akurasi dari masing – masing sampel pada setiap sub-kategori. Untuk menghitung jumlah karakter sesuai dari setiap sampel, digunakan persamaan 1. Setelah mendapatkan jumlah karakter sesuai, persamaan 1 digunakan untuk mendapatkan persentase akurasi pada persamaan 2.

$$\text{Karakter Sesuai} = \sum \text{Karakter yang diuji} - \sum \text{Karakter Salah} \quad (1)$$

$$\text{Akurasi} = \frac{\text{Karakter Sesuai}}{\sum \text{Karakter yang diuji}} \times 100\% \quad (2)$$

Tahap kelima adalah analisis hasil dan pembahasan. Hasil uji akurasi pada tahapan sebelumnya digunakan sebagai bahan analisis guna menemukan kelemahan dan keunggulan dari masing-masing layanan. Selain menyajikan data akurasi setiap sub-kategori, dari hasil studi literatur juga dijabarkan mengenai perbedaan fitur dan kapabilitas yang berisi: perbedaan fitur deteksi teks / OCR dan perbedaan batasan input gambar dari OCR. Tahap terakhir setelah analisis hasil dan pembahasan adalah menyimpulkan penelitian yang telah dilakukan, berdasarkan hasil analisis perbandingan Google Vision API dan Microsoft Computer Vision.

3. Hasil dan Pembahasan

Hasil dari penelitian ini adalah perbandingan antara Google Vision API dan Microsoft Computer Vision. Dari segi fitur, perbandingan antara kedua penyedia layanan OCR disajikan pada Tabel 2.

Tabel 2 Komparasi fitur deteksi teks

Fitur	Google Vision API	Microsoft Computer Vision
Dukungan bahasa (OCR <i>Printed text</i>)	133 dukungan bahasa (teks cetak yang dipetakan)	164 dukungan bahasa (teks cetak)
Bentuk tulisan	Teks cetak, tulisan tangan	Teks cetak, tulisan tangan
Mendukung SDK	Ya	Ya
Biaya	Gratis terbatas (1000/bulan), sesuai pemakaian	Gratis terbatas (5000/bulan), sesuai pemakaian
Integrasi dengan teknologi cloud	Ya	Ya
Dukungan bahasa pemrograman	REST, Go, Java, Node.js, Python, C#, PHP, Ruby, Javascript, .NET	REST API, Java, Node.js, Python, C#, Ruby, Javascript, .NET
Tipe tagihan	Pembayaran sesuai pemakaian	Pembayaran sesuai pemakaian dan transaksi tetap per bulan

Dari segi fitur, Microsoft Computer Vision unggul dalam dukungan bahasa, pemakaian layanan gratis terbatas, dan variasi tipe tagihan yang dapat dipilih pengguna. Untuk pemakaian layanan gratis terbatas, kuantitas yang ditawarkan oleh Microsoft Azure lebih besar yaitu 5000 *transactions*, namun terdapat limitasi transaction yang hanya 20 *transaction/minute*. Google Vision dengan kuantitas yang hanya 1000 *request*, menawarkan limitasi kuota yang tetap yaitu 1800 *request/minute*. Microsoft juga memberikan variasi tipe tagihan yang tetap, disebut sebagai *Commitment Tiers*.

Limitasi ukuran gambar yang ditentukan oleh Google adalah 20 MB berdasarkan studi literatur, namun setelah dilakukan pengujian terhadap gambar dengan ukuran 93 MB, fungsi OCR yang dijalankan

memberikan keluaran ukuran gambar tidak dapat melebihi 41,943,040 Bytes atau 40 MB. Sedangkan berdasarkan studi pustaka, Microsoft Computer Vision memiliki batasan ukuran gambar maksimal 500 MB dan dalam penelitian ini berhasil dilakukan pengujian terhadap gambar dengan ukuran 95,5 MB.

Hasil pengujian menunjukkan bahwa dimensi gambar maksimal dari Google Vision API adalah 75 megapixels atau 75.000.000 pixels, sedangkan untuk Microsoft Computer Vision batasan dimensi gambar adalah 100 megapixels atau setara dengan gambar berdimensi 10,000 pixels x 10,000 pixels. Limitasi format gambar yang didukung dari kedua layanan hampir sama, dibedakan dengan Google Vision API yang dapat melakukan OCR dengan format animated gif.

