

EKSTRAKSI INFORMASI NILAI GIZI PADA LABEL KEMASAN MENGUNAKAN CONFUSION-AWARE WEIGHTED LEVENSHTAIN DISTANCE DAN DIRECTION-AWARE SPATIAL ALIGNMENT

Ari Fuzzaman, Tito Pinandita, Elindra Ambar Pambudi, Maulida Ayu Fitriani

Teknik Informatika, Universitas Muhammadiyah Purwokerto
Jl. KH. Ahmad Dahlan, Dusun III, Dukuhwaluh, Kec. Kembaran,
Kabupaten Banyumas, Jawa Tengah 53182, Indonesia
zamanari01@gmail.com

ABSTRAK

Kesadaran masyarakat terhadap kandungan gizi produk makanan semakin meningkat, mendorong kebutuhan sistem ekstraksi informasi nilai gizi secara otomatis berbasis perangkat *mobile*. Permasalahan utama dalam otomatisasi pembacaan label nutrisi terletak pada kesalahan asosiasi spasial antara label dan nilai numerik, serta kesalahan substitusi karakter pada hasil OCR akibat kemiripan visual antar karakter. Penelitian ini bertujuan membangun ekstraksi informasi nilai gizi yang mampu membaca delapan nutrisi utama secara *real-time* dari citra kemasan makanan. Metode yang dikembangkan mengintegrasikan *Confusion-Aware Weighted Levenshtein Distance* (WLD) untuk koreksi teks OCR berbasis matriks konfusi karakter, dan *Direction-Aware Spatial Alignment* untuk penyelarasan nilai numerik berbasis *bounding box*, diimplementasikan dalam arsitektur MVVM dengan mekanisme *human-in-the-loop* dan rekomendasi *Generative AI*. Pengujian terhadap 30 sampel kemasan (240 *field*) menunjukkan rata-rata *Precision* 98,69%, *Recall* 82,07%, *F1-score* 89,61%, dan *Accuracy* 85,42%, dengan rata-rata MAE 4,296. Sistem terbukti andal pada kemasan datar, meskipun label semi-terstruktur dan kemasan melengkung masih menjadi tantangan yang perlu diatasi pada penelitian selanjutnya.

Kata kunci : OCR, Nilai Gizi, *Weighted Levenshtein Distance*, *Spatial Alignment*, *Human-in-the-loop*

1. PENDAHULUAN

Informasi nilai gizi pada kemasan makanan merupakan instrumen penting dalam pengambilan keputusan konsumsi, khususnya dalam konteks peningkatan kesadaran masyarakat terhadap pola makan sehat. Digitalisasi informasi melalui perangkat *mobile* mendorong pengembangan sistem otomatis untuk membaca dan menginterpretasikan label nutrisi secara *real-time*. Teknologi *Optical Character Recognition* (OCR) telah banyak digunakan untuk mengekstraksi teks dari citra kemasan makanan [1], termasuk pada penelitian yang memanfaatkan kombinasi deteksi objek dan OCR untuk mengidentifikasi kandungan nutrisi tertentu seperti natrium [2].

Namun demikian, label nilai gizi memiliki karakteristik semi-terstruktur menyerupai tabel tanpa garis eksplisit, sehingga hubungan antara label dan nilai numerik tidak selalu dapat diidentifikasi secara langsung melalui OCR mentah [3]. Permasalahan utama dalam ekstraksi label gizi tidak hanya terletak pada kesalahan pengenalan karakter, tetapi juga pada kesalahan asosiasi spasial antara label dan nilai. OCR sering menghasilkan substitusi karakter akibat kemiripan visual, seperti *g* terbaca sebagai *9*, *l* sebagai *1*, atau *o* sebagai *0*, yang berdampak langsung pada hilangnya informasi nutrisi tertentu [4], [5]. Selain itu, tanpa mekanisme penyalarsan berbasis koordinat, sistem cenderung mengambil nilai dari baris yang tidak sesuai atau dari bagian *footer* seperti kebutuhan energi harian (misalnya 2150 kkal). Pendekatan berbasis *deep learning* untuk analisis struktur

dokumen memang menunjukkan performa tinggi, namun memerlukan *dataset* pelatihan yang besar dan sumber daya komputasi signifikan [6].

Penelitian ini bertujuan membangun ekstraksi informasi nilai gizi yang mampu membaca delapan nutrisi utama secara akurat dan *real-time* dari citra kemasan makanan, dengan menggabungkan koreksi karakter berbasis konfusi visual dan penyalarsan spasial berbasis arah secara terintegrasi dalam satu *pipeline* ringan tanpa pelatihan model tambahan.

Oleh karena itu, penelitian ini mengusulkan pendekatan hibrida yang mengintegrasikan *Confusion-Aware Weighted Levenshtein Distance* untuk meningkatkan toleransi kesalahan karakter OCR, serta *Direction-Aware Spatial Alignment* berbasis *bounding box* untuk menjaga konsistensi relasi label-nilai pada struktur semi-terstruktur. Pendekatan ini dirancang bersifat deterministik, ringan, dan tidak memerlukan pelatihan model tambahan. Sistem diimplementasikan menggunakan arsitektur MVVM pada Android dan dilengkapi mekanisme *human-in-the-loop* untuk validasi manual sebelum integrasi dengan modul rekomendasi berbasis *Generative AI*.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Sejumlah penelitian sebelumnya telah berupaya mengembangkan ekstraksi informasi dan teks secara otomatis. Dalam konteks analisis kemasan, E.Kina [1] mengembangkan sistem deteksi alergen berbasis OCR yang dipadukan dengan *machine learning* untuk membaca informasi kemasan secara *real-time*. Khusus

untuk label nilai gizi, Febrianti Malelak dkk. [2] mengintegrasikan deteksi objek berbasis FPN dan Faster R-CNN untuk mengisolasi area informasi natrium guna meningkatkan akurasi spesifik sebelum dilakukan OCR.

