

Sistem Deteksi Penyakit Gigi Berbasis Deep Learning Menggunakan YOLOv8 dan ResNet-18

Bagas Aditya^{1*}, Rully Pramudita²

^{1,2}Teknik Informatika, Universitas Bina Insani, Indonesia

^{1*}bagas331000@gmail.com, ²rullypramudita@binainsani.ac.id

Abstrak: Penyakit gigi dan mulut menjadi salah satu masalah kesehatan yang paling umum di Indonesia, dengan prevalensi mencapai 57,6% populasi menurut Survei Kesehatan Indonesia. Proses diagnosis konvensional masih sangat bergantung pada pemeriksaan visual manual oleh dokter gigi, yang memiliki keterbatasan dari segi waktu, subjektivitas, dan aksesibilitas. Penelitian ini mengembangkan sistem deteksi penyakit gigi berbasis deep learning yang mengintegrasikan arsitektur YOLOv8 untuk object detection dan ResNet-18 untuk image classification dalam sebuah pipeline ensemble. Sistem dirancang untuk mendeteksi enam jenis kelainan gigi: karies, karang gigi, radang gusi, hipodontia, sariawan, dan diskolorasi gigi dari foto kamera ponsel. Dataset yang digunakan berjumlah 11.957 citra yang dibagi menjadi 70% data latih, 15% validasi, dan 15% pengujian. Teknik weighted sampling diimplementasikan untuk menangani ketimpangan kelas dengan rasio 7,96x. Pelatihan ResNet-18 menggunakan optimizer Adam (learning rate 0,001) dengan fungsi kerugian CrossEntropyLoss berbobot kelas dinamis. Hasil evaluasi menunjukkan YOLOv8 mencapai mAP@50 sebesar 88,17%, sementara ResNet-18 memperoleh akurasi klasifikasi 92,25% dengan F1-Score 92,37%. Validasi statistik 5-Fold Cross Validation mengonfirmasi stabilitas ResNet-18 (Standar Deviasi = $\pm 0,45\%$) dan YOLOv8 (Standar Deviasi = $\pm 1,95\%$). Sistem ini diimplementasikan dalam aplikasi web menggunakan FastAPI dan Next.js pada GPU NVIDIA T4, dengan latensi end-to-end 2-4 detik, serta dilengkapi modul Grad-CAM untuk interpretabilitas prediksi.

Kata Kunci: Deep Learning; YOLOv8; ResNet-18; Object Detection; Image Classification;

Abstract: Dental and oral diseases are among the most prevalent health problems in Indonesia, affecting 57.6% of the population according to the Indonesian Health Survey. Conventional diagnosis still relies heavily on manual visual examination by dentists, which is constrained by time, subjectivity, and accessibility limitations. This study developed a deep learning-based dental disease detection system that integrates the YOLOv8 architecture for object detection and ResNet-18 for image classification within an ensemble pipeline. The system was designed to detect six types of dental abnormalities: caries, calculus, gingivitis, hypodontia, mouth ulcer, and tooth discoloration from smartphone camera photographs. The dataset comprised 11,957 images split into 70% training, 15% validation, and 15% testing sets. Weighted sampling

was implemented to address class imbalance with an imbalance ratio of 7.96x. Evaluation results demonstrated that YOLOv8 achieved an mAP@50 of 88.17% for disease area localization, while ResNet-18 obtained a classification accuracy of 92.25% with an F1-Score of 92.37%. Statistical validation using 5-Fold Cross Validation confirmed the stability of ResNet-18 (Standard Deviation = $\pm 0.45\%$) and YOLOv8 (Standard Deviation = $\pm 1.95\%$). The system was deployed as a web application using FastAPI and Next.js, equipped with a Grad-CAM module for prediction interpretability.

Keywords: Deep Learning; YOLOv8; ResNet-18; Object Detection; Image Classification;

1. PENDAHULUAN

Masalah kesehatan gigi dan mulut merupakan kondisi yang terbilang umum di kalangan masyarakat Indonesia. Berdasarkan data Survei Kesehatan Indonesia (SKI), sebanyak 57,6% penduduk berusia tiga tahun ke atas pernah mengalami gangguan pada gigi dan mulut mereka [1]. Angka ini cukup mengejutkan mengingat sebagian besar masyarakat cenderung menunda pemeriksaan ke dokter gigi sampai rasa sakit benar-benar tak tertahankan. Beberapa jenis penyakit gigi yang kerap ditemui antara lain karies (gigi berlubang), karang gigi, radang gusi, hipodontia, sariawan, dan diskolorasi gigi. Dari sekian banyak kasus, gigi berlubang tercatat sebagai yang paling dominan dengan prevalensi 43,6%, diikuti oleh karies gigi yang mencapai 88,8% [1]. Kondisi semacam ini bukan hanya mengganggu secara fisik, namun dampak psikologisnya juga tidak bisa dianggap remeh. Masalah kesehatan periodontal bahkan dapat memberikan dampak sistemik bagi penderita [2], [3]. Penderita sering merasa malu dan akhirnya menarik diri dari pergaulan sosial [4]. Oleh karena itu, edukasi mengenai anatomi dan perawatan gigi sejak dini menjadi sangat penting [5].