Kode Program 1. Kode Program Google Vision

```
client = vision.ImageAnnotatorClient()
image = vision.Image(content=content)
response = client.text_detection(image=image)
texts = response.text_annotations
```

Kode Program 1 adalah kode yang digunakan untuk mengirim request dan mendapatkan response pembacaan teks kepada Google Vision. Request yang dikirim berbentuk JSON. Pada baris 1, variabel client diisi dengan melakukan instantiasi / pembuatan object ImageAnnotatorClient dari library vision. Baris 2 adalah inisialisasi variabel image yang berisi pemanggilan method dari library vision, dengan parameter content yang berisi gambar yang disimpan dalam memori. Variabel response adalah balikan dari method text_detection yang dimiliki oleh object client pada baris 1, dengan parameter image yang diinisialisasi pada baris 2. Response yang didapatkan berbentuk JSON, yang memiliki key text_annotations dengan value hasil pembacaan oleh Google Vision.

Kode Program 2. Kode Program Microsoft Computer Vision

```
computervision_client =
ComputerVisionClient(endpoint,
CognitiveServicesCredentials(key))
read_response =
computervision_client.read_in_stream(read_image, raw=True)
read_operation_location = read_response.headers["Operation-
```

```
Location"]
operation_id = read_operation_location.split("/")[-1]
read_result =
computervision_client.get_read_result(operation_id)
text_result = read_result.analyze_result.read_results
```

Kode Program 2 adalah kode yang digunakan untuk mengirim request dan mendapatkan response dari layanan Microsoft Computer Vision. Baris kode 1 dan 2 adalah inisialisasi library ComputerVisionClient dengan parameter endpoint dan kunci kredensial (url cognitive service dan kunci langganan yang telah dibuat dalam website). Baris kode 4 dan 5 digunakan untuk mendapatkan operation id yang terdapat pada url operation location. Baris kode 6 menjalankan method pada client untuk mendapatkan response keseluruhan. Selanjutnya diambil hanya hasil pembacaan teks, pada variabel text_result di baris kode ke 7.

Kode Program 1 dan 2 digunakan untuk melakukan pembacaan pada 120 sampel KTP-el. Sampel gambar KTP-el ditampilkan pada gambar 2 dan 3 dalam kategori penurunan kualitas gambar, sub-kategori *blur* dan *noise*.



(a) Sampel KTP-el *Blur* (b) Sampel KTP-el *Noise*
Gambar 2. Sampel KTP-el *Blur* dan *Noise*

Penggunaan kode program 1 dan 2 pada sampel gambar KTP-el menghasilkan kumpulan teks hasil pembacaan OCR. Baris-baris teks hasil pembacaan kemudian dibandingkan dengan karakter yang diuji untuk menentukan jumlah karakter yang tidak sesuai/salah. Jumlah karakter yang salah selanjutnya digunakan pada persamaan 1 untuk mendapatkan jumlah karakter sesuai.