Terkait dengan perbaikan hasil ekstraksi teks, Kadek dkk. [4] mengimplementasikan *Levenshtein Distance* dalam aplikasi Android berbasis OCR yang terbukti efektif dalam mentoleransi kesalahan minor akibat perbedaan karakter. Hal serupa juga dibuktikan oleh Had dkk. [5], yang menunjukkan bahwa pendekatan *post-processing* berbasis algoritma koreksi teks mampu meningkatkan akurasi OCR secara signifikan. Di sisi lain, untuk menangani struktur tata letak dokumen, Lymantama dkk. [3] mengusulkan pendekatan berbasis posisi (koordinat teks) yang terbukti lebih stabil dibandingkan pendekatan berbasis urutan teks semata dalam mengekstraksi tabel semi-terstruktur dari dokumen PDF.

Meskipun demikian, penelitian-penelitian tersebut menangani koreksi karakter dan penyelarasan spasial secara terpisah dan belum diujikan pada domain label nilai gizi kemasan makanan dalam konteks aplikasi *mobile real-time*. Koreksi teks yang ada belum mempertimbangkan pola substitusi karakter yang khas pada OCR kemasan, sementara pendekatan spasial yang ada dirancang untuk dokumen PDF statis, bukan untuk tata letak semi-terstruktur tanpa garis pembatas pada citra kamera. Kesenjangan inilah yang menjadi landasan utama penelitian ini.

2.2. Optical Character Recognition

Optical Character Recognition (OCR) adalah teknologi yang memungkinkan pengenalan dan ekstraksi teks dari dokumen atau citra digital yang dipindai, kemudian mengubahnya menjadi data yang dapat dianalisis, diedit, dan dicari [7]. Teknologi OCR telah banyak dimanfaatkan untuk membaca informasi pada kemasan makanan, terutama dalam konteks deteksi alergen dan kandungan nutrisi secara otomatis.

Namun demikian, performa OCR sangat dipengaruhi oleh kualitas citra masukan. Pada kemasan makanan, faktor seperti resolusi rendah, distorsi geometris akibat permukaan melengkung, variasi font, serta latar belakang bertekstur dapat menurunkan akurasi pengenalan karakter secara signifikan. Kondisi ini menyebabkan munculnya substitusi karakter yang sistematis akibat kemiripan visual. Oleh karena itu, tahap *post-processing* berbasis koreksi teks menjadi komponen yang tidak dapat diabaikan dalam sistem ekstraksi berbasis OCR pada domain kemasan.

2.3. Google ML Kit

Dalam pengembangan aplikasi berbasis *mobile*, efisiensi ekstraksi teks sangat bergantung pada kemampuan perangkat dalam memproses citra secara lokal. Google ML Kit hadir sebagai SDK *on-device machine learning* yang dioptimalkan untuk performa

tinggi tanpa memerlukan koneksi internet secara konstan.

Penelitian oleh Zufar dkk. [8] menunjukkan bahwa pemanfaatan *API Text Recognition* pada ML Kit mampu mengenali blok teks dengan tingkat keberhasilan yang tinggi, sekaligus membagi teks ke dalam elemen baris dan kolom yang mempermudah analisis spasial pada dokumen terstruktur.

Dengan demikian, Google ML Kit merupakan fondasi yang ideal bagi arsitektur aplikasi *mobile* yang membutuhkan fitur *Optical Character Recognition* (OCR) mandiri. Perpaduan antara efisiensi pemrosesan *on-device*, kemandirian dari akses internet, serta presisi dalam analisis tata letak teks menjadikannya solusi yang tangguh untuk mengekstraksi data secara cepat dan akurat langsung dari perangkat pengguna.

2.4. Levenshtein Distance

Algoritma *Levenshtein Distance* adalah suatu pengukuran untuk menghitung jumlah perbedaan antara dua kata [9]. Kesalahan karakter pada hasil OCR merupakan masalah umum, terutama pada dokumen dengan resolusi rendah atau distorsi visual. Salah satu pendekatan yang banyak digunakan untuk mengatasi masalah ini adalah algoritma *Levenshtein Distance*.

Pradana dkk. [10] memanfaatkan algoritma *Levenshtein Distance* untuk meningkatkan akurasi translasi karakter berbasis OCR. Penelitian tersebut menekankan pentingnya pencocokan terhadap kamus referensi guna meminimalkan kesalahan substitusi karakter. Selain itu, Had dkk. [5] menunjukkan bahwa pendekatan *post-processing* berbasis algoritma koreksi teks (seperti SymSpell dan *edit distance*) mampu meningkatkan akurasi OCR pada data paspor secara signifikan. Hal ini menunjukkan bahwa koreksi berbasis jarak teks merupakan tahap penting setelah proses OCR mentah.

Namun demikian, sebagian besar penelitian tersebut menggunakan *edit distance* standar tanpa mempertimbangkan bobot khusus untuk pasangan karakter yang sering tertukar akibat kesalahan OCR. Oleh karena itu, pendekatan *Weighted Levenshtein Distance* yang mempertimbangkan karakter *confusion* menjadi relevan untuk dikembangkan lebih lanjut.

2.5. Analisis Spasial dan Ekstraksi Tabel Semi-Terstruktur

Label nilai gizi pada kemasan makanan umumnya memiliki struktur semi-terstruktur menyerupai tabel tanpa garis pembatas yang jelas. Tantangan utama pada struktur ini adalah menjaga konsistensi hubungan antara label dan nilai numerik dalam baris yang sama.

Lymantama dkk. [3] mengusulkan pendekatan dual untuk ekstraksi tabel dari dokumen PDF *native* dan *ter-scan*, dengan memanfaatkan informasi spasial koordinat teks. Penelitian tersebut menunjukkan bahwa pendekatan berbasis posisi lebih stabil

dibandingkan pendekatan berbasis urutan teks semata. Selain itu, Saputra dkk. [11] dalam penelitiannya menggunakan *bounding box* dalam proses OCR untuk menjaga akurasi ekstraksi teks pada label nutrisi makanan yang sering kali memiliki tata letak padat dan tanpa garis.

