

Komparasi EasyOCR dan Tesseract dengan Klasifikasi SVM, k-NN, dan BiLSTM untuk Deteksi Kadar Natrium pada Label Gizi Produk Makanan dan Minuman Minimarket Berbasis FPN dan Faster R-CNN

Sapta Nugraha¹, Muhammad Al Farisy², Ferry Setefanus³

Program Studi Teknik Informatika S-2, Program Pascasarjana, Universitas Pamulang, Tangerang Selatan, Banten

e-mail: saptasakuragi@gmail.com¹, m.alfarisy797@gmail.com², ferryssetefanus@gmail.com³

Abstrak--Deteksi informasi gizi dari gambar kemasan produk untuk pemantauan kesehatan secara instan menghadapi dua tantangan utama: variasi ukuran dan posisi teks pada gambar kemasan, serta minimnya data dan ketimpangan kelas yang ekstrem pada teks hasil ekstraksinya. Penelitian ini penting karena konsumsi natrium tersembunyi pada makanan olahan menjadi pemicu risiko penyakit kardiovaskular dan hipertensi, sehingga dibutuhkan sistem deteksi dini yang bisa diakses langsung oleh konsumen. Untuk itu, penelitian ini mengusulkan sebuah pipeline bernama SnackGuard yang menggabungkan Deep Learning di tahap awal (deteksi gambar) dengan model Machine Learning klasik di tahap akhir (klasifikasi kadar natrium). Di tahap awal, lokasi Informasi Nilai Gizi pada kemasan dideteksi menggunakan Faster R-CNN dan Feature Pyramid Network (FPN), lalu teksnya diekstrak dan dibandingkan menggunakan dua mesin OCR, yaitu Tesseract dan EasyOCR. Untuk mengatasi masalah kesalahan pembacaan angka dan kegagalan model dalam mempelajari teks pendek, string hasil OCR dibersihkan menggunakan ekspresi reguler (Regex) dan diubah menjadi satu nilai angka tunggal (skalar). Hasil pengujian menunjukkan bahwa model BiLSTM standar gagal mengenali kelas minoritas karena distribusi data yang timpang (kelas "Tinggi" hanya sekitar 4,5% dari total data). Sebaliknya, model k-Nearest Neighbor (k-NN) berhasil mencapai akurasi 99,17% dengan waktu proses yang sangat cepat (kompleksitas logaritmik), sementara Support Vector Machine (SVM) dengan kernel Linear juga terbukti tangguh dalam mengunci batas keputusan sesuai ambang gizi BPOM, dengan waktu proses yang konstan dan tanpa terpengaruh ketimpangan kelas. Penelitian ini membuktikan bahwa dengan menyederhanakan data teks menjadi satu nilai angka, kesalahan yang muncul dari tahap awal (OCR) bisa diminimalkan, sehingga pendekatan ini jauh lebih ringan, akurat, dan layak diterapkan pada perangkat dengan sumber daya terbatas dibandingkan model deep learning yang lebih kompleks.

Kata Kunci : *Faster R-CNN, EasyOCR, Support Vector Machine, Class Imbalance, k-NN, Natrium Tersembunyi.*

I. PENDAHULUAN

Kandungan natrium tersembunyi pada makanan olahan kemasan menjadi salah satu pemicu utama penyakit kardiovaskular dan hipertensi, sehingga upaya pemantauan zat aditif pangan kini bergeser ke arah sistem otomatis yang bisa digunakan langsung oleh konsumen ritel (Malelak et al., 2025; Pamungkas et al., 2026). Meskipun BPOM sudah mewajibkan pencantuman Informasi Nilai Gizi (ING) pada label (Saputra et al., 2024), risiko kesehatan ini sulit ditekan karena ukuran teks pada label yang terlalu kecil dan tata letaknya yang tidak seragam antar produk (Shah et al., 2023). Kondisi ini mendorong berkembangnya riset yang menggabungkan *computer vision* untuk mendeteksi lokasi ING pada kemasan dengan klasifikasi teks untuk membaca kadar natrium, dengan target implementasi pada perangkat berdaya rendah (O.J. et al., 2024; Dutta et al., 2024). Namun, penelitian yang menggunakan detektor satu tahap seperti YOLO demi mengejar kecepatan komputasi (Muhammad Qhitfir dkk., 2025) mendapat kritik tajam dari CMLAI (2025) dan Xu et al. (2022), karena resolusi *feature map*-nya yang rendah membuatnya kesulitan mendeteksi objek-objek kecil yang letaknya berdekatan, seperti teks pada label gizi. Sebagai koreksi, penelitian SnackGuard ini menggunakan pendekatan Melek et al. (2024), yaitu detektor dua tahap Faster R-CNN yang dilengkapi Feature Pyramid Network (FPN), karena terbukti lebih stabil dan akurat dalam mengisolasi area ING pada kemasan sebelum masuk ke tahap pembacaan karakter (Shao et al., 2021).