Kategori	Jumlah karakter sesuai Google Vision/Microsoft Computer Vision									
	KTP 1	KTP 2	KTP 3	KTP 4	KTP 5	KTP 6	KTP 7	KTP 8	KTP 9	KTP 10
Resolusi tinggi	356/ 368	401/ 406	364/ 373	329/ 335	342/ 357	342/ 352	366/ 377	366/ 369	359/ 365	366/ 372
Resolusi sedang	351/ 361	398/ 405	370/ 376	376/ 375	349/ 359	345/ 353	370/ 374	366/ 378	358/ 362	370/ 376
Resolusi rendah	347/ 349	396/ 405	360/ 375	320/ 333	348/ 350	339/ 351	366/ 372	363/ 372	354/ 360	363/ 377
Peningkatan kecerahan	351/ 359	401/ 406	376/ 379	325/ 328	354/ 360	350/ 354	376/ 364	372/ 373	358/ 364	368/ 374
Peningkatan kontras	356/ 368	387/ 406	371/ 366	320/ 327	354/ 359	346/ 354	377/ 366	365/ 374	357/ 358	369/ 374
Penajaman	350/ 362	392/ 404	356/ 345	337/ 337	334/ 337	333/ 338	369/ 368	365/ 366	366/ 353	365/ 372
Grayscale	350/ 363	401/ 405	371/ 374	327/ 328	347/ 359	345/ 356	368/ 381	367/ 373	356/ 366	366/ 374
Blur	338/ 336	379/ 398	363/ 363	303/ 302	343/ 345	324/ 331	212/ 310	283/ 326	301/ 336	305/ 341
Pengurangan kecerahan	342/ 359	399/ 398	357/ 371	301/ 336	337/ 355	336/ 348	360/ 368	361/ 368	344/ 362	324/ 341
Noise	344/ 341	142/ 370	350/ 356	292/ 307	336/ 344	334/ 339	81/ 260	330/ 339	195/ 333	302/ 339
Citra dirotasi	348/ 360	401/ 403	372/ 372	268/ 335	350/ 358	350/ 349	372/ 377	285/ 371	361/ 363	369/ 375
Perspektif buruk	356/ 364	402/ 406	374/ 375	325/ 340	350/ 360	347/ 355	370/ 379	379/ 369	361/ 363	371/ 380

Gambar 3. Jumlah karakter sesuai setiap kategori

Gambar 3 adalah hasil perhitungan jumlah karakter sesuai menggunakan persamaan 1 pada Google Vision/Microsoft Computer Vision. Angka yang dicetak tebal menunjukkan jumlah karakter sesuai yang lebih unggul dalam membaca KTP-el. Jumlah karakter sesuai selanjutnya digunakan dalam persamaan 2 untuk mendapatkan persentase akurasi masing-masing sampel KTP-el. Jumlah karakter yang diuji pada setiap KTP-el ditunjukkan pada Tabel 4.

Tabel 4. Jumlah karakter yang diuji

No	Sampel KTP-el	Jumlah karakter yang diuji
1.	KTP-el 1	371
2.	KTP-el 2	406
3.	KTP-el 3	388
4.	KTP-el 4	341
5.	KTP-el 5	362
6.	KTP-el 6	361
7.	KTP-el 7	394
8.	KTP-el 8	397
9.	KTP-el 9	378
10.	KTP-el 10	384

Setelah didapatkan persentase akurasi dari setiap sampel KTP-el, dihitung rata-rata kategori dengan membagi total persentase akurasi dari setiap kategori. Rata – rata persentase akurasi tersebut disajikan dalam diagram 1.

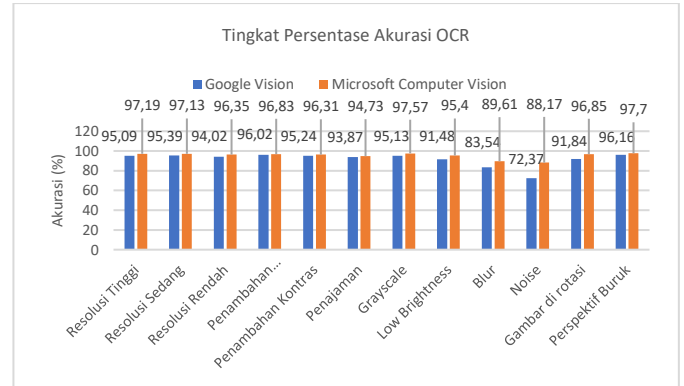


Diagram 1. Tingkat Persentase Akurasi OCR

Hasil pengujian 120 sampel KTP-el, yang dibagi menjadi empat kategori yaitu resolusi, peningkatan kualitas gambar, penurunan kualitas gambar, dan orientasi gambar, menunjukkan bahwa Microsoft Computer Vision memiliki akurasi yang lebih baik daripada Google Vision API. Pada kategori resolusi, Google Vision API memiliki rata-rata akurasi yang lebih rendah dua persen daripada Microsoft Computer Vision. Hasil uji akurasi menunjukkan bahwa KTP-el dengan resolusi rendah atau berdimensi 360 pixel x 240 pixel masih menunjukkan tingkat akurasi yang cukup tinggi, mengalami penurunan akurasi satu persen dari resolusi tinggi dan sedang.