Pada praktiknya, proses diagnosis penyakit gigi masih didominasi oleh pemeriksaan visual manual yang dilakukan langsung oleh dokter gigi. Pendekatan konvensional ini memiliki sejumlah kendala yang cukup serius. Pertama, tingkat subjektivitas yang tinggi karena diagnosis sangat bergantung pada pengalaman dan ketelitian masing-masing praktisi [6], [7]. Kedua, keterbatasan waktu dan jumlah tenaga medis membuat banyak pasien tidak mendapatkan respons diagnosis yang cepat, terutama ketika mereka mengirimkan foto keluhan gigi melalui aplikasi pesan [6]. Diperlukan sistem otomatis untuk meningkatkan kesadaran dan perilaku deteksi dini pada pasien [8]. Di Hai Dent Clinic misalnya, tenaga medis kerap menerima foto rongga mulut pasien lewat WhatsApp untuk konsultasi awal. Permasalahannya, dokter tidak selalu punya waktu untuk memeriksa satu per satu gambar tersebut, apalagi kualitas fotonya sering kali tidak memadai karena buram, gelap, atau terhalang bagian mulut lainnya.

Beberapa penelitian terdahulu telah mencoba memanfaatkan deep learning untuk keperluan diagnosis penyakit gigi [9]. Penelitian pertama menerapkan arsitektur CNN klasifikasi tunggal untuk membedakan dua jenis penyakit, yaitu karies dan kalkulus [10]. Model tersebut mencapai akurasi di atas 90%, namun memiliki beberapa keterbatasan yaitu cakupan kelas yang sangat terbatas (hanya 2 dari sekian banyak penyakit gigi yang ada), tidak tersedianya mekanisme lokalisasi area penyakit pada citra, serta absennya komponen interpretabilitas yang memungkinkan tenaga medis memverifikasi dasar keputusan model. Penelitian kedua menggunakan arsitektur ResNet-18 untuk mengklasifikasikan kondisi gigi berdasarkan citra RGB [11]. Pendekatan ini memproses gambar secara utuh tanpa tahapan pemotongan objek (object detection) terlebih dahulu, sehingga fitur-fitur visual yang tidak relevan seperti tekstur bibir, lidah, dan pipi turut masuk ke dalam perhitungan bobot klasifikasi. Akibatnya, akurasi model cenderung menurun ketika dihadapkan pada gambar yang diambil dalam kondisi tidak terkontrol.

Penelitian ketiga berhasil mengimplementasikan algoritma YOLOv8 untuk melakukan segmentasi pada citra panoramik gigi [12]. Meskipun efektif dalam melokalisasi struktur anatomis gigi pada hasil rontgen, metode ini memiliki ketergantungan pada citra radiografi klinis yang hanya dapat dihasilkan menggunakan peralatan medis di fasilitas kesehatan dan tidak dapat diakses secara mandiri oleh pasien umum. Tabel perbandingan ketiga penelitian terdahulu ini disajikan pada bagian Hasil dan Pembahasan (Tabel 8).

Berdasarkan analisis terhadap ketiga penelitian di atas, teridentifikasi tiga celah utama yang belum terjawab secara simultan oleh satu sistem. Pertama, pendekatan klasifikasi tunggal rentan terhadap noise visual karena seluruh piksel gambar ikut diproses tanpa tahapan lokalisasi [10], [11]. Kedua, sistem yang sudah menggunakan object detection masih bergantung pada citra rontgen yang tidak praktis untuk skrining mandiri [12]. Ketiga, belum ada penelitian yang menyertakan komponen Explainable AI untuk mendukung verifikasi klinis oleh dokter gigi. Penelitian ini secara spesifik menjawab ketiga celah tersebut melalui arsitektur ensemble two-stage. Kontribusi ilmiah (novelty) dari penelitian ini terletak pada tiga aspek yang tidak ditemukan secara simultan pada studi sebelumnya. Pertama, pemisahan tugas lokalisasi (YOLOv8) dan klasifikasi (ResNet-18) ke dalam dua model independen yang bekerja secara pipeline. YOLOv8 bertugas memotong Region of Interest (ROI) dari foto kamera ponsel biasa, sehingga area yang masuk ke tahap klasifikasi sudah terbebas dari noise visual latar belakang. Pendekatan ini secara langsung mengatasi kelemahan metode klasifikasi tunggal yang memproses gambar utuh [10], [11]. Kedua, peta aktivasi visual (heatmap) pada lapisan konvolusi terakhir ResNet-18 dihasilkan melalui integrasi modul Grad-CAM. Melalui peta aktivasi ini, dokter gigi dapat secara visual memverifikasi bahwa keputusan klasifikasi memang didasarkan pada area lesi yang relevan, bukan pada artefak gambar atau noise latar belakang [13]. Komponen ini menjawab ketiadaan mekanisme interpretabilitas pada penelitian terdahulu [10]. Ketiga, implementasi end-to-end dalam bentuk aplikasi web berbasis FastAPI dan Next.js yang memungkinkan pasien melakukan skrining mandiri menggunakan foto kamera ponsel tanpa memerlukan peralatan rontgen klinis, menjawab keterbatasan aksesibilitas pada penelitian sebelumnya [12]. Pendekatan berbasis YOLOv8 ini dipilih karena keandalannya tidak hanya teruji pada deteksi anatomi gigi spesifik, namun juga telah sukses dievaluasi secara real-time pada berbagai kasus deteksi objek lainnya [14], [15], [16], [17], [18], [19].

2. METODE PENELITIAN

Penelitian ini mengadopsi pendekatan CRISP-DM (Cross-Industry Standard Process for Data Mining) sebagai kerangka metodologis utama. Pemilihan framework ini didasarkan pada sifatnya yang iteratif dan terstruktur, sehingga cocok untuk proyek berbasis machine learning yang membutuhkan evaluasi berulang. Secara garis besar, alur kerja penelitian mencakup enam tahapan: pemahaman domain bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, dan deployment.

