

Perancangan Sistem Klasifikasi Cacat Produk Berbasis Vision Language Model Pada Aplikasi Web untuk Proses Quality Control Tanpa Training Dataset

Mohammad Zacky Valentino Putra¹, Imelda Regina Siswi², Sahdila Adinda Putri³, Dicky Apdillah⁴

^{1,2,3,4}Program Studi Teknik Informatika, Universitas Asahan, Jl. Jend. Ahmad Yani, Kisaran, Sumatera Utara

E-mail: mohammadzackyvalentinoputra@gmail.com

* Corresponding Author

 <https://doi.org/10.31004/jerkin.v5i1.6598>

ARTICLE INFO

Article history

Received: 10 Juli 2026

Revised: 15 Juli 2026

Accepted: 21 Juli 2026

Kata Kunci

Vision Language Model; Quality Control; Klasifikasi Cacat Produk; GPT-4o Vision; Zero-Shot Learning; Aplikasi Web

Keywords

Vision Language Model; Quality Control; Product Defect Classification; GPT-4o Vision; Zero-Shot Learning; Web Application



ABSTRACT

Proses quality control (QC) pada industri manufaktur umumnya bergantung pada inspeksi visual manual yang rentan terhadap kelelahan dan subjektivitas manusia. Pendekatan berbasis deep learning konvensional memerlukan dataset berlabel dalam jumlah besar serta proses pelatihan model yang memakan waktu dan biaya. Penelitian ini merancang arsitektur sistem berbasis aplikasi web yang memanfaatkan Vision Language Model (VLM) siap pakai, khususnya GPT-4o Vision, untuk mengklasifikasikan cacat produk secara otomatis tanpa memerlukan proses pelatihan dataset. Metode yang digunakan adalah zero-shot visual reasoning melalui pendekatan zero-shot inference berbasis deskripsi tekstual. Sistem dirancang menggunakan three-tier architecture yang mencakup antarmuka pengguna berbasis web, lapisan logika bisnis, dan integrasi API VLM. Evaluasi kelayakan konseptual dilakukan melalui analisis komparatif, skenario use-case, dan matriks risiko sistem. Hasil perancangan menunjukkan bahwa pendekatan ini mampu mengurangi kebutuhan anotasi data, mempercepat deployment, dan menghasilkan output klasifikasi yang dapat dijelaskan. Penelitian ini memberikan kontribusi berupa blueprint perancangan sistem QC berbasis VLM sebagai acuan implementasi di lingkungan industri

Quality control (QC) in manufacturing industries typically relies on manual visual inspection prone to human fatigue and subjectivity. Conventional deep learning approaches require large labeled datasets and time-consuming model training. This study designs a web-based system architecture leveraging GPT-4o Vision to automatically classify product defects without dataset training using zero-shot inference. The system uses a three-tier architecture: a web-based UI, a business logic layer, and VLM API integration. Conceptual feasibility is evaluated through comparative analysis, use-case scenarios, and a risk matrix. Results show this approach reduces annotation requirements, accelerates deployment, and produces explainable outputs. This study contributes a VLM-based QC blueprint for industrial implementation.



This is an open access article under the CC-BY-SA license.

How to Cite Rino Lengan el at (2026) Pendampingan Pengembangan Bahan Ajar Digital Berbasis Potensi Lokal dengan Canva Bagi Mahasiswa PGSD PSDKU Universitas Pattimura <https://doi.org/10.31004/jerkin.v5i1.6598>

PENDAHULUAN

Industri manufaktur modern menghadapi tekanan yang semakin besar untuk meningkatkan efisiensi dan akurasi dalam proses pengendalian kualitas (quality control). Inspeksi visual manual yang selama ini menjadi andalan memiliki keterbatasan fundamental, di antaranya tingkat kelelahan operator, inkonsistensi penilaian antarpersonel, serta ketidakmampuan memproses volume produksi tinggi secara

real-time [1]. Data dari laporan industri menunjukkan bahwa kesalahan inspeksi manual dapat mencapai 20–30% dalam kondisi produksi padat [2].