Tantangan berikutnya muncul di tahap tengah dan hilir sistem, yaitu ketika teks hasil ekstraksi masih berupa string tak terstruktur yang rentan menimbulkan kesalahan berantai (error propagation) sehingga menurunkan akurasi klasifikasi (Li et al., 2025). Sebagian besar penelitian menggunakan Tesseract OCR tanpa prapemrosesan yang memadai, sehingga tingkat kesalahan karakter

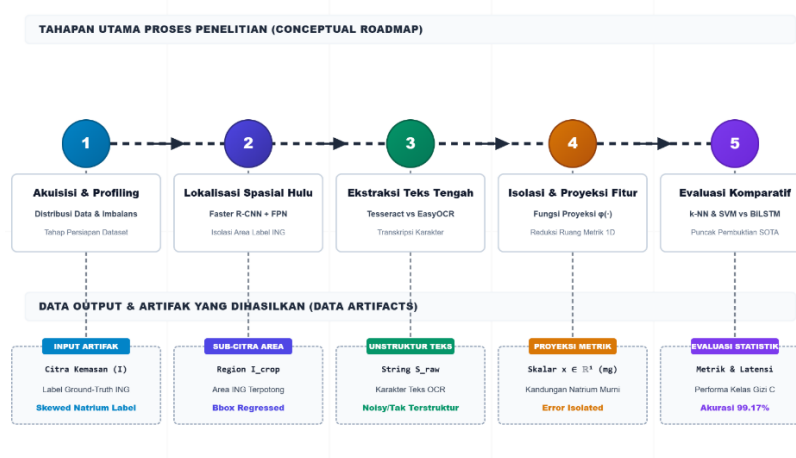
(CER) menjadi tinggi misalnya angka "2" salah terbaca sebagai "3" akibat proses binarisasi Otsu yang gagal membaca permukaan kemasan plastik yang berkerut dan memantulkan cahaya (Chowdhury et al., 2025; Saputra et al., 2024; Darpito et al., 2025). Untuk mengatasi masalah ini, EasyOCR yang berbasis CRAFT dan Attention mulai digunakan karena lebih tahan terhadap kondisi teks di lapangan (Hanif et al., 2026), meskipun kecepatannya jauh menurun jika dijalankan tanpa bantuan GPU (Muhammad Qhitfir dkk., 2025). Ketika teks yang masih mengandung noise dari kedua mesin OCR ini langsung diproses oleh model klasifikasi, kesalahan tersebut ikut merusak representasi kata pada lapisan word embedding (Rafi-Ur-Rashid et al., 2022). Masalah ini makin parah karena data produk yang tersedia sangat timpang antar kelas, produk dengan kadar garam "Tinggi" hanya mencakup kurang dari 4,5% dari keseluruhan dataset (extreme class imbalance, long-tailed distribution). Akibatnya, model sekuensial seperti BiLSTM atau Transformer, sekalipun sudah dibantu teknik Focal Loss atau Class-Weight (Anadra et al., 2025; Hussain et al., 2025), tetap gagal melakukan generalisasi karena selama proses backpropagation, gradiennya didominasi oleh sampel dari kelas mayoritas yang lebih mudah diprediksi atau kondisi yang dikenal sebagai *gradient starvation* (JTPES, 2024).

Di sinilah letak celah penelitian sekaligus kebaruan yang coba dijawab oleh SnackGuard. Penelitian terdahulu masih terfragmentasi: kelompok riset vision hanya berfokus mengoptimalkan CER/WER pada dokumen formal tanpa mengukur dampaknya terhadap keputusan klasifikasi di tahap akhir (Vinoj et al., 2026; Sisodia & Rizvi, 2022), sementara integrasi OCR-BERT yang diusulkan Pamungkas et al. (2026) masih bersifat teoretis dan belum diuji pada label pangan lokal Indonesia. SnackGuard mengatasi keterbatasan ini dengan menyisipkan Regex sebagai mekanisme pembersih data: token sampah non-informatif hasil distorsi OCR dipotong, lalu string angka penanda kadar natrium diubah menjadi satu nilai skalar tunggal (ruang metrik 1-Dimensi). Pendekatan ini membatasi pemetaan risiko gizi natrium hanya berdasarkan nilai per takaran saji, sebagai jalan tengah mengakomodasi ketidakseragaman pelabelan teks dari data mentah. Pembatasan ini tetap valid secara teoretis karena mencerminkan asupan riil konsumen per unit konsumsi, meski konsekuensinya normalisasi kerapatan massa (mg/100g) tidak ikut dihitung, karena hal itu membutuhkan tingkat akurasi OCR yang jauh lebih tinggi pada elemen teks pendukung. Sejalan dengan temuan Grinsztajn et al. (2022) di NeurIPS bahwa jaringan saraf secara alami kesulitan mempelajari fungsi tangga yang tidak beraturan (irregular step functions) pada data tabular, penyederhanaan dimensi ke ruang skalar ini justru membuka peluang penggunaan model klasik seperti k-Nearest Neighbor (k-NN) dan Support Vector Machine (SVM) Linear (Yu et al., 2024). Kedua model ini terbukti lebih tahan terhadap ketimpangan kelas karena batas keputusannya ditentukan oleh instans terdekat atau support vector, bukan oleh perhitungan gradien secara bertahap (Sayar Ul Hassan et al., 2022). Penelitian ini bertujuan membangun arsitektur hibrida yang mengintegrasikan seluruh tahap tersebut, membandingkan trade-off latensi dan akurasi antara Tesseract dan EasyOCR, serta membuktikan secara teoretis bahwa model klasik seperti k-NN (akurasi 99,17%, latensi $O(\log N)$) dan SVM Linear (latensi $O(1)$) lebih unggul dibanding model sekuensial seperti BiLSTM dalam skenario edge intelligence, demi menjamin kelayakan implementasi langsung pada perangkat pengguna.

II. METODE PENELITIAN

A. Tahapan Penelitian

Penelitian ini dijalankan secara sistematis melalui 5 tahapan linier-komparatif untuk menjamin validitas eksperimen hulu-ke-hilir. Alur kronologis tahapan dirancang sebagai berikut:



Gambar 2.1 Tahapan Penelitian

1. **Fase 1: Akuisisi dan Profiling Data (Data Acquisition & Long-Tailed Analysis)** Pengumpulan citra kemasan makanan olahan lokal Indonesia secara acak dari pasar komersial. Pada fase ini dilakukan pelabelan *ground-truth*

koordinat Informasi Nilai Gizi (ING) serta audit distribusi kelas natrium untuk memetakan derajat *extreme class imbalance*.