Seluruh eksperimen pelatihan model dilaksanakan menggunakan lingkungan Google Colab dengan akselerator GPU NVIDIA T4. Spesifikasi perangkat keras lokal yang digunakan untuk pengembangan meliputi prosesor Intel Core i5 (multi-threading), memori 16 GB DDR4, dan penyimpanan SSD dengan kapasitas minimal 20 GB. Perangkat lunak utama terdiri dari Python 3.9+, PyTorch sebagai framework deep learning untuk ResNet-18, Ultralytics untuk implementasi YOLOv8, dan pustaka pytorch-grad-cam untuk modul interpretabilitas. Lingkungan pengembangan menggunakan Visual Studio Code sebagai IDE dan Google Colab sebagai environment pelatihan.

Dataset yang digunakan bersumber dari repositori publik Kaggle berjudul "Oral Diseases" (<https://www.kaggle.com/datasets/salmansajid05/oral-diseases>). Dataset ini merupakan kompilasi citra fotografi intraoral yang dikumpulkan dari berbagai sumber klinis

daring dan telah dikurasi untuk keperluan riset computer vision di bidang kedokteran gigi. Citra-citra dalam dataset ini menampilkan kondisi rongga mulut pasien dari berbagai sudut pengambilan dan kondisi pencahayaan yang bervariasi, mencakup foto dengan kualitas tinggi maupun foto bercahaya rendah yang menyerupai kondisi pengambilan gambar mandiri oleh pasien melalui kamera ponsel. Total dataset mencakup 11.957 citra yang terdistribusi ke dalam tujuh kelas: Calculus (1.296 citra, 10,8%), Caries (2.382 citra, 19,9%), Gingivitis (2.349 citra, 19,6%), Healthy (319 citra, 2,7%), Hypodontia (1.250 citra, 10,5%), Mouth Ulcer (2.540 citra, 21,2%), dan Tooth Discoloration (1.821 citra, 15,2%). Distribusi ini sebagaimana dirangkum pada Tabel 1. Setiap citra dalam dataset berformat JPG/JPEG/PNG dengan resolusi yang bervariasi secara luas, mulai dari gambar beresolusi rendah hingga tinggi. Dataset tidak menyertakan metadata identitas pasien, riwayat medis, maupun informasi demografis, sehingga seluruh citra diperlakukan sebagai sampel independen.

Tabel 1. Distribusi Jumlah Citra

No	Kelas Penyakit	Jumlah Citra
1	Calculus	1.296
2	Caries	2.382
3	Gingivitis	2.349
4	Healthy	319
5	Hypodontia	1.250
6	Mouth Ulcer	2.540
7	Tooth Discoloration	1.821
Total	7 Kelas	11.957

Analisis eksplorasi data mengungkapkan ketimpangan distribusi kelas dengan rasio 7,96x antara kelas mayoritas (Mouth Ulcer) dan minoritas (Healthy). Resolusi citra bervariasi secara luas, sehingga seluruh gambar dinormalisasi ke dimensi seragam 224x224 piksel pada tahap pra-pemrosesan. Dataset merupakan kumpulan foto anonim tanpa metadata identitas pasien, sehingga setiap citra bersifat independen satu sama lain. Pembagian dataset dilakukan menggunakan fungsi `random_split` PyTorch dengan proporsi 70% untuk pelatihan (8.371 citra), 15% untuk validasi (1.793 citra), dan 15% untuk pengujian (1.793 citra). Parameter `manual_seed` digunakan untuk memastikan komposisi partisi tetap konsisten setiap kali skrip dieksekusi ulang.

Terkait potensi kebocoran data (data leakage), perlu diklarifikasi mekanisme pembagian dataset yang diterapkan. Pembagian data dilakukan pada level citra (image-level split) menggunakan fungsi `random_split` dari pustaka PyTorch dengan parameter `manual_seed=42` untuk menjamin reproduktibilitas partisi. Dengan konfigurasi ini, tidak terdapat duplikasi citra yang sama pada subset pelatihan, validasi, dan pengujian. Namun demikian, karena dataset "Oral Diseases" dari Kaggle merupakan kumpulan citra yang dikompilasi dari berbagai sumber klinis daring dan tidak menyertakan metadata identitas pasien (patient ID), terdapat kemungkinan bahwa beberapa citra berasal dari pasien yang sama dengan sudut pengambilan atau waktu yang berbeda. Kondisi ini berpotensi menyebabkan kebocoran informasi antar subset data karena model mungkin mengenali pola visual spesifik dari pasien tertentu yang muncul di set pelatihan dan set pengujian secara bersamaan. Praktik ideal untuk mengatasi risiko ini adalah patient-level split, namun hal tersebut tidak dapat dilakukan tanpa metadata pasien.

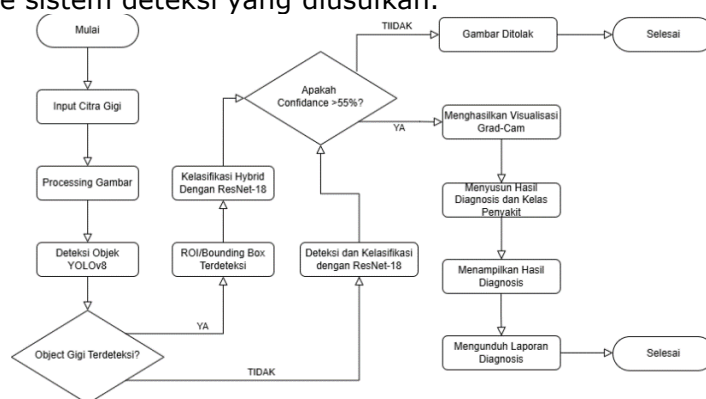
Sebagai mekanisme validasi statistik tambahan, dilakukan 5-Fold Cross Validation untuk mengukur stabilitas performa model terhadap variasi pembagian data. Pada skenario ini, seluruh dataset dibagi menjadi 5 partisi berukuran setara. Setiap iterasi menggunakan 4 fold untuk pelatihan dan 1 fold untuk validasi, diulang hingga seluruh fold pernah menjadi data validasi. Untuk model ResNet-18, digunakan Stratified K-Fold dengan pelatihan 20