Pendekatan kecerdasan buatan berbasis computer vision dan deep learning telah menawarkan solusi yang menjanjikan. Namun, implementasinya masih menghadapi hambatan berupa kebutuhan akan dataset berlabel dalam jumlah besar, proses anotasi yang mahal, serta keharusan melatih ulang model setiap kali jenis produk atau cacat berubah [3]. Hal ini menjadikan pendekatan konvensional kurang fleksibel untuk industri dengan variasi produk tinggi.

Kemunculan Vision Language Model (VLM) generasi terbaru seperti GPT-4o Vision [4], Gemini Vision, dan LLaVA [5] membuka paradigma baru dalam computer vision. Model-model ini dilatih pada data visual dan teks berskala miliaran sehingga dapat dimanfaatkan langsung tanpa pelatihan tambahan melalui zero-shot inference [6].

Penelitian ini bertujuan merancang arsitektur sistem berbasis aplikasi web yang memanfaatkan VLM untuk mengklasifikasikan cacat produk dalam konteks quality control. Fokus utama adalah perancangan konseptual, alur sistem, serta evaluasi kelayakan dari perspektif rekayasa sistem.

METODE

Penelitian ini menggunakan pendekatan Design Science Research (DSR), yaitu metodologi yang berfokus pada perancangan dan evaluasi artefak untuk memecahkan masalah praktis di domain tertentu [15]. DSR dipilih karena penelitian ini menghasilkan artefak desain berupa blueprint sistem yang dapat divalidasi kelayakannya.

Tahapan Penelitian

Tabel 1. Tahapan Design Science Research

No.	Aktivitas	Output
1	Identifikasi masalah & motivasi	Problem statement
2	Definisi tujuan solusi	Spesifikasi kebutuhan
3	Perancangan artefak	Arsitektur & alur kerja sistem
4	Demonstrasi konseptual	Skenario use-case
5	Evaluasi kelayakan	Analisis komparatif & risiko
6	Komunikasi hasil	Laporan jurnal

Teknik Pengumpulan Data

Data yang digunakan bersifat sekunder, meliputi: (1) studi literatur sistematis dari jurnal terkait computer vision, VLM, dan quality control tahun 2019–2024; (2) dokumentasi teknis API GPT-4o Vision dari OpenAI; (3) standar industri ISO 9001 terkait inspeksi kualitas; dan (4) studi kasus implementasi VLM di domain serupa.

Metode Evaluasi Kelayakan

Evaluasi kelayakan dilakukan melalui tiga pendekatan: (1) Analisis Komparatif membandingkan VLM zero-shot dengan CNN konvensional; (2) Skenario Use-Case mensimulasikan alur inspeksi; dan (3) Matriks Risiko mengidentifikasi dan merencanakan mitigasi risiko teknis dan operasional.

HASIL DAN PEMBAHASAN

Gambaran Umum Arsitektur

Sistem dirancang menggunakan Three-Tier Architecture yang memisahkan tanggung jawab antara antarmuka pengguna, logika bisnis, dan layanan eksternal. Pemisahan ini memastikan skalabilitas, kemudahan pemeliharaan, dan fleksibilitas integrasi VLM.

Tabel 2. Arsitektur Three-Tier Sistem VLM QC

Layer	Komponen	Fungsi Utama
Presentation	Web Browser	Upload gambar, tampil hasil, dashboard
Application	Backend API	Validasi, preprocessing, parsing output VLM

Service	GPT-4o Vision	Inferensi zero-shot, klasifikasi cacat
Data	Database	Penyimpanan citra, log audit, laporan

Alur Kerja Sistem

Alur kerja sistem terdiri dari tujuh tahap berurutan yang membentuk pipeline inspeksi end-to-end dengan mekanisme penanganan kesalahan untuk keandalan operasional.

Tabel 3. Pipeline Alur Kerja Sistem Inspeksi VLM

No	Tahap	Deskripsi Proses
1	Input Citra	Upload citra JPEG/PNG/WebP min. 512x512px
2	Preprocessing	Resize, normalisasi, konversi Base64
3	Konstruksi Instruksi	Instruksi terstruktur berdasarkan jenis produk dan kategori cacat
4	Inferensi VLM	GPT-4o Vision: zero-shot reasoning & klasifikasi cacat
5	Parsing Output	Ekstraksi jenis cacat, keparahan, lokasi, confidence score
6	Keputusan QC	PASS / REJECT / REVIEW berdasarkan threshold
7	Logging	Simpan ke database untuk audit & analisis tren