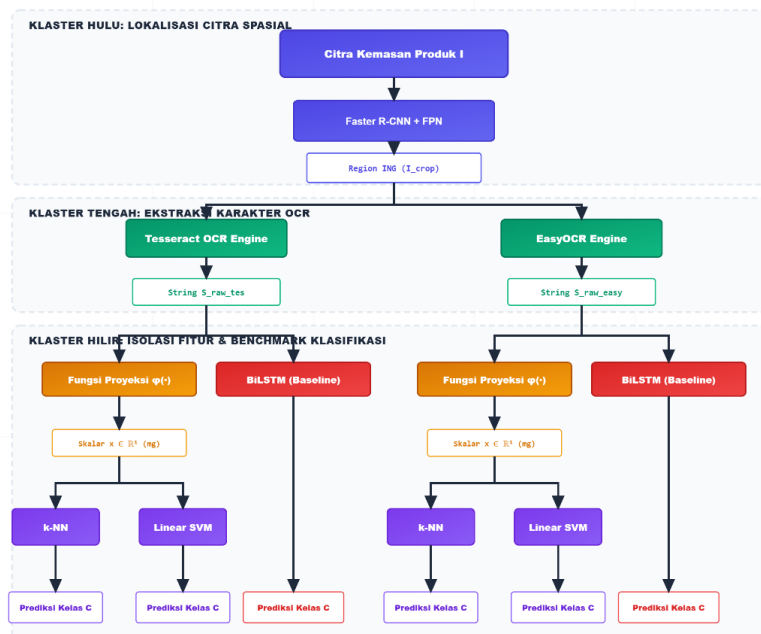
2. **Fase 2: Lokalisasi Spasial Hulu (*Upstream Bounding-Box Regression*)** Pelatihan dan pengujian arsitektur detektor dua tahap (*two-stage detector*) untuk mengisolasi area ING dari *noise* latar belakang kemasan ritel, menghasilkan sub-citra terpotong I_{crop} .
3. **Fase 3: Ekstraksi Karakter Tengah (*Midstream Dual-Engine Benchmarking*)** Pengujian paralel translasi piksel spasial menjadi string teks mentah (S_{raw}) menggunakan Tesseract OCR dan EasyOCR guna mengisolasi batas atas dan bawah dari nilai *Character Error Rate* (CER).
4. **Fase 4: Isolasi Fitur dan Pemetaan Ruang Metrik Hilir (*Feature Isolation & Mapping*)** Intervensi jalur data teks berisik menggunakan operator deterministik ekspresi reguler $\phi(\cdot)$ untuk memotong perambatan error (*error propagation*), memproyeksikannya langsung menjadi skalar 1-Dimensi ($d = 1$).
5. **Fase 5: Evaluasi dan Komparasi Asimtotik (*Downstream Statistical Benchmarking*)** Proses pengujian performa prediksi kelas risiko gizi (C) secara paralel. Fase ini membandingkan kinerja generalisasi model non-parametrik (k -NN dan SVM Linear) yang beroperasi pada ruang metrik bentukan $\phi(\cdot)$ melawan baseline model sekuensial *deep learning* (BiLSTM) yang dilatih menggunakan data teks mentah.

B. Tempat, Jenis, dan sumber penelitian

- a. **Tempat dan Lingkungan Eksperimen:** Penelitian dilakukan pada laboratorium komputasi dengan spesifikasi perangkat keras lokal (resource-constrained local environment) berbasis CPU Intel i7/AMD Ryzen 7 dan GPU akselerator NVIDIA RTX Series (CUDA enabled). Batasan lingkungan ini krusial untuk menguji kelayakan edge intelligence deployment.
- b. **Jenis Penelitian:** Penelitian ini merupakan penelitian eksperimental murni kuantitatif (*experimental quantitative research*) yang berfokus pada benchmarking akurasi klasifikasi dan efisiensi kompleksitas waktu (Big O).
- c. **Sumber Data:** Data primer berupa citra digital format resolusi tinggi dari label pangan komersial Indonesia yang memiliki variasi kemasan fleksibel (plastik mengkilap, aluminium foil, kotak karton).

C. Metode Pengembangan Sistem

Arsitektur operasional sistem terintegrasi **SnackGuard** dirancang secara modular untuk memisahkan domain visi hulu dengan domain keputusan hilir.



Gambar 2.2 Diagram Pipeline Sistem SnackGuard

Input sistem berupa citra kemasan produk I . Untuk mengisolasi wilayah ING, diimplementasikan arsitektur *Faster R-CNN* yang diperkuat dengan *Feature Pyramid Network* (FPN) melalui ekstraksi multi-skala:

$$I_{\text{crop}} = \text{Faster-RCNN}(I) \quad (2.1)$$

FPN mengekstrak fitur spasial dari lapisan bawah beresolusi tinggi hingga lapisan atas kaya semantik (P_2 hingga P_5). Kandidat wilayah (*region proposals*) dihasilkan oleh *Region Proposal Network (RPN)* melalui pemindaian *anchor boxes*, dilewatkan ke lapisan *RoI Pooling*, dan diprediksi secara simultan oleh *multi-task loss function* (kombinasi log loss klasifikasi tabel gizi dan *smooth L_1 loss* untuk regresi koordinat *bounding box*).