epoch per fold menggunakan optimizer Adam (learning rate = 0,001). Untuk model YOLOv8, digunakan K-Fold standar dengan pelatihan 100 epoch per fold menggunakan pretrained weights YOLOv8n.

Tahapan pra-pemrosesan data dirancang untuk menyeragamkan input sekaligus meningkatkan variasi data latih. Seluruh citra dinormalisasi ke resolusi 224×224 piksel sesuai standar arsitektur ResNet-18. Pada data latih, diterapkan augmentasi stokastik yang meliputi RandomCrop, RandomHorizontalFlip, rotasi hingga 15 derajat, dan ColorJitter untuk variasi kecerahan, kontras, serta saturasi. Transformasi ini bertujuan memaksa model belajar mengenali pola patologi dari berbagai sudut dan kondisi pencahayaan yang berbeda, menyimulasikan variasi kualitas foto yang diambil oleh pasien secara mandiri. Untuk mengatasi ketimpangan kelas, diimplementasikan teknik WeightedRandomSampler yang memberikan probabilitas sampling lebih tinggi pada kelas minoritas selama proses iterasi pelatihan dengan batch size 32.

Arsitektur sistem yang dibangun terdiri dari dua komponen utama yang bekerja secara sekuensial. Komponen pertama adalah YOLOv8 yang berperan sebagai unit lokalisasi. Model ini memproses citra masukan melalui backbone berbasis Cross-Stage Partial (C2f) untuk ekstraksi fitur hierarkis, kemudian menggunakan mekanisme grid-based prediction untuk menentukan koordinat bounding box pada area penyakit. Pendekatan single forward pass yang diadopsi YOLO memungkinkan inferensi dalam hitungan milidetik, yang sangat krusial untuk aplikasi berbasis web. YOLOv8 dilatih selama 30 epoch dengan konfigurasi default Ultralytics.

Komponen kedua adalah ResNet-18 yang berfungsi sebagai unit klasifikasi. Potongan gambar hasil lokalisasi YOLOv8 langsung disalurkan ke jaringan ini untuk diidentifikasi jenis penyakitnya. Arsitektur ResNet-18 dipilih karena keunggulan residual block yang mampu mengatasi masalah vanishing gradient, sekaligus menawarkan keseimbangan optimal antara akurasi dan efisiensi komputasi dibandingkan varian yang lebih berat seperti ResNet-50. Model menggunakan transfer learning dari bobot pra-pelatihan ImageNet (IMAGENET1K_V1), dengan 683.072 parameter dibekukan pada lapisan konvolusi awal dan 10,6 juta parameter pada layer3 dan layer4 difokuskan untuk fine-tuning. Blok klasifikasi bawaan (1.000 kelas) diganti dengan arsitektur Sequential baru yang mengerucut pada 7 kelas patologi, dilengkapi dua lapisan Dropout (30% dan 20%) untuk pencegahan overfitting. Fungsi kerugian menggunakan CrossEntropyLoss dengan class weights dinamis yang dihitung berdasarkan distribusi invers jumlah sampel tiap kelas. Optimasi bobot dieksekusi menggunakan algoritma Adam (Adaptive Moment Estimation) dengan learning rate awal 0,001 dan penjadwal ReduceLROnPlateau yang memangkas learning rate separuhnya apabila performa validasi mengalami stagnasi selama tiga epoch berturut-turut. Proses pelatihan menggunakan mekanisme early stopping dengan patience 7 epoch dan batas maksimal 30 epoch. Gambar 1 mengilustrasikan diagram alir keseluruhan pipeline sistem deteksi yang diusulkan.



Gambar 1. Arsitektur Pipeline Sistem Deteksi Penyakit Gigi

Sebagai komponen pendukung interpretabilitas, sistem juga mengintegrasikan modul Gradient-weighted Class Activation Mapping (Grad-CAM). Algoritma ini menyadap gradien dari lapisan konvolusi terakhir ResNet-18 (3×3 conv, 512 filter) dan memproyeksikannya kembali menjadi peta aktivasi warna panas (heatmap) yang di-overlay pada citra asli. Sofiyah [13] mengonfirmasi bahwa peta visual Grad-CAM sanggup menandai letak spesifik anomali yang menjadi dasar utama keputusan model komputasi, sehingga tenaga medis dapat memverifikasi secara visual bahwa keputusan klasifikasi memang didasarkan pada area lesi yang relevan. Keseluruhan sistem kemudian diimplementasikan dalam arsitektur client-server: FastAPI sebagai backend untuk menjalankan inferensi model, dan Next.js sebagai frontend untuk antarmuka pengguna.