Berdasarkan penelitian tersebut, pendekatan berbasis *alignment* spasial menjadi komponen penting dalam ekstraksi label gizi yang memiliki struktur semi-teratur.

2.6. Arsitektur Sistem Mobile Berbasis MVVM

Pengembangan aplikasi Android modern banyak menggunakan arsitektur Model–View–ViewModel (MVVM) untuk meningkatkan modularitas dan pemisahan logika bisnis dari antarmuka pengguna.

Ulhaq dkk. [12] menerapkan arsitektur MVVM pada aplikasi pencatatan kalori harian dan menunjukkan bahwa pola ini meningkatkan *maintainability* serta skalabilitas sistem. Pratama dkk. [13] juga mengkaji optimasi arsitektur MVVM pada aplikasi Android berbasis *RESTful API*, yang menunjukkan peningkatan kinerja dan efisiensi pemrosesan data.

Arsitektur ini relevan dalam penelitian karena sistem tidak hanya melakukan OCR, tetapi juga menyimpan data nutrisi, memproses rekomendasi AI, dan menampilkan hasil secara dinamis kepada pengguna.

2.7. Human-in-the-loop

Pendekatan *human-in-the-loop* memungkinkan pengguna untuk melakukan validasi atau koreksi terhadap hasil ekstraksi otomatis. Lazaros dkk. [14] menunjukkan bahwa integrasi manusia dalam *pipeline machine learning* dapat meningkatkan kualitas hasil ekstraksi dengan memadukan pengawasan manusia dan keputusan otomatis di berbagai tahap pemrosesan.

Dalam konteks ekstraksi teks, mekanisme koreksi manual oleh pengguna memungkinkan peningkatan akurasi akhir meskipun sistem OCR tidak sepenuhnya sempurna. Pendekatan ini meningkatkan reliabilitas sistem tanpa menambah kompleksitas model.

2.8. Generative AI

Perkembangan *Generative AI* membuka peluang baru dalam sistem rekomendasi nutrisi yang adaptif. Yang dkk. [15] memperkenalkan *ChatDiet*, sebuah kerangka kerja berbasis *Large Language Models* (LLM) yang mampu memberikan rekomendasi makanan yang dipersonalisasi dengan menggabungkan data nutrisi populasi dan profil fisiologis pengguna secara spesifik. Sementara itu, Deng dkk. [16] mengembangkan *NutriRec*, sebuah sistem berbasis LLM multimodal yang mengintegrasikan penalaran nutrisi dengan pola diet kolaboratif untuk menghasilkan saran makanan yang lebih terukur. Di sisi lain, Zhang dkk. [17] menunjukkan bahwa penggunaan LLM yang

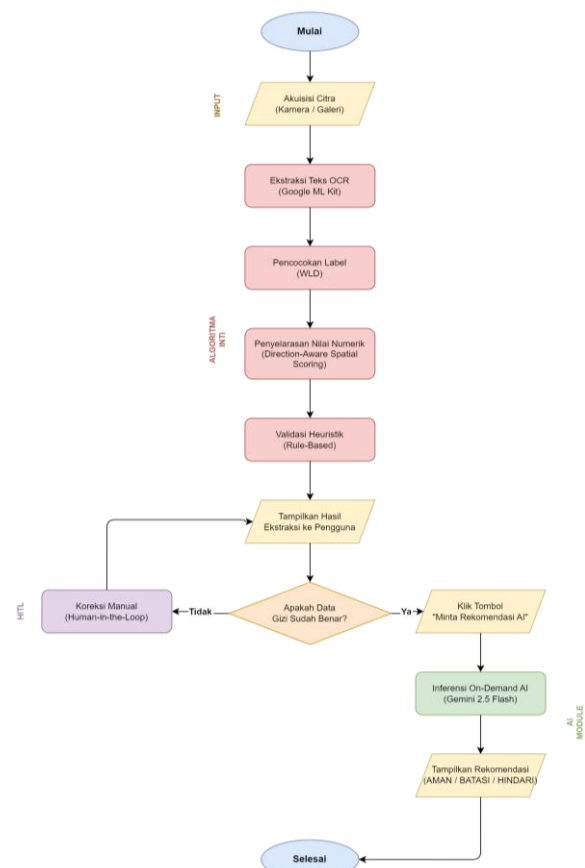
ikombinasikan dengan *Knowledge Graph* dapat membantu sistem dalam menangani batasan diet yang sangat ketat melalui strategi relaksasi konstrain yang dinamis.

Integrasi antara ekstraksi nilai gizi otomatis dan modul rekomendasi berbasis *Generative AI* memberikan nilai tambah pada sistem yang dikembangkan dalam penelitian ini, memungkinkan transformasi data mentah dari label makanan menjadi saran tindakan yang relevan bagi pengguna.

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan eksperimental dengan membangun sistem ekstraksi nilai gizi berbasis Android yang bekerja secara *real-time*. Sistem dirancang untuk membaca label nutrisi kemasan makanan melalui citra, mengekstraksi delapan nutrisi utama, serta menghasilkan rekomendasi konsumsi berbasis profil pengguna.

Secara umum, sistem ini terdiri dari 5 modul utama, yakni akuisisi citra kemasan makanan, proses *Optical Character Recognition* (OCR) menggunakan Google ML Kit, koreksi label berbasis *Confusion-Aware Weighted Levenshtein Distance*, penyelarasan nilai numerik menggunakan *Direction-Aware Spatial Alignment*, dan rekomendasi nutrisi berbasis *Generative AI*. Alur kerja sistem secara keseluruhan dapat dilihat pada Gambar 1, sedangkan penjelasan rinci setiap modul diuraikan pada sub-bagian.