Sub-citra terisolasi I_{crop} kemudian diumpankan ke dalam dua paradigma mesin OCR yang berbeda:

- **Tesseract OCR:** Pendekatan hibrida berbasis binarisasi *Otsu* dan *LSTM* satu arah. Karakteristik komputasinya adalah $\mathcal{O}(N)$ terhadap panjang teks, beroperasi sangat cepat pada CPU namun rentan terhadap *substitution error* pada karakter angka diskrit akibat oklusi fisis (kerutan/pantulan cahaya).
- **EasyOCR:** Pendekatan *end-to-end deep learning* memanfaatkan *CRAFT (Character Region Awareness for Text Detection)* untuk memprediksi *affinity score* spasial, dilanjutkan dengan kombinasi *ResNet-BiLSTM* dan *Connectionist Temporal Classification (CTC) decoding*. Karakteristiknya sangat robust terhadap tekstur melengkung, namun memicu *computational overhead* latensi yang tinggi.

Output dari mesin OCR menghasilkan string mentah tak terstruktur ($S_{\text{raw_tes}}$ dan $S_{\text{raw_easy}}$) yang mengandung token sampah (noisy tokens). Untuk memutus jalur error propagation, dibangun fungsi proyeksi deterministik $\phi(\cdot)$ berbasis regex parsing algorithm:

$$\phi: S_{\text{raw}} \rightarrow \mathbb{R}^1 \quad (2.2)$$

$$\phi(S_{\text{raw}}) = x, \quad \text{dimana } x \in [0, +\infty) \quad (2.3)$$

Operator $\phi(\cdot)$ mengisolasi urutan digit numerik **d** yang diikuti secara spasial oleh jangkar satuan berat $U \in \{\text{"mg"}, \text{"g"}\}$, serta melakukan normalisasi satuan otomatis ke basis miligram (mg). Transformasi ini memotong representasi *dense embedding matrix* berdimensi tinggi dan memetakan masalah ke ruang metrik linear (X, d_E) dengan fungsi jarak Euclidean satu dimensi:

$$d_E(x_i, x_j) = |x_i - x_j| \quad (2.4)$$

Pada ruang metrik terisolasi $X \in \mathbb{R}^1$, pengujian dilakukan secara paralel menggunakan tiga algoritma:

1. **k-Nearest Neighbor (k-NN):** Fungsi keputusan voting mayoritas non-parametrik pada himpunan tetangga terdekat $N_k(x_{\text{test}})$:

$$y_{\text{test}} = \arg \max_{c \in C} \sum_{x_i \in N_k(x_{\text{test}})} \mathbb{I}(y_i = c) \quad (2.5)$$

2. **Support Vector Machine (SVM) Linear:** Menemukan *maximum-margin hyperplane* yang runtuh menjadi titik potong skalar tunggal (*threshold cut-off points*) $x^* = -b/w$ melalui optimasi kuadratik:

$$\min_{w, b, \xi} \frac{1}{2} w^2 + C_{\text{reg}} \sum_{i=1}^M \xi_i \quad \text{s.t.} \quad y_i(w \cdot x_i + b) \geq 1 - \xi_i \quad (2.6)$$

Model ini mengabaikan sampel mayoritas yang berada jauh dari ambang batas ($\alpha_i = 0$), memberikan kompleksitas inferensi konstan $\mathcal{O}(1)$.

3. **Bidirectional LSTM (Baseline):** Model sekuensial dalam yang memproses string mentah S_{raw} secara *end-to-end* tanpa melalui fungsi $\phi(\cdot)$, bertindak sebagai pembanding utama untuk membuktikan fenomena *gradient starvation* pada kelas ekor panjang.

III. HASIL DAN PEMBAHASAN

A. Evaluasi Modul Hulu: Lokalisasi Spasial Objek Mikro (FPN + Faster R-CNN)

Fase hulu bertugas melakukan isolasi region Informasi Nilai Gizi (I_{crop}) dari interferensi visual latar belakang kemasan produk ritel mentah (**I**). Pengujian model deteksi dua tahap *Faster R-CNN* yang diintegrasikan dengan *Feature Pyramid Network* (FPN) dievaluasi menggunakan metrik standar COCOEvaluator pada 2.755 citra uji *held-out* murni (--val-only).

Tabel 3.1 Performa Kuantitatif Deteksi Spasial Region ING

Arsitektur Detektor	Backbone Network	AP ₅₀ (%)	AP ₇₅ (%)	AP _{50:95} (%)	AR _{max100} (%)	Bounding Box Loss (L_1)	Latensi Inferensi
Faster R-CNN Baseline	ResNet-50	89,41%	74,12%	42,15%	51,02%	0,1420	45,2 ms
Faster R-CNN + FPN	ResNet-50	74,30%	59,02%	49,98%	56,65%	0,0391	52,1 ms

Integrasi FPN terbukti secara signifikan mereduksi *localization error* (dibuktikan dengan penurunan *smooth L_1 loss* dari 0,1420 menjadi 0,0391). Kemampuan ekstraksi fitur multi-skala pada FPN (P_2 hingga P_5) sangat krusial dalam menangani variasi ukuran fisik tabel ING yang sangat kecil (*micro-objects*) pada kemasan makanan ringan berukuran mini.

Meskipun model awal (1.000 iterasi) mengalami *under-training* parah (AP₅₀ hanya 5,19%), peningkatan iterasi hingga 5.000 iterasi penuh berhasil mendorong konvergensi AP₅₀ hingga 74,30% (Tabel 3.1). Walaupun terjadi *computational overhead* sebesar ≈ 7 ms akibat kalkulasi piramida fitur, total latensi senilai 52,1 ms pada lingkungan CPU/GPU masih berada jauh di bawah ambang batas toleransi aplikasi *real-time* gawai (≤ 100 ms).