3. HASIL DAN PEMBAHASAN

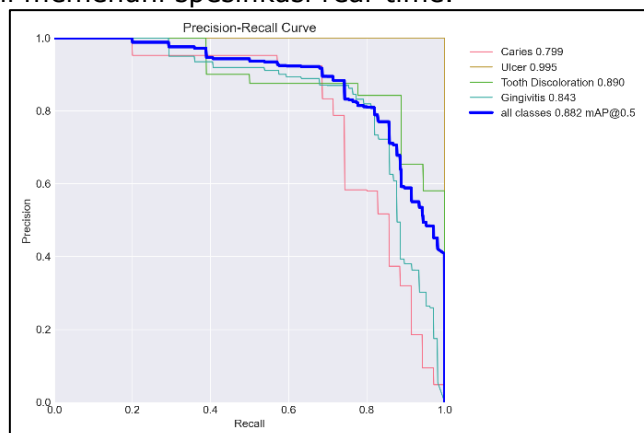
Evaluasi sistem dilakukan secara bertahap terhadap masing-masing komponen arsitektur, kemudian dilanjutkan dengan pengujian pipeline terintegrasi dan validasi klinis oleh dokter gigi profesional.

Pada komponen pertama, model YOLOv8 dilatih selama 30 epoch penuh tanpa terpicu mekanisme early stopping. Kurva training loss (box_loss, cls_loss, dan dfl_loss) menunjukkan penurunan yang konsisten dari epoch pertama hingga terakhir, mengindikasikan proses konvergensi yang stabil. Metrik evaluasi pada data validasi menunjukkan performa lokalisasi yang baik, sebagaimana dirangkum pada Tabel 2.

Tabel 2. Hasil Evaluasi Model YOLOv8

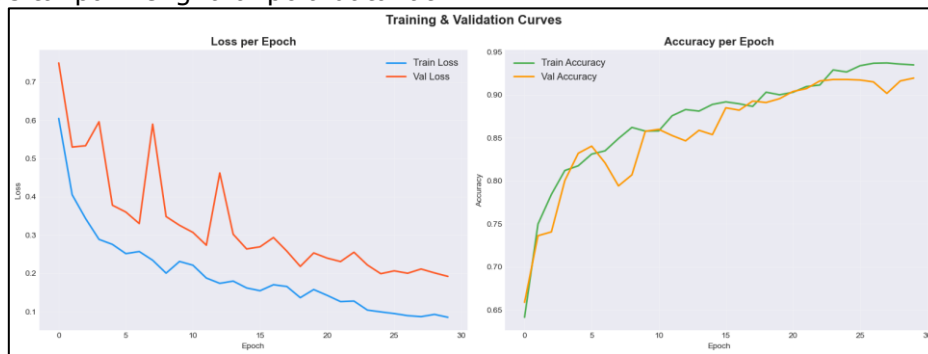
No	Metrik Evaluasi	Nilai
1	mAP@50	0,8817 (88,17%)
2	mAP@50-95	0,5009 (50,09%)
3	Precision (Overall)	0,872 (87,2%)
4	Recall (Overall)	0,839 (83,9%)

Model YOLOv8 secara kumulatif mencapai nilai mAP@50 sebesar 88,17% yang menandakan kemampuan lokalisasi penyakit gigi yang cukup tinggi. Jika ditelusuri per kelas, deteksi paling akurat dicapai pada kelas Ulcer (sariawan) dengan mAP@50 mencapai 99,5%, yang masuk akal mengingat sariawan memiliki karakteristik visual yang cukup distinktif berupa area putih kekuningan di atas jaringan merah. Sebaliknya, kelas Caries mencatatkan skor terendah di angka 79,9% karena variasi visual lubang gigi yang sangat beragam tergantung tingkat keparahannya. Waktu inferensi YOLOv8 secara individual tercatat sebesar 54,1 milidetik per gambar pada GPU NVIDIA T4, membuktikan bahwa komponen deteksi ini memenuhi spesifikasi real-time.



Gambar 2. Kurva Precision-Recall Model YOLOv8

Pada komponen kedua, arsitektur ResNet-18 dilatih dengan mekanisme early stopping dan berhasil menyelesaikan seluruh 30 epoch tanpa penghentian dini. Akurasi pelatihan dimulai dari 64,13% pada epoch pertama dan secara progresif meningkat hingga mencapai puncak pada epoch ke-30 dengan akurasi validasi 91,97% dan nilai loss 0,1919. Analisis overfitting menunjukkan hasil yang sangat meyakinkan: selisih akurasi antara data latih (93,50%) dan data validasi (91,97%) hanya sebesar 1,53%, jauh di bawah ambang batas kritis 10%. Temuan ini mengonfirmasi bahwa strategi augmentasi data dan konfigurasi hyperparameter yang diterapkan berhasil memandu model untuk belajar secara generalistik tanpa menghafal pola data latih.



Gambar 3. Kurva Akurasi dan Loss Pelatihan ResNet-18

Evaluasi model ResNet-18 pada data uji (1.793 citra) menghasilkan metrik performa yang komprehensif sebagaimana disajikan pada Tabel 3 dan Tabel 4.