Gambar 1. Alur sistem ekstraksi nilai gizi berbasis Android

Berdasarkan alur pada Gambar 1, sistem dimulai dari akuisisi citra label kemasan menggunakan kamera perangkat Android. Citra kemudian diproses oleh mesin OCR Google ML Kit untuk menghasilkan blok teks beserta koordinat *bounding box*-nya. Selanjutnya, modul WLD mencocokkan setiap blok teks dengan kamus nutrisi untuk mengidentifikasi label gizi, sementara modul *Direction-Aware Spatial Alignment* memilih kandidat nilai numerik yang paling relevan secara spasial. Hasil ekstraksi kemudian divalidasi secara heuristik dan ditampilkan kepada pengguna melalui antarmuka *human-in-the-loop* untuk koreksi manual sebelum dikirim ke modul *Generative AI* guna menghasilkan rekomendasi konsumsi yang dipersonalisasi.

3.1. Pengumpulan Data

Data penelitian ini merupakan citra digital label informasi nilai gizi (*Nutrition Facts*) yang diperoleh melalui akuisisi langsung (*primary data collection*). Proses pengambilan citra dilakukan menggunakan perangkat bergerak Samsung Galaxy M21 dengan spesifikasi kamera utama beresolusi 48 MP dan Redmi Note 9 dengan spesifikasi kamera utama beresolusi 64 MP. Pemilihan perangkat ini bertujuan untuk merepresentasikan kemampuan perangkat kelas menengah (*mid-range*) yang menjadi target pengguna utama sistem, serta untuk menguji efisiensi pemrosesan *on-device* menggunakan Google ML Kit.

Total *dataset* yang dikumpulkan berjumlah 30 label kemasan produk makanan dan minuman yang beredar di Indonesia dengan total *field* 240. Sampel tersebut dipilih menggunakan teknik *purposive sampling* dengan variasi karakteristik sebagai berikut:

- Variasi Struktur Tata Letak, sampel mencakup label dengan format tabel bergaris penuh dan format semi-terstruktur (tanpa garis pembatas vertikal/horizontal) yang sering menyebabkan kesalahan asosiasi spasial.
- Variasi Kondisi Fisik, sampel diambil dari berbagai jenis material kemasan, termasuk kotak karton (permukaan datar) dan plastik fleksibel atau kaleng (permukaan melengkung/distorsi silindris).
- Variasi Pencahayaan, citra diambil dalam kondisi pencahayaan ruangan normal untuk mensimulasikan skenario penggunaan nyata oleh konsumen.

Seluruh data citra kemudian divalidasi dan diberi label manual (*ground truth*) mencakup delapan atribut nutrisi utama (Energi Total, Energi dari Lemak, Lemak Total, Lemak Jenuh, Protein, Karbohidrat Total, Gula, dan Garam/Natrium) untuk keperluan evaluasi akurasi sistem. Contoh representasi variasi *dataset* label kemasan yang digunakan dapat dilihat pada Gambar 2.



Gambar 2. Contoh variasi *dataset* label kemasan: (a) tabel bergaris penuh, (b) semi-terstruktur, (c) permukaan melengkung

Berdasarkan Gambar 2 terlihat bahwa variasi karakteristik visual pada label kemasan menciptakan tantangan yang berbeda-beda dalam proses ekstraksi informasi nilai gizi. Keberagaman ini sengaja dimasukkan ke dalam *dataset* untuk menguji sejauh mana sistem mampu mempertahankan akurasi pada kondisi dunia nyata.

3.2. Akuisisi dan Prapemrosesan Citra

Proses dimulai dengan pengambilan citra menggunakan pustaka *CameraX*. Sistem menerapkan fitur kontrol iluminasi secara opsional untuk meminimalkan *noise* visual dari elemen desain kemasan yang tidak relevan. Fitur ini menggunakan (*LED Flash*) guna meningkatkan kontras teks tabel gizi terhadap latar belakang kemasan, memastikan *engine* mendapat masukan citra dengan pencahayaan yang stabil.

3.3. Ekstraksi dan *Direction-Aware Spatial Scoring*

Sistem menggunakan Google ML Kit untuk ekstraksi teks secara *on-device*. Mengingat label gizi sering kali memiliki tata letak semi-terstruktur tanpa garis pembatas yang jelas, diusulkan algoritma *Direction-Aware Spatial Alignment*. Algoritma ini menghitung skor kedekatan (S) antara pusat koordinat *bounding box* label gizi (L) dan kandidat nilai numerik (V). Pertama, dihitung selisih koordinat pusat antara label dan kandidat nilai:

Kandidat dieliminasi apabila berada di luar rentang *filter* arah, yaitu $-40 \leq \Delta y \leq 90$ dan $\Delta x \geq -40$. Skor dasar dihitung menggunakan *Weighted Manhattan distance* dengan bobot asimetris yang memberikan penalti lebih besar pada sumbu vertikal, karena pada tabel gizi kandidat nilai cenderung sejajar horizontal dengan labelnya:

$$\Delta x = x_V - x_L, \quad \Delta y = y_V - y_L \quad (1)$$

Skor dasar kemudian dihitung menggunakan *Weighted Manhattan distance* dengan pembobotan asimetris (nilai Δy diberi bobot 2 dan Δx diberi bobot 0,5):

$$S_{base}(L, V) = 2|\Delta y| + 0,5|\Delta x| \quad (2)$$

Selanjutnya diterapkan dua faktor reduksi arah. Faktor α memberikan prioritas pada kandidat yang berada di sisi kanan label, dan faktor β memberikan prioritas pada kandidat yang sejajar secara vertikal dengan label:

$$\alpha = \begin{cases} 0,8 & \text{jika } \Delta x \geq 0 \\ 1,0 & \text{lainnya} \end{cases}, \quad \beta = \begin{cases} 0,85 & \text{jika } |\Delta y| \leq 50 \\ 1,0 & \text{lainnya} \end{cases} \quad (3)$$

Skor akhir diperoleh dengan menerapkan kedua faktor reduksi pada skor dasar:

$$S(L, V) = S_{base} \cdot \alpha \cdot \beta \quad (4)$$

Nilai numerik terpilih (V^*) adalah kandidat dengan skor S minimum dari seluruh kandidat yang lolos *filter* arah:

$$V^* = \arg \min V \in \mathcal{H} \quad S(L, V) \quad (5)$$

Pembobotan asimetris pada skor dasar (bobot 2 untuk Δy dan 0,5 untuk Δx) dipilih karena pada struktur label gizi, nilai numerik umumnya berada di baris yang sama dengan labelnya. Oleh karena itu, penyimpangan vertikal yang besar jauh lebih kritis dibandingkan penyimpangan horizontal, karena dapat menyebabkan asosiasi nilai dari baris nutrisi yang berbeda. Ilustrasi mekanisme pemilihan kandidat nilai dapat dilihat pada Gambar 3.