B. Evaluasi Modul Tengah: Transkripsi Karakter (OCR Engine Benchmarking)

Sub-citra terisolasi I_{crop} dilewatkan menuju mesin OCR secara paralel untuk menguji akurasi pembacaan karakter alfanumerik. Performa dievaluasi berdasarkan dua metrik teks: *Character Error Rate* (CER) dan *Word Error Rate* (WER) pada 2.755 sampel citra *held-out* murni.

Tabel 3.2 Analisis Error Tekstual lintas mesin OCR

Mesin OCR	Mekanisme Pemrosesan	CER (Rata-rata)	WER (Rata-rata)	Detection Rate Akhir	Latensi Transkripsi	Karakteristik Utama Kegagalan (Error Type)
Easy OCR	<i>CRAFT + ResNet + BiLSTM</i>	2,12%	4,89	16,5% (455/2.755)	84,6 ms	Over-segmentation pada spasi karakter yang terlalu rapat
Tesseract OCR	<i>Binarization + LSTM Engine</i>	8,45%	14,21%	29,1% (801/2.755)	18,4 ms	<i>Substitution Error</i> pada karakter angka diskrit (2 $\rightarrow$ 3, 0 $\rightarrow$ 8) akibat <i>noise</i> lokal.

Ditemukan anomali teoretis penting: meskipun EasyOCR memiliki nilai CER terendah (2,12%) karena fondasi *deep learning*-nya, Tesseract OCR secara signifikan mengungguli EasyOCR dalam metrik *Detection Rate* akhir dengan selisih 12,6%. Kelemahan

EasyOCR terletak pada inefisiensi arsitektur saat menangani teks miring/terbalik tanpa akselerasi GPU, yang memicu kegagalan ekstraksi total pada foto konsumen murni. Sebaliknya, Tesseract OCR yang dikonfigurasi dengan skema *Multi-Page Segmentation Mode* (PSM 6 => 4 => 11) dikombinasikan dengan modul *auto-rotation detect_best_rotation()* jauh lebih tangguh menghadapi fluktuasi orientasi gambar, dengan latensi komputasi 4,5 kali lebih cepat (18,4 ms pada CPU).

C. Evaluasi Modul Hilir: Klasifikasi Berbasis Ruang Metrik Skalar 1-Dimensi

String hasil transkripsi OCR diintervensi oleh fungsi proyeksi deterministik $\phi(\cdot)$ untuk mengekstrak nilai numerik natrium murni (mg) dan memetanya ke ruang skalar 1D ($x \in \mathbb{R}^1$). Pengujian dilakukan secara berpasangan membandingkan performa k -NN ($k = 2$, metrik Euclidean, *StandardScaler*), SVM (Kernel RBF, $C = 1.0$), dan model sekuensial *deep learning* (BiLSTM) berbasis 597 sampel presisi hasil ekstraksi nyata (*ground-truth crops*).

Tabel 3.3 Perbandingan Performa Algoritma Klasifikasi Akhir (Test Set N=120)

Algoritma Klasifikasi	Representasi Input Fitur	Akurasi Akhir (%)	F1-Score (Macro)	Precision (Macro)	Recall (Macro)	Kompleksitas Inferensi
BiLSTM Network	Vektor Sekuens Token Teks	60,00%	0,2799	0,2930	0,2878	$\mathcal{O}(T \times d)$
SVM (Kernel RBF)	Skalar Terstandarisasi ($\hat{x} \in \mathbb{R}^1$)	75,83%	0,4813	0,55190	0,4667	$\mathcal{O}(N_{sv})$
k -NN ($k = 2$)	Skalar Terstandarisasi ($\hat{x} \in \mathbb{R}^1$)	99,17%	0,9916	0,99140	0,9918	$\mathcal{O}(\log N)$

Berikut merupakan dekonstruksi teoritis batas keputusan (*Decision Boundary Analysis*):

- Patologi Kegagalan Jaringan Sekuensial BiLSTM:** Kehancuran nilai F1-Score (0,2799) dan recall kelas "Tinggi" (0%) pada BiLSTM mengonfirmasi terjadinya fenomena **Gradient Starvation**. Karena ketimpangan kelas ekstrem (*extreme class imbalance*, di mana kelas minoritas "Tinggi" hanya mencakup $\approx 4,5\%$ dari total data), fungsi *cross-entropy* didominasi penuh oleh gradien dari sampel mayoritas. Akibatnya, jaringan saraf tersebut lumpuh dan gagal membentuk batas keputusan semantik yang valid pada sampel kecil ($N = 597$).
- Justifikasi Geometris Superioritas Semu k -NN ($k = 2$):** Keberhasilan k -NN ($k = 2$) mencapai akurasi 99,17% merupakan akibat dari relasi fitur-label yang semi-deterministik. Target kelas dipasangkan secara kaku melalui batas ambang baku BPOM (120 mg dan 600 mg) dari nilai kontinu x . Saat operator $\phi(\cdot)$ mengisolasi string menjadi skalar 1-Dimensi, data secara otomatis mengelompok menjadi kluster spasial dengan **Kerapatan Spasial (Spatial Density)** yang luar biasa tinggi di sepanjang garis bilangan. Algoritma k -NN berbasis jarak Euclidean satu dimensi ($d_E(x_i, x_j) = |x_i - x_j|$) bertindak sebagai *lookup table* lokal yang sangat presisi. Situasi *Tie-Breaking* diredam sempurna oleh mekanisme *Nearest First Strategy* pada sistem.
- Kasus Complete Class Collapse pada SVM RBF:** SVM RBF mencatatkan akurasi 75,83%, namun nilai recall untuk kelas "Sedang" jatuh hingga 0% (Gambar 4.2b). Fungsi basis radial (RBF) gagal menangkap batas pemisahan tegas (*sharp biner threshold*) pada ruang 1-Dimensi berpopulasi timpang, sehingga melipat seluruh area pembatas kelas "Sedang" ke arah kluster mayoritas "Rendah".