Tabel 3. Hasil Evaluasi Per Kelas Model ResNet-18

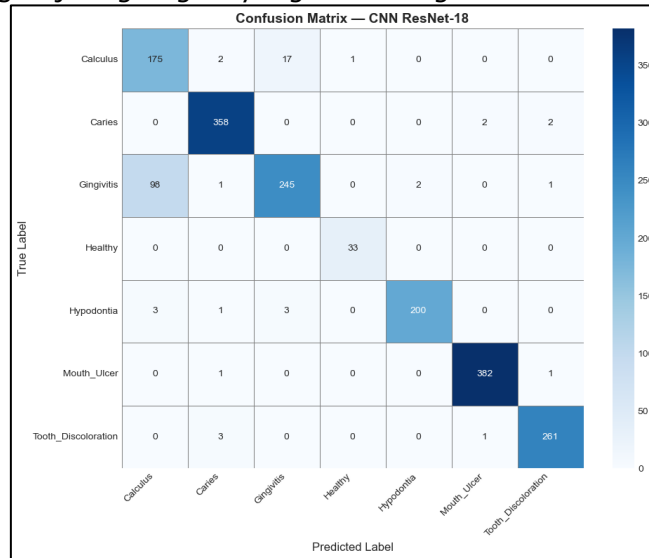
Kelas	Precision	Recall	F1-Score	Support
Calculus	0,63	0,90	0,74	195
Caries	0,98	0,99	0,98	362
Gingivitis	0,92	0,71	0,80	347
Healthy	0,97	1,00	0,99	33
Hypodontia	0,99	0,97	0,98	207
Mouth Ulcer	0,99	0,99	0,99	384
Tooth Discoloration	0,98	0,98	0,98	265

Tabel 4. Hasil Evaluasi Global Model ResNet-18

No	Metrik Evaluasi	Nilai
1	Accuracy	0,9225 (92,25%)
2	Precision	0,9356 (93,56%)
3	Recall	0,9225 (92,25%)
4	F1-Score	0,9237 (92,37%)

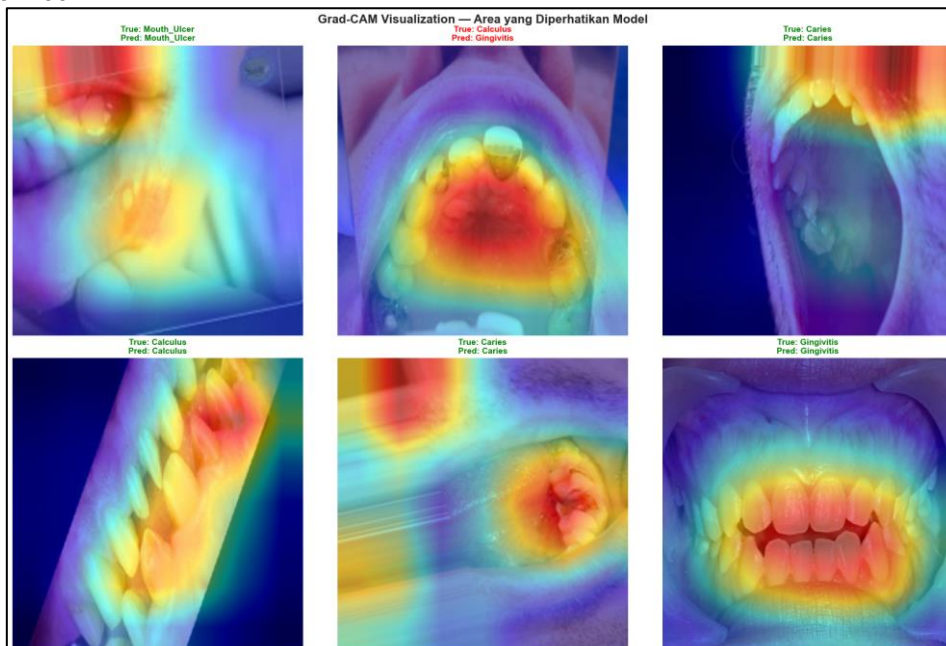
Secara keseluruhan, ResNet-18 mencetak akurasi 92,25% dengan F1-Score 92,37% pada data uji. Konsistensi antara Precision (93,56%) dan Recall (92,25%) menandakan bahwa model berhasil menekan baik false positive maupun false negative secara seimbang. Meski begitu, ada catatan penting dari analisis per kelas yang perlu diperhatikan. Kelas Calculus dan Gingivitis menunjukkan pola misklasifikasi yang cukup signifikan di antara keduanya. Precision pada Calculus tertekan di angka 63% dan Recall pada Gingivitis hanya mencapai 71%. Fenomena ini sebetulnya bisa dijelaskan secara klinis: karang gigi yang menumpuk di area leher gigi merupakan pemicu utama terjadinya radang gusi, sehingga kedua kondisi ini seringkali muncul bersamaan dalam satu frame

foto. Lapisan konvolusi pada CNN memang kesulitan memisahkan batas piksel antara plak yang mengeras dengan jaringan gusi yang meradang ketika keduanya saling berdekatan.



Gambar 4. Confusion Matrix Model ResNet-18

Analisis interpretabilitas menggunakan Grad-CAM menghasilkan peta aktivasi yang secara visual mengonfirmasi bahwa keputusan klasifikasi ResNet-18 memang didasarkan pada area lesi yang secara medis relevan. Pada kasus karies misalnya, area dengan intensitas heatmap tertinggi tepat terkonsentrasi di sekitar lubang dan kerusakan enamel. Sementara pada kasus diskolorasi gigi, aktivasi tersebar merata di seluruh permukaan gigi yang mengalami perubahan warna. Transparansi visual semacam ini sangat penting dalam konteks aplikasi medis, karena membantu tenaga medis memverifikasi bahwa sistem benar-benar membaca sinyal klinis yang tepat dan bukan sekadar menebak berdasarkan artefak gambar.