Berdasarkan Gambar 3 Algoritma bekerja dengan menemukan skor terendah berdasarkan jarak berbobot, yang mana cenderung ke kanan, dengan (+dx) dan penyelarasan yang hampir vertikal.

Gambar 3. Ilustrasi *Direction-Aware Spatial Alignment*: area *filter* arah dan skor kedekatan antar kandidat nilai

3.4. Koreksi Teks dengan *Weighted Levenshtein Distance* (WLD)

Penanganan kesalahan karakter OCR akibat kemiripan visual (misalnya: 'g' terbaca sebagai '9', atau 'l' sebagai '1') dilakukan menggunakan metode *Weighted Levenshtein Distance* (WLD). Berbeda dengan *Levenshtein* standar, WLD memberikan biaya substitusi yang lebih kecil pada pasangan karakter yang sering tertukar (*confusion pairs*) berdasarkan matriks konfusi yang didefinisikan sebagai:

$$c(a, b) = \begin{cases} 0 & \text{jika } a = b \\ w_a(a, b) & \text{jika } (a, b) \in \mathcal{C} \\ 1 & \text{lainnya} \end{cases} \quad (6)$$

dengan $\mathcal{C}$ adalah himpunan *confusion pairs*, dan $w_a(a, b)$ adalah bobot konfusi antara karakter a dan b . Nilai bobot yang digunakan: 0,2 untuk pasangan ('g', '9') dan ('o', '0'), serta 0,3 untuk pasangan ('l', '1') dan ('s', '5'). Rekurens pemrograman dinamis untuk menghitung WLD adalah:

$$D(i, j) = \min \begin{cases} D(i-1, j) + 1 & (\text{hapus}) \\ D(i, j-1) + 1 & (\text{sisip}) \\ D(i-1, j-1) + c(s_1[i], s_2[j]) & (\text{substitusi}) \end{cases} \quad (7)$$

dengan kondisi awal $D(i, 0) = i$ dan $D(0, j) = j$. Nilai akhir WLD antara string s_1 dan s_2 adalah $D(|s_1|, |s_2|)$. Skor kemiripan (*similarity*) kemudian dinormalisasi terhadap panjang alias referensi:

$$\text{sim}(t, a) = 1 - \frac{WLD(t, a)}{|a|} \quad (8)$$

dengan t adalah subtingkat teks OCR dan a adalah alias referensi dari kamus nutrisi. Sebuah label dicocokkan apabila memenuhi ambang batas kemiripan:

$$\text{sim}(t, a) > \tau, \quad \tau = 0,7 \quad (9)$$

Penetapan bobot konfusi dilakukan berdasarkan tingkat kemiripan visual antar karakter pada hasil OCR. Berikut ditampilkan pada Tabel 1.

Tabel 1. *Confusion pairs* dan Bobot Substitusi WLD

Pasangan (a, b)	Bobot $w_a(a,b)$	Alasan Kemiripan Visual
('g', '9') dan ('9', 'g')	0,2	Bentuk lengkung bawah sangat mirip pada font sans-serif
('o', '0') dan ('0', 'o')	0,2	Bentuk oval hampir identik pada resolusi rendah
('l', '1') dan ('1', 'l')	0,3	Garis vertikal tipis, mirip pada kemasan bertekstur
('s', '5') dan ('5', 's')	0,3	Lekukan serupa pada font kecil beresolusi rendah

Dari Tabel 1 dapat dilihat pasangan ('g', '9') dan ('o', '0') diberi bobot lebih kecil (0,2) karena kemiripan bentuk yang sangat tinggi sehingga substitusinya lebih dapat ditoleransi. Sebaliknya, pasangan ('l', '1') dan ('s', '5') diberi bobot lebih besar (0,3) karena meskipun sering tertukar, perbedaan visualnya masih cukup membedakan konteks label dari nilai numerik.

3.5. Validasi Heuristik dan Pendekatan *Human-in-the-loop* (HITL)

Meskipun koreksi otomatis menggunakan WLD telah diterapkan, sistem tetap memerlukan mekanisme validasi untuk menangani ambiguitas data yang lolos dari *filter* algoritma. Oleh karena itu, diterapkan pendekatan dua lapis, yaitu Validasi Heuristik (System-Level) yang menerapkan aturan logika standar (*rule-based*) untuk menormalkan data sebelum ditampilkan ke pengguna seperti konversi unit dengan mendeteksi "gram", "g", dan "mg" lalu pembersihan simbol dengan menghapus karakter non numerik seperti titik dua ":", tanda kurung "()", maupun simbol lainnya. Selain itu, logika ambang batas juga diterapkan dengan mendeteksi nilai anomali, seperti gula yang melebihi total energi yang akan ditandai dengan *low confidence*.

Selanjutnya adalah validasi manual *Human-in-the-loop* (HITL) diintegrasikan untuk memberikan kendali akhir kepada pengguna. Sebelum data dikirim ke modul AI, pengguna disajikan antarmuka pratinjau (*preview interface*) yang menampilkan hasil ekstraksi. Pengguna dapat melakukan verifikasi dan koreksi manual jika terdapat kesalahan bacaan yang fatal. Mekanisme ini memastikan bahwa *prompt* yang dikirim ke *Generative AI* bersih dari kesalahan data sehingga meningkatkan reliabilitas rekomendasi yang dihasilkan.