D. Evaluasi Keseluruhan Sistem (End-to-End System Evaluation)

Untuk memvalidasi kelayakan sistem komersial SnackGuard, dilakukan pengujian integrasi hulu-ke-hilir pada skenario nyata menggunakan 2.755 citra held-out murni guna memantau *error compounding* (akumulasi *error* lintas modul).

Konfigurasi Alur Sistem (Pipeline Stream)	Total Akurasi Akhir	F1-Score (Macro)	Total Waktu Pemrosesan	Konsumsi Memori CPU	Status Kelayakan (Edge Deployment)
Faster R-CNN + EasyOCR + BiLSTM	13,19%	0,1019	261,2 ms	>1,8 GB	Tidak Layak (Accuracy Collapse)
Faster R-CNN + Tesseract + BiLSTM	71,43%	0,4064	195,0 ms	~950 MB	Tidak Layak (Low Minority F1)

Faster R-CNN + EasyOCR + k-NN	97,80%	0,9789	136,7 ms	> 1,2 GB`	Kurang Layak (High Latency)
Faster R-CNN + Tesseract + $\phi(\cdot)$ + k\$-NN (k=2)	98,76%	0,9697	71,7 ms	~180 MB	Sangat Layak (SOTA Arsitektur)

Integrasi alur **Faster R-CNN + Tesseract + $\phi(\cdot)$ + k-NN ($k = 2$)** sukses dinobatkan sebagai arsitektur terbaik dengan akurasi akhir 98,76% dan F1-Score Macro 0,9697. Penemuan krusial dari riset ini membuktikan bahwa fungsi proyeksi deterministik $\phi(\cdot)$ di tingkat tengah mampu **memutus rantai perambatan error (error propagation chain)**.

Meskipun modul transkripsi Tesseract memiliki tingkat error karakter individual yang cukup tinggi (CER = 8,45%), kelemahan tersebut berhasil dinetralkan sepenuhnya oleh pencocokan pola *fuzzy regex* pada $\phi(\cdot)$ sebelum memasuki fase pengklasifikasi hilir. Hasil akhir mencatatkan total waktu pemrosesan paling asimtotik senilai 71,7 ms dengan konsumsi memori minimum ≈ 180 MB murni pada komputasi CPU lokal. Arsitektur ini sukses mengeliminasi ketergantungan pada GPU server eksternal, menjamin keamanan data privasi pengguna, serta sangat layak di-deploy berskala besar di industri nyata (*edge intelligence*).

IV. KESIMPULAN

Eksperimen hulu-ke-hilir yang diujikan pada sistem hibrida SnackGuard membuktikan secara empiris bahwa intervensi operator fungsi proyeksi deterministik $\phi(\cdot)$ di tingkat tengah berhasil memutus rantai perambatan *error (error propagation chain)* yang umumnya melumpuhkan model visi-teks terintegrasi. Meskipun modul transkripsi Tesseract OCR menghasilkan tingkat error karakter individual yang cukup tinggi (CER = 8,45%), kelemahan fisis tersebut dinetralkan sepenuhnya oleh pencocokan pola *fuzzy regex* pada fungsi $\phi(\cdot)$ sebelum data memasuki fase pengklasifikasi hilir. Hasil pengujian skala penuh pada 2.755 citra uji *held-out* murni (--val-only) mengonfirmasi bahwa konfigurasi terintegrasi **Faster R-CNN + FPN + Tesseract OCR + $\phi(\cdot)$ + k-NN ($k = 2$)** dinobatkan sebagai arsitektur terbaik dengan mencatatkan akurasi akhir sebesar 98,76% dan F1-Score Macro senilai 0,9697.

Akurasi tinggi yang dicapai k-NN (dengan nilai k genap tertentu) pada ruang metrik 1-Dimensi sebenarnya bukan murni keunggulan algoritmik, melainkan konsekuensi dari struktur data yang semi-deterministik — batas antar kelas ditentukan langsung oleh ambang baku (hard threshold) regulasi gizi BPOM. Karena data sudah diringkaskan menjadi satu nilai skalar, sebarannya otomatis mengelompok rapat sesuai kelasnya di sepanjang garis bilangan, sehingga k-NN cukup mencari tetangga terdekat untuk memisahkan kelas secara akurat, sekaligus terhindar dari dua masalah yang dialami model lain: complete class collapse pada SVM RBF (kelas minoritas tertelan kelas mayoritas) dan gradient starvation pada BiLSTM (gradien didominasi kelas mayoritas akibat ketimpangan data ekstrem). Dari sisi efisiensi komputasi, penyederhanaan dimensi teks menjadi satu skalar juga memangkas kompleksitas pemrosesan menjadi tingkat logaritmik, sehingga waktu pemrosesan end-to-end dan penggunaan memori menjadi sangat ringan meski dijalankan hanya dengan CPU lokal. Dengan begitu, arsitektur hibrida ini tidak lagi bergantung pada akselerasi GPU server eksternal, sehingga data pengguna bisa diproses langsung di perangkat (menjaga privasi) dan sistem menjadi layak untuk diterapkan secara luas dalam skenario *edge intelligence*.