Dalam proses perhitungan akurasi, pada resolusi rendah ditemukan kesalahan pembacaan karakter seperti karakter Tgl yang berubah menjadi Tol, karakter RT salah dikenali menjadi FIT, karakter e salah dikenali menjadi a, dan karakter U yang salah dikenali menjadi LI. Pada setiap kategori resolusi juga ditemukan pembacaan karakter titik dua dan spasi yang tidak konsisten, terutama pada penyedia layanan OCR Google Vision API.

Pada penelitian yang dilakukan oleh Angga Maulana Purba, Agus Harjoko, dan Moh Edi Wibowo, dilakukan deteksi teks pada KTP-el menggunakan algoritma MSER (Maximally Stable External Regions) dengan kombinasi horizontal RLSA. Penelitian tersebut menghasilkan kesimpulan akurasi terbaik

dengan metode tersebut sebesar 84,49 persen [13], 10,6 persen dan 12,7 persen lebih rendah dari hasil uji akurasi Google Vision dan Microsoft Computer Vision pada gambar dengan resolusi tinggi.

Penelitian lain yang dilakukan oleh Dewinta Zulhida Putri, Diyah Puspitaningrum, dan Yudi Setiawan menggunakan Tesseract Tess-two dan jaro-winkler spring pada kartu nama memberikan pernyataan tentang karakter-karakter yang tidak terdapat pada kartu nama yang ikut terekstraksi dengan OCR [14]. Hal ini sejalan dengan tantangan dalam mengekstraksi data pada KTP-el. KTP-el memiliki watermark yang berupa teks yang seringkali salah diinterpretasikan oleh OCR sebagai informasi yang akan diekstraksi. Kesalahan pembacaan watermark dapat diminimalisir dengan melakukan peningkatan kualitas gambar.

Hasil uji akurasi pada kategori peningkatan kualitas gambar menunjukkan bahwa penambahan kecerahan mempengaruhi kenaikan tingkat akurasi pada Google Vision. Selain penambahan kecerahan, konversi gambar ke dalam bentuk grayscale atau abu-abu juga menaikkan tingkat akurasi pada kedua penyedia layanan. Namun untuk penajaman gambar gagal untuk menaikkan tingkat akurasi kedua penyedia layanan. Hal ini terjadi karena penajaman gambar dapat membuat karakter dalam gambar menjadi menyatu. Penyatuan karakter ini disebabkan oleh teknik penajaman yang membuat tepi objek menjadi lebih tebal. Sehingga pada sampel KTP-el yang berbayang, teknik ini membuat beberapa karakter pada gambar dengan jarak yang sempit menjadi menyatu.

Kategori penurunan kualitas gambar berhasil menurunkan performa pengenalan teks OCR. Gambar KTP-el dengan kecerahan yang dikurangi sebanyak 50, menurunkan tingkat akurasi dari Google Vision API sebanyak 3,91 persen dari persentase akurasi maksimal. Pada Microsoft Computer Vision tingkat akurasi berkurang 1,79 persen. Sub kategori blur dan noise berhasil membuat tingkat akurasi turun hingga dibawah 90 persen. Dua penurunan kualitas gambar ini berhasil membuat beberapa sampel KTP-el tidak dikenali.

```
PROVINSILAMPUNG
KOTA BANDAR LAMPUNG
# #####
BRARIANALIGLOK GENOGEK
PELLIARAMAHASSISWA
AIR HIP
KOTA BANDAR
```

Karakter di atas adalah hasil uji sampel KTP-el yang diberikan Gaussian Noise. Dari 394 karakter pada KTP-el, Google Vision API hanya berhasil mengenali 81 karakter saja. Dengan ditambahkannya noise atau blur pada KTP-el, banyak karakter yang salah diinterpretasikan, hingga hilangnya beberapa blok-blok karakter yang seharusnya dikenali. Pada kategori ini, Microsoft Computer Vision unggul di ketiga sub kategori, dengan selisih persentase akurasi 3,92 persen untuk sub kategori low brightness, 6,07 persen untuk sub kategori blur, dan 15,8 persen untuk sub kategori noise.