Spesifikasi Kebutuhan Sistem

Tabel 4. Spesifikasi Kebutuhan Fungsional dan Non-Fungsional

Jenis	Kode	Deskripsi Kebutuhan
F	F-01	Sistem menerima citra produk JPEG, PNG, WebP
F	F-02	Menghasilkan klasifikasi cacat dengan keparahan dan confidence score
F	F-03	Memberikan keputusan QC: PASS / REJECT / REVIEW
F	F-04	Menyimpan riwayat dan mendukung ekspor laporan
NF	NF-01	Waktu respons tidak melebihi 15 detik per inspeksi
NF	NF-02	Mendukung minimal 10 pengguna simultan
NF	NF-03	Antarmuka responsif, akses via browser tanpa instalasi
NF	NF-04	Ketersediaan sistem minimal 99% pada jam operasional

Kategori Cacat Produk yang Dapat Diklasifikasikan

Berdasarkan studi literatur dan analisis kebutuhan industri, sistem dirancang untuk mengklasifikasikan sembilan kategori cacat produk umum yang dapat diidentifikasi secara visual.

Tabel 5. Kategori Cacat Produk dan Tingkat Keparahan

No	Kategori	Deskripsi Visual	Keparahan
1	Goresan	Bekas gesekan linier pada permukaan	Minor–Major
2	Retak	Keretakan struktural produk	Major–Critical
3	Penyok	Deformasi permukaan akibat benturan	Minor–Major
4	Perubahan Warna	Perbedaan warna dari standar	Minor–Major
5	Kontaminasi	Material asing pada permukaan	Major–Critical
6	Deformasi	Perubahan bentuk dari desain	Critical

7	Komponen Hilang	Bagian yang seharusnya ada tidak ada	Critical
8	Kesalahan Dimensi	Ukuran tidak sesuai toleransi	Major–Critical
9	Kekasaran Permukaan	Tekstur tidak sesuai standar finishing	Minor

Analisis Komparatif VLM vs CNN Konvensional

Perancangan sistem berbasis VLM menawarkan keunggulan strategis dibandingkan pendekatan CNN konvensional sebagaimana ditunjukkan pada Tabel 6 berikut.

Tabel 6. Perbandingan VLM Zero-Shot vs CNN Konvensional

Aspek	CNN Konvensional	VLM Zero-Shot
Kebutuhan Dataset	Ribuan citra berlabel	Tidak diperlukan
Waktu Persiapan	Minggu hingga bulan	Hari hingga minggu
Adaptasi Produk Baru	Pelatihan ulang wajib	Modifikasi instruksi
Explainability	Terbatas (black-box)	Tinggi (natural language)
Biaya Infrastruktur	Tinggi (GPU)	Sedang (pay-per-use)
Akurasi Cacat Langka	Rendah (underfitting)	Lebih baik
Kemandirian Koneksi	Dapat offline	Butuh internet
Konsistensi Output	Sangat konsisten	Perlu validasi

Dari perbandingan di atas, pendekatan VLM unggul pada aspek kecepatan deployment dan fleksibilitas adaptasi produk, sangat relevan untuk industri kecil-menengah yang tidak memiliki sumber daya GPU atau dataset besar [11].

Analisis Risiko dan Mitigasi

Tabel 7. Matriks Risiko dan Strategi Mitigasi

Risiko	Prob.	Dampak	Strategi Mitigasi
Gangguan koneksi / API	Sedang	Tinggi	Retry otomatis, antrian offline, fallback manual
Kebocoran data citra	Rendah	Tinggi	Enkripsi TLS, VLM on-premise
Inkonsistensi output VLM	Sedang	Sedang	Schema validation, temperature rendah
Biaya API tinggi	Tinggi	Sedang	Caching, batching, model lokal
Cacat tidak dikenali VLM	Sedang	Sedang	Deskripsi visual detail, eskalasi manual

Posisi Penelitian terhadap Kajian Terdahulu

Penelitian ini berbeda dari studi terdahulu. Jiang et al. [7] mengoptimalkan CNN untuk akurasi tinggi namun membutuhkan dataset besar, sedangkan rancangan ini mengeliminasi ketergantungan tersebut. Cao et al. [9] mengimplementasikan VLM pada komponen elektronik secara empiris, sementara penelitian ini berfokus pada perancangan arsitektur sistem komprehensif untuk konteks industri yang lebih luas.