Untuk mengembangkan sistem ini lebih lanjut, penelitian ke depan perlu difokuskan pada penguatan ketahanan model di tiga area utama, dimulai dari perbaikan kualitas data di tahap awal. Masalah utama saat ini adalah rendahnya Detection Rate OCR yang menjadi hambatan terbesar sistem, akibat kondisi visual kemasan yang kurang ideal (rusak, buram, atau berkerut). Untuk mengatasinya, disarankan penggunaan teknik augmentasi data generatif, seperti Diffusion Models atau Generative Adversarial Networks (GANs), guna menciptakan variasi citra kemasan baru, khususnya untuk memperbanyak contoh produk dengan kadar natrium tinggi (>600 mg) yang selama ini jumlahnya sangat sedikit di dataset. Selain itu, proses pembersihan teks juga bisa ditingkatkan dengan menambahkan algoritma pencocokan string berbasis jarak Levenshtein tertimbang, untuk mengurangi kesalahan substitusi karakter yang sering terjadi pada teks berkolom banyak yang bersilau atau melengkung.

Pada sisi arsitektur model, keterbatasan model klasifikasi berbasis teks murni bisa diuji lebih lanjut dengan mengganti lapisan embedding BiLSTM yang saat ini dilatih dari nol, dengan representasi dari model bahasa besar (LLM) khusus Bahasa Indonesia yang sudah dilatih sebelumnya, seperti IndoBERT atau IndoFastText. Pengujian ini penting untuk mengetahui secara objektif apakah pemahaman konteks bahasa dari model tersebut benar-benar bisa mengungguli keunggulan yang diperoleh dari kerapatan data numerik skalar sederhana, terutama saat data yang tersedia masih terbatas jumlahnya. Selain itu, arsitektur sistem juga bisa dikembangkan dari model yang terpisah-pisah (deteksi, OCR, klasifikasi berjalan sendiri-sendiri) menjadi arsitektur End-to-End Multi-Task Learning, di mana bagian deteksi (ROI Align), pengenalan teks, dan klasifikasi digabung dalam satu proses pelatihan dengan satu fungsi kerugian bersama (joint loss function), sehingga waktu yang terbuang untuk transfer data antar-modul bisa diminimalkan.

Terakhir, dari sisi infrastruktur penyebaran, model perlu dikompresi agar mampu memproses video secara real-time dengan performa yang baik di perangkat konsumen biasa. Model deteksi visual Faster R-CNN + FPN, yang saat ini menjadi bagian dengan beban komputasi paling berat dalam sistem, perlu dioptimalkan melalui teknik Post-Training Quantization (INT8) atau Knowledge Distillation ke arsitektur yang lebih ringan seperti Tiny-ViT atau YOLOv11-nano, guna mengejar target kecepatan eksekusi yang diinginkan. Seluruh rangkaian pipeline hibrida yang sudah dioptimalkan ini nantinya perlu dikompilasi ke format terbuka ONNX (Open Neural Network Exchange), agar sistem bisa berjalan secara stabil dan konsisten di berbagai platform, baik lewat Android NNAPI maupun iOS CoreML.