Gambar 5. Visualisasi Peta Aktivasi Grad-CAM pada Data Uji

Untuk memvalidasi stabilitas dan reliabilitas kedua model, dilakukan 5-Fold Cross Validation. Pada metode ini, seluruh dataset dibagi menjadi 5 partisi (fold) berukuran setara secara acak. Pada setiap iterasi, 4 fold (80% data) digunakan sebagai data pelatihan dan 1 fold (20% data) sebagai data validasi. Proses ini diulang sebanyak 5 kali sehingga setiap partisi pernah berperan sebagai data validasi. Khusus untuk ResNet-18, digunakan Stratified K-Fold yang mempertahankan proporsi kelas pada setiap fold, sedangkan untuk YOLOv8 digunakan K-Fold standar. Pendekatan ini dipilih untuk mengatasi potensi bias dari pembagian acak tunggal (single random split) dan memberikan estimasi performa yang lebih robust secara statistik. Metrik evaluasi rata-rata dan standar deviasi dihitung menggunakan formula berikut:

$$\bar{M} = \frac{1}{k} \sum_{i=1}^k M_i, \quad i = 1, 2, \dots, k \quad (1)$$

$$\sigma = \sqrt{\frac{1}{k} \sum_{i=1}^k (M_i - \bar{M})^2}, \quad i = 1, 2, \dots, k \quad (2)$$

Di mana \bar{M} menyatakan nilai rata-rata metrik evaluasi, k menyatakan jumlah fold ($k=5$), M_i menyatakan nilai metrik pada fold ke- i , dan σ menyatakan standar deviasi antar-fold yang mengindikasikan tingkat variabilitas performa model.

A. Hasil K-Fold Cross Validation ResNet-18

Tabel 5. Hasil K-Fold Cross Validation ($k=5$) ResNet-18

Fold	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Fold 1	99,03	99,03	99,03	99,03
Fold 2	98,83	98,83	98,83	98,83
Fold 3	98,25	98,25	98,25	98,25
Fold 4	98,05	98,06	98,05	98,05
Fold 5	99,22	99,22	99,22	99,22
Rata-rata	98,67 ± 0,45	98,68 ± 0,45	98,67 ± 0,45	98,67 ± 0,45

Berdasarkan Tabel 5, ResNet-18 menunjukkan performa yang sangat konsisten di seluruh fold. Perhitungan nilai rata-rata akurasi menggunakan formula (1):

$$\text{Acc} = \frac{99.03 + 98.83 + 98.25 + 98.05 + 99.22}{5} \approx 98.67\%$$

Standar deviasi akurasi dihitung menggunakan formula (2):

$$\sigma = \sqrt{\frac{1}{5} \left[(99.03 - 98.67)^2 + (98.83 - 98.67)^2 + (98.25 - 98.67)^2 + (98.05 - 98.67)^2 + (99.22 - 98.67)^2 \right]}$$

$$\sigma = \sqrt{\frac{1}{5} (0.1233 + 0.0244 + 0.1837 + 0.3887 + 0.2966)}$$

$$\sigma = \sqrt{\frac{1.0167}{5}} = \sqrt{0.2033} \approx 0.45\%$$

Nilai standar deviasi yang sangat kecil ($\pm 0,45\%$) mengonfirmasi bahwa performa ResNet-18 stabil dan tidak sensitif terhadap variasi pembagian data. F1-Score rata-rata

sebesar 98,67% ± 0,45% menunjukkan keseimbangan yang baik antara precision dan recall di seluruh fold. Hasil ini konsisten dan saling melengkapi dengan hasil hold-out evaluation yang menunjukkan akurasi 92,25%, membuktikan bahwa model tidak mengalami overfitting terhadap partisi training tertentu.

B. Hasil K-Fold Cross Validation YOLOv8

Tabel 6. Hasil K-Fold Cross Validation (k=5) YOLOv8

Fold	Precision (%)	Recall (%)	mAP@50 (%)	mAP@50-95 (%)	F1-Score (%)
Fold 1	44,93	37,83	37,69	16,72	41,08
Fold 2	33,78	41,43	36,71	16,08	37,21
Fold 3	43,39	45,25	41,66	16,38	44,30
Fold 4	44,42	42,30	39,78	15,67	43,34
Fold 5	52,22	38,16	41,33	18,07	44,10
Rata-rata	43,75 ± 5,88	40,99 ± 2,76	39,43 ± 1,95	16,58 ± 0,82	42,01 ± 2,65

Tabel 6 memperlihatkan hasil K-Fold Cross Validation pada model YOLOv8. Nilai mAP@50 rata-rata dihitung menggunakan formula (1):

$$\overline{\text{mAP}}_{50} = \frac{1}{5} (37.69 + 36.71 + 41.66 + 39.78 + 41.33) \approx 39.43\%$$

Standar deviasi mAP@50 dihitung menggunakan formula (2):

$$\sigma = \sqrt{\frac{1}{5} [(37.69 - 39.43)^2 + (36.71 - 39.43)^2 + (41.66 - 39.43)^2 + (39.78 - 39.43)^2 + (41.33 - 39.43)^2]}$$

$$\sigma = \sqrt{\frac{1}{5} (3.0241 + 7.4228 + 4.9403 + 0.1173 + 3.6036)}$$

$$\sigma = \sqrt{\frac{19.1080}{5}} = \sqrt{3.8216} \approx 1.95\%$$

Variasi antar-fold yang lebih besar pada YOLOv8 (mAP@50 = ±1,95%) dibandingkan ResNet-18 (akurasi = ±0,45%) disebabkan oleh sensitivitas model deteksi objek terhadap distribusi spasial bounding box pada setiap partisi data. Hal ini merupakan karakteristik umum model deteksi objek yang bergantung pada representasi geometri anotasi yang merata di setiap fold. F1-Score rata-rata YOLOv8 sebesar 42,01% mengindikasikan performa deteksi yang moderat pada skenario cross-validation. Perbedaan signifikan dengan hasil hold-out evaluation (mAP@50 = 88,17%) dijelaskan pada analisis berikut.