3.6. Inferensi Cerdas Berbasis On-Demand AI

Data nutrisi yang telah diekstraksi dan divalidasi kemudian dikirimkan ke model *Generative AI*. Sistem menyusun *prompt* yang menyertakan data nutrisi mentah beserta profil pengguna. AI melakukan penalaran (*reasoning*) untuk menghasilkan analisis apakah produk tersebut aman dikonsumsi atau perlu

dibatasi, memberikan saran yang lebih personal dibandingkan sistem berbasis aturan statis.

3.7. Metode Evaluasi

Pengujian kinerja sistem dilakukan menggunakan dua pendekatan evaluasi utama yaitu evaluasi klasifikasi teks dengan mengukur keberhasilan sistem dalam mengenali dan mengoreksi teks label nutrisi menggunakan metrik *Precision*, *Recall*, *F1-score*, dan *Accuracy*.

a. Precision

$$Precision = \frac{TP}{TP + FP} \quad (10)$$

b. Recall

$$Recall = \frac{TP}{TP + FN} \quad (11)$$

c. F1-score

$$F1-Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (12)$$

d. Accuracy

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (13)$$

Selanjutnya adalah evaluasi akurasi numerik (MAE) untuk mengukur rata-rata selisih absolut antara nilai gizi hasil ekstraksi sistem ($\hat{y}$) dengan nilai sebenarnya (y) pada kemasan untuk delapan parameter nutrisi.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (14)$$

Nilai MAE yang rendah mengindikasikan bahwa sistem mampu mengekstraksi angka kuantitatif dengan presisi tinggi, yang krusial untuk perhitungan asupan harian pengguna.

4. HASIL DAN PEMBAHASAN

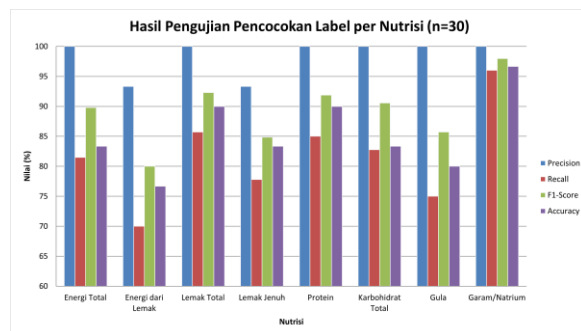
Pengujian dilakukan terhadap 30 sampel label kemasan makanan yang dipilih secara *purposive sampling* dengan mempertimbangkan variasi tata letak dan kondisi fisik kemasan. Evaluasi mencakup dua aspek utama: akurasi pencocokan label nutrisi menggunakan metrik *Precision*, *Recall*, *F1-score*, dan *Accuracy*; serta akurasi ekstraksi nilai numerik menggunakan *Mean Absolute Error* (MAE). Selain itu dilakukan analisis kasus kegagalan dan pengujian antarmuka *human-in-the-loop*.

4.1. Hasil Pengujian Pencocokan Label (WLD)

Pencocokan label dilakukan dengan membandingkan hasil deteksi WLD terhadap *ground truth* yang dianotasi manual. Label dianggap berhasil jika nilai kemiripan WLD melampaui ambang batas $\tau = 0,7$. Hasil per nutrisi disajikan pada Gambar 4.

Berdasarkan Gambar 4, label Garam/Natrium memperoleh *F1-score* tertinggi sebesar 97,96% dengan *Precision* dan *Accuracy* masing-masing 100,00% dan 96,67%. Hal ini disebabkan oleh kata

“Natrium” yang memiliki pola karakter distinktif sehingga jarang mengalami konfusi dengan label nutrisi lain. Sebaliknya, Energi dari Lemak mencatat *F1-score* terendah sebesar 80,00% dengan *Recall* 70,00%, yang mengindikasikan bahwa sistem masih melewatkan sebagian nilai pada label tersebut. Penyebab utamanya adalah variasi penulisan pada kemasan, seperti “Energi Lemak”, “Kal. dari Lemak”, atau sekadar “From Fat”, yang memerlukan toleransi lebih tinggi dari ambang batas WLD yang ditetapkan ($\tau = 0,7$). Secara keseluruhan, sistem mencapai rata-rata *Precision* 98,69%, *Recall* 82,07%, *F1-score* 89,61%, dan *Accuracy* 85,42% di atas 240 *field* evaluasi, yang menunjukkan performa pencocokan label yang memadai.



Gambar 4. Hasil pengujian pencocokan label per nutrisi

4.2. Hasil Ekstraksi Nilai Numerik (MAE)

Evaluasi akurasi nilai numerik dilakukan dengan menghitung MAE antara nilai yang diekstraksi sistem dengan *ground truth* per nutrisi dari 30 sampel. Hasil selengkapnya disajikan pada Tabel 2.

Tabel 2. Hasil MAE Ekstraksi Nilai Numerik per Nutrisi

Nutrisi	Satuan	MAE	Error Relatif (%)
Energi Total	kcal	17,833	15,84%
Energi dari Lemak	kcal	5,733	13,16%
Lemak Total	g	1,217	22,31%
Lemak Jenuh	g	1,217	44,24%
Protein	g	1,900	69,09%
Karbohidrat Total	g	3,267	21,05%
Gula	g	1,033	12,92%
Garam/Natrium	mg	2,167	1,11%
Rata-rata	-	4,296	24,97%

Berdasarkan Tabel 2, Energi Total mencatat MAE tertinggi secara absolut sebesar 17,833 kkal dengan *error* relatif 15,84%, yang wajar mengingat satuan kkal memiliki rentang nilai jauh lebih besar dibandingkan nutrisi lain. Protein menunjukkan *error* relatif tertinggi (69,09%) meskipun MAE-nya hanya 1,900 g, karena nilai protein pada sampel umumnya sangat kecil sehingga penyimpangan kecil menghasilkan persentase *error* yang besar. Di sisi lain, Garam/Natrium memiliki *error* relatif terendah (1,11%) meski satuannya mg, yang berarti sistem

sangat andal dalam membaca nilai natrium secara numerik. Gula mencatat MAE terkecil (1,033 g) yang menunjukkan konsistensi ekstraksi nilai numerik pada nutrisi tersebut. Secara keseluruhan, rata-rata MAE sebesar 4,296 dengan *error* relatif 24,97% menunjukkan bahwa sistem mampu mengekstraksi nilai numerik dengan tingkat presisi yang cukup untuk perhitungan asupan harian.