Pada pengujian OCR kategori orientasi gambar, gambar yang dirotasi mempengaruhi akurasi OCR dari Google Vision API sebanyak 3,55 persen dan Microsoft Computer Vision 0,34 persen. Hal ini menandakan OCR engine yang dimiliki oleh kedua penyedia layanan dapat membaca gambar yang dirotasi. Pada hasil OCR pada gambar yang dirotasi 45°, 135°, dan 330°, Google Vision API memberikan urutan hasil pemindaian yang tidak konsisten dimana data hasil pemindaian dimulai dari blok pekerjaan atau berlaku hingga.

Karakteristik urutan pemindaian yang tidak konsisten pada berbagai sampel KTP-el menggunakan Google Vision API memiliki kesamaan dengan pernyataan pada hasil penelitian yang dilakukan oleh Meredita Susanty, dan Herminarto Nugroho saat mengimplementasikan OCR berbasis Tesseract pada sertifikat UTBK (Ujian Tertulis Berbasis Komputer). Tesseract OCR dapat mendeteksi gambar terbalik maupun gambar yang dirotasi 180°. Namun rotasi gambar dan dimensi gambar dapat menyebabkan perbedaan urutan kata hasil pemindaian [15]. Sebaliknya, Microsoft Computer Vision memberikan hasil pemindaian dengan urutan yang konsisten dan dengan akurasi yang optimal. Sub kategori perspektif buruk tidak menurunkan tingkat akurasi pada layanan OCR kedua platform.

Hasil uji dari Google Vision API dan Microsoft Computer Vision pada 120 sampel KTP-el dari 4 kategori menunjukkan bahwa Microsoft Computer Vision unggul dalam perbandingan uji akurasi dengan Google Vision API. Kekurangan Google Vision API pada pengujian ini terdapat pada ketahanannya dalam mendeteksi gambar dengan blur dan noise.

4. Kesimpulan dan Saran

Berdasarkan penelitian yang telah dilakukan, diambil beberapa kesimpulan. Dari segi fitur, Microsoft Computer Vision unggul dalam dukungan bahasa, penggunaan gratis terbatas, dan tipe tagihan *commitment tiers*. Google Vision API memiliki keunggulan pada variasi dukungan bahasa pemrograman. Pada limitasi Microsoft Computer Vision juga relatif unggul kecuali pada bagian request quota, dimana Google Vision API memiliki kuota 1800 request/minute.

Untuk tingkat akurasi pada pengenalan KTP-el, Microsoft Computer Vision memiliki tingkat akurasi yang lebih tinggi daripada Google Vision API dengan selisih 1 – 2 persen, dengan tambahan catatan bahwa degradasi kualitas gambar akan mengurangi tingkat akurasi secara signifikan. Tingkat selisih persentase akurasi 3,92 persen untuk sub kategori low brightness, 6,07 persen untuk sub kategori blur, dan 15,8 persen untuk sub kategori noise. Kedua penyedia layanan memiliki kemampuan yang mumpuni untuk melakukan OCR, terutama pada pengenalan citra KTP-el.

Masing – masing memiliki keunggulan dan kelemahan yang telah dijabarkan dan dapat menjadi pertimbangan bagi pengguna untuk memilih layanan yang tepat sesuai kebutuhan. Pada penelitian lebih lanjut dapat menggunakan hasil penelitian ini sebagai referensi dan bahan perbandingan untuk teknologi OCR lainnya dan dapat menambahkan kategori pengujian, variabel yang dibandingkan, dan jenis font yang lebih beragam.