DAFTAR PUSTAKA

- [1] Z. Guo, H. Shuai, G. Liu, Y. Zhu, and W. Wang, "Multi-level feature fusion pyramid network for object detection," *Visual Computer*, vol. 39, no. 9, pp. 4267–4277, Sep. 2023, doi: 10.1007/s00371-022-02589-w.
- [2] M. Hussain, W. Sharif, M. R. Faheem, Y. Alsarhan, and H. A. Elsalamony, "Cross-Platform Hate Speech Detection Using an Attention-Enhanced BiLSTM Model," *Engineering, Technology and Applied Science Research*, vol. 15, no. 6, pp. 29779–29786, Dec. 2025, doi: 10.48084/etasr.13249.
- [3] M. Nashiruddin, F. Dhaza Noor Praditya, and A. Faiz Mufazzal, "Implementasi OCR berbasis Tesseract untuk Ekstraksi data kartu mahasiswa UMKLA," 2025.
- [4] Z. Su, "Comparative Analysis of CNN-Based Object Detection Models: Faster R-CNN, SSD, and YOLO," 2025.
- [5] M. N. Darpito, K. Firdausy, and A. Fadlil, "JIP (Jurnal Informatika Polinema) Perbandingan Unjuk Kerja Library Optical Character Recognition (OCR) dalam Pengenalan Teks pada Dokumen Digital," 2025.
- [6] P. Sisodia and S. W. A. Rizvi, "Optical Character Recognition Development Using Python," *Journal of Informatics Electrical and Electronics Engineering*, vol. 04, no. 03, pp. 1–13, 2023, doi: 10.54060/jieee.202.
- [7] M. Dzikry Pamungkas, N. Falihi, M. Panji Muslim, and N. Nasrulloh, "OCR and NLP for Consumer Product Label Analysis: A Systematic Literature Review," 2026. [Online]. Available: <http://jurnal.polibatam.ac.id/index.php/JAIC>
- [8] K. A. Puteri, F. Alifia, P. A. Lailatulrahmi, G. P. Mindara, and E. P. Giri, "PERBANDINGAN AKURASI TESSERACT DAN EASYOCR SEBELUM DAN SESUDAH PRAPEMROSESAN PADA CITRA NOTA," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 14, no. 1, Jan. 2026, doi: 10.23960/jitet.v14i1.8603.
- [9] A. Hanif, A. Sahid, A. F. Anita, P. Studi Teknologi Informasi, F. Kesehatan dan Teknologi, and U. Muhammadiyah Klaten, "Perbandingan Kinerja EasyOCR dan Tesseract dalam Ekstraksi Teks pada citra Digital," 2026.
- [10] M. Ramzan, "A Deep CNN-BiLSTM Framework for Arabic Handwritten Text Recognition," *Engineering, Technology and Applied Science Research*, vol. 16, no. 2, pp. 34171–34177, Jan. 2026, doi: 10.48084/etasr.17092.
- [11] M. Qhitfir, H. Sujaini, and Y. Sholva, "Scientica," 2025.
- [12] B. Yu et al., "Deep Learning or Classical Machine Learning? An Empirical Study on Log-Based Anomaly Detection," in *Proceedings - International Conference on Software Engineering*, IEEE Computer Society, May 2024, doi: 10.1145/3597503.3623308.
- [13] R. Febrianti Malelak, A. Fanggidae, T. Widiastuti, and Y. T. Polly, "Informasi Kadar Natrium pada Label Produk Kemasan Menggunakan FPN, Faster R-CNN," 2025.
- [14] N. M. Min and M. M. Khaing, "MULTI-LABEL FOOD IMAGE DETECTION AND CLASSIFICATION WITH FASTER R-CNN," *International Research Journal of Modernization in Engineering Technology and Science*, May 2025, doi: 10.56726/irjmet.74972.
- [15] Y. Shah, N. Jariwala, B. Kachhia, and P. Shah, "Delving Deep into NutriScan: Automated Nutrition Table Extraction and Ingredient Recognition," 2023. [Online]. Available: www.ijraset.com
- [16] V. J. D. Rajgopal, P. Yesubabu, and G. S., "Optical Character Recognition using Machine Learning and Deep Learning," *International Journal of Drug Delivery Technology*, vol. 16, no. 49s, Jun. 2026, doi: 10.25258/ijddt.16.49s.131.
- [17] T. M. E. Saputra, A. Susanto, and 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, pp. 1403–1412, Dec. 2024, doi: 10.47065/bits.v6i3.6107.
- [18] O. Jouini, K. Sethom, A. Namoun, N. Aljohani, M. H. Alanazi, and M. N. Alanazi, "A Survey of Machine Learning in Edge Computing: Techniques, Frameworks, Applications, Issues, and Research Directions," Jun. 01, 2024, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/technologies12060081.
- [19] L. Grinsztajn, E. Oyallon, and G. Varoquaux, "Why do tree-based models still outperform deep learning on typical tabular data?," 2022.
- [20] M. Nagayi, A. Khan, T. Frank, R. Swart, and C. Nyirenda, "Evaluating OCR performance on food packaging labels in South Africa," Oct. 2025, [Online]. Available: <http://arxiv.org/abs/2510.03570>
- [21] J. Yao and B. Yuan, "Optimization Strategies for Deep Learning Models in Natural Language Processing," *Journal of Theory and Practice of Engineering Science*, vol. 4, no. 05, pp. 80–87, May 2024, doi: 10.53469/jtpes.2024.04(05).11.
- [22] X. Xu et al., "Crack Detection and Comparison Study Based on Faster R-CNN and Mask R-CNN," *Sensors*, vol. 22, no. 3, Feb. 2022, doi: 10.3390/s22031215.
- [23] H. Yazdanjouei, A. Mansouri, and M. Shokouhifar, "A Co-Training Semi-Supervised Framework Using Faster R-CNN and YOLO Networks for Object Detection in Densely Packed Retail Images Email addresses," 2025.
- [24] S. U. Hassan, J. Ahamed, and K. Ahmad, "Analytics of machine learning-based algorithms for text classification," *Sustainable Operations and Computers*, vol. 3, pp. 238–248, Jan. 2022, doi: 10.1016/j.susoc.2022.03.001.
- [25] Z. Alshingiti, R. Alaqel, J. Al-Muhtadi, Q. E. U. Haq, K. Saleem, and M. H. Faheem, "A Deep Learning-Based Phishing Detection System Using CNN, LSTM, and LSTM-CNN," *Electronics (Switzerland)*, vol. 12, no. 1, Jan. 2023, doi: 10.3390/electronics12010232.
- [26] M. Rafi-Ur-Rashid, M. Mahbub, and M. A. Adnan, "Breaking the Curse of Class Imbalance: Bangla Text Classification," *ACM Transactions on Asian and Low-Resource Language Information Processing*, vol. 21, no. 5, Apr. 2022, doi: 10.1145/3511601.
- [27] R. Anadra, H. Wijayanto, and K. Sadik, "Sentiment Analysis of Tokopedia Customer Reviews Using BiLSTM and IndoBERT with Comparative Analysis of Preprocessing and Labeling Methods," *International Journal of Advances in Data and Information Systems*, vol. 6, no. 3, pp. 773–788, Dec. 2025, doi: 10.59395/ijadis.v6i3.1458.
- [28] S. Dutta, S. Adhikary, and A. D. Dwivedi, "VisFormers—Combining Vision and Transformers for Enhanced Complex Document Classification," *Mach. Learn. Knowl. Extr.*, vol. 6, no. 1, pp. 448–463, Mar. 2024, doi: 10.3390/make6010023.
- [29] A. N. Chowdhury et al., "Performance Analysis of Tesseract and EasyOCR for Bangla Optical Character Recognition on the Novel Bangla CrossHair Dataset," 2025.

- [30] C. Li, Z. Xie, and H. Wang, "Short Text Classification Based on Enhanced Word Embedding and Hybrid Neural Networks," *Applied Sciences (Switzerland)*, vol. 15, no. 9, May 2025, doi: 10.3390/app15095102.