4.3. Analisis Kasus Gagal

Dari 30 sampel yang diuji, terdapat sejumlah kasus di mana sistem menghasilkan ekstraksi yang tidak akurat. Kegagalan dikategorikan berdasarkan tahap terjadinya *error* dalam *pipeline* sistem. Distribusi penyebab kegagalan disajikan pada Tabel 3 dan contoh visual kasus berhasil serta gagal ditampilkan pada Tabel 4.

Tabel 3. Distribusi Penyebab Kegagalan Ekstraksi

Kategori Kegagalan	Deskripsi	Jumlah	Persentase
Kesalahan substitusi karakter OCR	Karakter terbaca salah oleh ML Kit, WLD tidak berhasil mengoreksi	5	14,3 %
Misalignment spasial	Nilai dipilih dari baris nutrisi yang berbeda akibat <i>layout</i> tidak beraturan atau kemasan duplikasi	22	62,9 %
Distorsi citra kemasan	Kemasan silindris/melengkung menyebabkan karakter tidak terbaca	6	17,1 %
False positive (ambiguitas)	Sistem memprediksi nilai pada <i>field</i> yang seharusnya nol (GT=0)	2	5,7%
Total	-	35	100 %

Berdasarkan Tabel 3, dari total 35 *field* yang tidak berhasil diekstraksi dengan benar (FP=2, FN=33), *misalignment* spasial menjadi penyebab dominan dengan 22 kasus (62,9%). Pola kegagalan ini paling banyak terjadi pada kemasan dengan label semi-terstruktur tanpa garis pembatas, seperti pada sampel roma sandwich, ramen yes, dan tic-tic, di mana algoritma *Direction-Aware Spatial Alignment* memilih nilai dari baris nutrisi yang berdekatan akibat jarak vertikal antar baris yang sangat kecil. Kesalahan distorsi kemasan silindris menyumbang 6 kasus (17,1%), terutama pada kaleng melengkung (sampel nabati dan abc squash) yang menyebabkan karakter terpotong atau tidak terbaca oleh ML Kit. Kesalahan substitusi karakter OCR berkontribusi 5 kasus (14,3%), seperti gula pada sampel pop mie yang terbaca 8 alih-alih 3 akibat kemiripan visual pada citra *noise*. Sementara itu, 2 kasus *false positive* (5,7%) terjadi ketika sistem memprediksi nilai pada *field* yang seharusnya nol, umumnya akibat ambiguitas posisi *bounding box* pada label dengan tata letak padat.

Tabel 4. Contoh Kasus Berhasil vs Gagal (*side by side*)

Gambar	Output	Deskripsi
	Klasifikasi Hasil (Untuk untuk E40) Energi Total 40.0 kkal ✓ Energi dari Lemak 0.0 kkal ✓ Lemak Total 0.0 g ✓ Lemak Jenuh 0.0 g ✓ Karbohidrat Total 0.0 g ✓ Gula 0.0 g ✓ Protein 0.0 g ✓ Natrium 240.0 mg ✓ Klasifikasi Hasil (Untuk untuk E40)	Karakter terbaca dengan benar karena citra jelas dan label terstruktur
	Klasifikasi Hasil (Untuk untuk E40) Energi Total 0.0 kkal ✓ Energi dari Lemak 0.0 kkal ✓ Lemak Total 0.0 g ✓ Lemak Jenuh 0.0 g ✓ Karbohidrat Total 0.0 g ✓ Gula 0.0 g ✓ Protein 0.0 g ✓ Natrium 0.0 mg ✓ Klasifikasi Hasil (Untuk untuk E40)	Gagal terbaca karena bentuk silindris dan label tidak terstruktur

Tabel 4 menampilkan perbandingan kasus berhasil dan gagal secara visual. Pada kasus berhasil, seluruh 8 *field* nutrisi berhasil diekstraksi dengan benar karena citra tajam, tabel terstruktur, dan tidak ada distorsi permukaan. Sebaliknya, pada kasus gagal sistem gagal mengekstraksi lebih dari 4 *field* akibat kombinasi permukaan silindris dan kondisi label yang tidak terstruktur.

Dari total 30 sesi pengujian, sebanyak 18 sampel (60%) memerlukan minimal satu koreksi manual melalui antarmuka HITL sebelum data dianggap valid. Rata-rata jumlah *field* yang dikoreksi per sesi adalah 1,94 *field*. Kasus yang paling banyak memerlukan koreksi adalah sampel dengan kemasan silindris dan semi-terstruktur, sebagaimana tercermin dari distribusi kegagalan pada Tabel 3. Mekanisme HITL terbukti efektif dalam meningkatkan keandalan data sebelum dikirim ke modul *Generative AI*, memastikan *prompt* yang dihasilkan bebas dari kesalahan ekstraksi yang dapat memengaruhi kualitas rekomendasi nutrisi.