5. Daftar Pustaka

- [1] Thammarak, K., Sirisathitkul, Y., Kongkla, P. and Intakosum, S., 2022. Automated Data Digitization System for Vehicle Registration Certificates Using Google Cloud Vision API. *Civil Engineering Journal*, 8(7), pp.1447-1458. DOI: <https://doi.org/10.28991/CEJ-2022-08-07-09>.
- [2] Guntara, R.G., 2022. Aplikasi Pengenalan Citra Wajah di KTP Menggunakan Google Cloud Vision API dan Kairos API Berbasis Android. *ILKOMNIKA: Journal of Computer Science and Applied Informatics*, 4(2), pp.198-207. DOI: <https://doi.org/10.28926/ilkomnika.v4i2.504>.
- [3] Toha, M.R. and Triayudi, A., 2022. PENERAPAN MEMBACA TULISAN DI DALAM GAMBAR MENGGUNAKAN METODE OCR BERBASIS WEBSITE (STUDI KASUS: e-KTP). *JST (Jurnal Sains dan Teknologi)*, 11(1), pp.175-183. DOI: <https://doi.org/10.23887/jstundiksha.v11i1.42279>.
- [4] Thammarak, K., Kongkla, P., Sirisathitkul, Y. and Intakosum, S., 2022. Comparative analysis of Tesseract and Google Cloud Vision for Thai vehicle registration certificate. *International Journal of Electrical and Computer Engineering*, 12(2), pp.1849-1858.
- [5] ElHanafi, A.M., Nurmadi, R., Tommy, T. and Siregar, R., 2020. Pemaparan Teknologi Biometrika Dan Google Cloud Vision Api Di SMK Dwi Tunggal 2 Tanjung. *Jurnal TUNAS*, 1(2), pp.62-65. DOI: <http://dx.doi.org/10.30645/jtunas.v1i2.15>.
- [6] Takapente, C.B., Sompie, S.R. and Poekoel, V.C., 2018. Implementasi Azure Cognitive Service Untuk Aplikasi Pengkategorian Foto. *Jurnal Teknik Informatika*, 13(4), pp. 1-8. DOI: <https://doi.org/10.35793/jti.13.4.2018.28093>.

- [7] G. Sugiarta, D. P. Andini, and S. Hidayatullah, 2021. Ekstraksi Informasi/Data e-KTP Menggunakan *Optical Character Recognition* Convolutional Neural Network. *JTERA (Jurnal Teknologi Rekayasa)*, 6(1), pp. 1-6.
- [8] Pathak, A., Ruhela, A., Saroha, A.K. and Bhardwaj, A., 2019. Examining Robustness of Google Vision API Based on the Performance on Noisy Images. *International Journal of Computer Science and Engineering (JCSE)*, 7(3), pp. 89-93, 2019.
- [9] Mangundap, J.J., Tasripan, T. and Kusuma, H., 2022. Sistem Pengenalan Text Pada Kemasan Obat untuk Membantu Penyandang Tunanetra dengan Keluaran Suara. *Jurnal Teknik ITS*, 11(2), pp.A128-A133. DOI: <http://dx.doi.org/10.12962/j23373539.v11i2.90393>.
- [10] Ruspandi, R.B., Sompie, S.R. and Kambey, F.D., 2018. Implementasi Azure Cognitive Service untuk Aplikasi Speech Recognition. *Jurnal Teknik Informatika*, 13(4), pp. 1-10. DOI: <https://doi.org/10.35793/jti.13.4.2018.28091>.
- [11] El Maghraby, A., 2021. Improving Custom Vision cognitive services model. *Journal of the ACS*, 11. 12(1), pp. 36-63.
- [12] Muhammad, F.D., 2021. Penggunaan e-KTP untuk Registrasi Otomatis Memanfaatkan Sistem OCR Dengan Metode Template Matching Correlation. *Media Jurnal Informatika*, 12(2), pp.57-60.
- [13] Purba, A.M., Harjoko, A. and 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), pp.177-188. DOI: <https://doi.org/10.22146/ijccs.41259>.
- [14] Putri, D.Z., Puspitaningrum, D. and Setiawan, Y., 2018. Konversi Citra Kartu Nama ke Teks Menggunakan Teknik OCR dan Jaro-Winkler Distance. *Jurnal Teknoinfo*, 12(1), pp.1-6. DOI: <https://doi.org/10.33365/jti.v12i1.35>.
- [15] Susanty, M. and Nugroho, H., 2020. Optical Character Recognition Implementation for Admission System in Universitas Pertamina. *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, 11(1), pp.165-170. DOI: <https://doi.org/10.24176/simet.v11i1.3838>.