Kontribusi utama: (1) blueprint arsitektur tiga lapis untuk integrasi VLM dalam sistem QC; (2) alur kerja end-to-end dengan mekanisme fallback; dan (3) analisis komparatif serta matriks risiko sebagai panduan praktis adopsi VLM di industri.

KESIMPULAN

Penelitian ini telah berhasil merancang arsitektur sistem berbasis aplikasi web untuk klasifikasi cacat produk dalam proses quality control dengan memanfaatkan Vision Language Model GPT-4o Vision tanpa memerlukan training dataset. Kesimpulan yang dapat ditarik:

1. Arsitektur three-tier yang dirancang mampu secara konseptual mengintegrasikan VLM ke dalam alur kerja quality control industri dengan pemisahan tanggung jawab yang jelas.
2. Pipeline inspeksi tujuh tahap menyediakan alur kerja komprehensif dengan mekanisme penanganan kesalahan yang menjamin keandalan operasional.
3. Analisis komparatif menunjukkan VLM zero-shot unggul pada aspek eliminasi kebutuhan dataset, kecepatan deployment, dan explainability dibandingkan CNN konvensional.
4. Matriks risiko mengidentifikasi lima risiko utama beserta strategi mitigasi sebagai panduan implementasi di lingkungan industri.

Untuk penelitian selanjutnya disarankan: (1) implementasi dan evaluasi empiris menggunakan dataset industri nyata; (2) eksplorasi VLM open-source on-premise seperti LLaVA; (3) pengembangan mekanisme active learning berbasis feedback operator; dan (4) integrasi dengan sistem kamera industri untuk inspeksi real-time.

UCAPAN TERIMA KASIH

(Baik orang maupun lembaga yang berkontribusi dalam proses penulisan artikel dapat disebutkan di sini).

REFERENSI

- [1] Pham, H., & Alcock, R. (2003). Reducing Human Error in Manufacturing. *Human Factors and Ergonomics in Manufacturing*, 13(4), 319–335.
- [2] Kumar, S., Singh, R., & Sharma, A. (2021). Human Error in Manual Visual Inspection. *International Journal of Advanced Manufacturing Technology*, 114(7), 2023–2041.
- [3] Bergmann, P., et al. (2019). MVTec AD — A Comprehensive Dataset for Unsupervised Anomaly Detection. *Proc. CVPR*, 9592–9600.
- [4] OpenAI. (2024). GPT-4o System Card. OpenAI Technical Report. <https://openai.com/research/gpt-4o-system-card>
- [5] Liu, H., Li, C., Wu, Q., & Lee, Y. J. (2023). Visual Instruction Tuning. *Advances in NeurIPS*, 36, 34892–34916.
- [6] Brown, T. B., et al. (2020). Language Models are Few-Shot Learners. *Advances in NeurIPS*, 33, 1877–1901.
- [7] Jiang, X., Peng, Z., & Wang, H. (2022). Deep Learning-Based Visual Defect Detection: A Review. *Journal of Manufacturing Systems*, 65, 691–708.
- [8] Radford, A., et al. (2021). Learning Transferable Visual Models from Natural Language. *Proc. ICML*, 8748–8763.
- [9] Cao, Z., Chen, Y., & Liu, W. (2024). VLM-Inspect: Zero-Shot Industrial Defect Classification. *IEEE Trans. Industrial Informatics*, 20(3), 2145–2157.
- [10] Menon, S., & Vondrick, C. (2023). Visual Classification via Description from LLMs. *ICLR 2023*.
- [11] Hevner, A. R., et al. (2004). Design Science in IS Research. *MIS Quarterly*, 28(1), 75–105.
- [12] Nielsen, J. (1994). *Usability Engineering*. Morgan Kaufmann Publishers.
- [13] Roth, K., et al. (2022). Towards Total Recall in Industrial Anomaly Detection. *Proc. CVPR*, 14318–14328.
- [14] Google. (2024). Gemini 1.5 Technical Report. Google DeepMind. <https://deepmind.google/technologies/gemini>