5. KESIMPULAN DAN SARAN

Penelitian ini berhasil membangun ekstraksi nilai gizi menggunakan kombinasi *Confusion-Aware Weighted Levenshtein Distance* dan *Direction-Aware Spatial Alignment* yang bekerja secara *on-device* tanpa memerlukan pelatihan model tambahan. Pengujian terhadap 30 sampel kemasan (240 *field*) menghasilkan rata-rata *Precision* 98,69%, *Recall* 82,07%, *F1-score* 89,61%, dan *Accuracy* 85,42% untuk pencocokan label, serta rata-rata MAE 4,296 dengan *error* relatif 24,97% untuk ekstraksi nilai numerik. Label

Garam/Natrium mencatat performa tertinggi (*F1* 97,96%) sedangkan Energi dari Lemak menjadi yang terendah (*F1* 80,00%) akibat variasi penulisan pada kemasan. Analisis kegagalan menunjukkan bahwa *misalignment* spasial pada label semi-terstruktur merupakan penyebab dominan (62,9%), diikuti distorsi kemasan silindris (17,1%) dan kesalahan substitusi karakter OCR (14,3%). Mekanisme *human-in-the-loop* terbukti efektif mengurangi propagasi kesalahan ke modul *Generative AI*, di mana 60% sampel memerlukan minimal satu koreksi manual dengan rata-rata 1,94 *field* per sesi. Untuk penelitian selanjutnya, disarankan pengembangan model deteksi area label gizi berbasis *deep learning* untuk meningkatkan *robustness* pada kemasan silindris, serta eksplorasi teknik *image rectification* guna meminimalkan distorsi geometris sebelum proses OCR dilakukan.

DAFTAR PUSTAKA

- [1] E. Kina, "Real-time food allergen detection using OCR-enhanced machine learning techniques," *PeerJ Comput. Sci.*, vol. 11, hlm. e3338, Des 2025, doi: 10.7717/peerj-cs.3338.
- [2] R. F. Malelak, A. Fanggidae, T. Widiastuti, dan Y. T. Polly, "Deteksi Informasi Kadar Natrium pada Label Produk Kemasan Menggunakan FPN, Faster R-CNN, dan OCR," *Jurnal Buana Informatika*, vol. 16, no. 2, hlm. 155–165, 2025.
- [3] W. V. Lymanama, Z. Rustam, dan F. Novkaniza, "Optimizing Table Extraction from PDFs: A Dual," dalam *Proceedings of the 2024 Brawijaya International Conference (BIC 2024)*, 2025, hlm. 337.
- [4] I. Kadek dkk., "Esscore: An OCR-Based Android App for Scoring Short Handwritten Answer Using Levenshtein Distance," 2025. [Daring]. Tersedia pada: <http://jurnal.polibatam.ac.id/index.php/JAIC>
- [5] I. S. Had, W. Maulana Baihaqi, dan D. Putriana Nuramanah Kinding, "Improving Tesseract OCR Accuracy Using SymSpell Algorithm on Passport Data," *Sinkron*, vol. 9, no. 1, hlm. 374–381, Jan 2025, doi: 10.33395/sinkron.v9i1.14395.
- [6] F. S. P. Azzahra dan C. Prianto, "Tinjauan Literatur Sistematis: Analisis Implementasi Kecerdasan Buatan untuk Verifikasi Dokumen," *Jurnal Ilmiah Sistem Informasi*, vol. 4, no. 3, hlm. 417–430, 2025.
- [7] S. Faizullah, M. S. Ayub, S. Hussain, dan M. A. Khan, "A Survey of OCR in Arabic Language: Applications, Techniques, and Challenges," *Applied Sciences*, vol. 13, no. 7, hlm. 4584, Apr 2023, doi: 10.3390/app13074584.
- [8] M. V. Zufar dan N. Pratiwi, "Implementasi YOLOv11 dan Google ML Kit untuk Pembacaan Struk pada Aplikasi Keuangan Mobile," *METIK JURNAL (AKREDITASI SINTA 3)*, vol. 9, no. 2, hlm. 302–311, 2025.

- [9] M. Lihawa, A. D. Hartanto, N. Norhikmah, D. Prabowo, I. N. Fajri, dan W. Widayani, "Implementation of the Levenshtein Distance Algorithm and the Regular Search Expression Method for Detecting Typors in Javascript," *Sistemasi: Jurnal Sistem Informasi*, vol. 12, no. 2, hlm. 452–464, 2023.
- [10] M. G. Pradana, H. B. Seta, N. Irzavika, P. H. Saputro, dan R. Rusiyono, "Levenshtein Distance Algorithm in Javanese Character Translation Machine Based on Optical Character Recognition," *JOIV: International Journal on Informatics Visualization*, vol. 9, no. 4, hlm. 1411–1418, 2025.
- [11] T. M. E. Saputra, A. Susanto, dan B. J. Carmelita, "Implementation of tesseract ocr and bounding box for text extraction on food nutrition labels," *Building of Informatics, Technology and Science (BITS)*, vol. 6, no. 3, hlm. 1403–1412, 2024.
- [12] A. Z. Ulhaq, A. Z. Adilukito, S. M. P. G. Neru, M. D. Agisfio, dan others, "Aplikasi Pencatatan Kalori Harian Berbasis Android Dengan Arsitektur MVVM," *Computer Science (CO-SCIENCE)*, vol. 5, no. 1, hlm. 26–34, 2025.
- [13] A. N. Pratama dan F. Ardiani, "Optimization of model-view-viewmodel (mvvm) architecture pattern and restfull api on android-based e-learning application," *Int J Comput Appl*, vol. 185, 2023.
- [14] K. Lazaros, A. G. Vrahatis, dan S. Kotsiantis, "Human-in-the-Loop Artificial Intelligence: A Systematic Review of Concepts, Methods, and Applications," *Entropy*, vol. 28, no. 4, hlm. 377, Mar 2026, doi: 10.3390/e28040377.
- [15] Z. Yang dkk., "ChatDiet: Empowering personalized nutrition-oriented food recommender chatbots through an LLM-augmented framework," *Smart Health*, vol. 32, hlm. 100465, 2024.
- [16] X. Deng dan W. Tu, "NutriRec: Multimodal Large Language Models Enhanced Personalized Dietary Recommendation," *IEEE Access*, 2026.
- [17] P. Zhang dkk., "Adaptive Constraint Relaxation in Personalized Nutrition Recommendations: An LLM-Driven Knowledge Graph Retrieval Approach," dalam *AMIA Annual Symposium Proceedings*, 2025, hlm. 1529