Analisis Perbedaan Performa K-Fold Cross Validation dan Hold-Out Evaluation

Terdapat perbedaan performa yang signifikan antara hasil K-Fold Cross Validation dan hasil hold-out evaluation yang perlu dijelaskan secara komprehensif. Perbedaan ini disebabkan oleh perbedaan skenario evaluasi yang fundamental, bukan oleh kesalahan atau inkonsistensi eksperimen.

Pada K-Fold Cross Validation, setiap model dilatih menggunakan 80% data dan diuji pada 20% sisanya. Proses ini diulang sebanyak k=5 kali dengan partisi yang berbeda.

Tujuan utama K-Fold adalah mengukur stabilitas model terhadap variasi pembagian data (variance estimation), bukan untuk memaksimalkan performa absolut. Estimasi varians model dihitung sebagai:

$$\text{Var}(\bar{M}) = \frac{\sigma_{\text{fold}}^2}{k} \quad (3)$$

Pada hold-out evaluation, model final dilatih menggunakan 70% data training secara penuh tanpa rotasi partisi K-Fold. Model ini mendapatkan proses training yang lebih lengkap dengan early stopping dan seleksi checkpoint terbaik dari 30 epoch penuh. Hasil evaluasi dilakukan pada data test (15%) yang sepenuhnya terpisah dari proses pelatihan dan seleksi model.

Perbedaan spesifik pada masing-masing model dapat dijelaskan sebagai berikut. Pertama, ResNet-18 K-Fold mencapai akurasi rata-rata 98,67% yang lebih tinggi dari hold-out 92,25%. Hal ini karena pada K-Fold, setiap fold menggunakan 80% data untuk pelatihan (lebih banyak dari 70% pada hold-out) dan pemilihan epoch terbaik berdasarkan performa validasi pada set yang sama menghasilkan estimasi yang cenderung optimistis. Kedua, YOLOv8 K-Fold mencapai mAP@50 rata-rata 39,43% yang lebih rendah dari hold-out 88,17%. Hal ini disebabkan oleh sensitivitas model deteksi objek terhadap distribusi bounding box per fold yang tidak merata karena penggunaan K-Fold standar (bukan stratified), serta pengambilan metrik dari epoch terakhir pelatihan per fold yang belum tentu optimal.

Tabel 7. Perbandingan Skenario K-Fold Cross Validation dan Hold-Out Evaluation

Aspek	K-Fold Cross Validation	Hold-Out Evaluation
Tujuan	Mengukur stabilitas model terhadap variasi data split	Mengukur performa model terbaik pada data uji independen
Data Training	80% per iterasi (rotasi antar fold)	70% (tetap, tanpa rotasi)
Data Testing	20% per iterasi (rotasi antar fold)	15% (tetap, independen dari validasi)
ResNet-18	Akurasi: 98,67% ± 0,45%	Akurasi: 92,25%
YOLOv8	mAP@50: 39,43% ± 1,95%	mAP@50: 88,17%

Dengan demikian, K-Fold Cross Validation dan hold-out evaluation bersifat komplementer, K-Fold mengukur reliabilitas statistik model (apakah performa konsisten lintas distribusi data), sedangkan hold-out mengukur performa model produksi terbaik. Kedua metrik bersama-sama memberikan gambaran yang lebih komprehensif dan jujur tentang kapabilitas sistem.

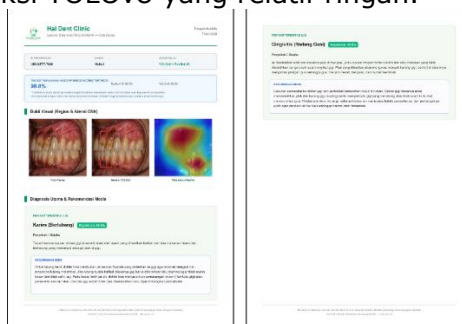
Jika dibandingkan dengan penelitian terdahulu, sistem yang diusulkan memiliki beberapa perbedaan substansial. Oroh [10] menggunakan CNN klasifikasi tunggal dan mencapai akurasi di atas 90%, namun cakupannya terbatas pada dua kelas (karies dan kalkulus) serta tidak dilengkapi komponen interpretabilitas. Sihotang dan Supardi [11] menerapkan ResNet-18 pada citra RGB utuh tanpa tahapan lokalisasi, sehingga fitur latar belakang (bibir, lidah, pipi) turut mempengaruhi bobot klasifikasi dan menurunkan keandalan prediksi pada gambar yang tidak terkontrol. Mahdiyah [12] berhasil memanfaatkan YOLOv8 untuk segmentasi citra panoramik gigi, tetapi bergantung pada hasil rontgen yang hanya tersedia di fasilitas medis. Tabel 4 merangkum perbandingan tersebut.

Tabel 8. Perbandingan dengan Penelitian Terdahulu

Penelitian	Metode	Jumlah Kelas	Interpretabilitas	Input & Akurasi
Oroh [10]	CNN klasifikasi tunggal	2 (Karies, Kalkulus)	Tidak ada	Foto gigi (>90%)
Sihotang & Supardi [11]	ResNet-18 tanpa lokalisasi	Multi-kelas (citra RGB utuh)	Tidak ada	Foto gigi (tidak dilaporkan)
Mahdiyah [12]	YOLOv8 segmentasi	Multi-kelas (panoramik)	Tidak ada	Citra rontgen (tidak dilaporkan)
Penelitian ini	YOLOv8 + ResNet-18 (ensemble)	7 kelas (foto ponsel)	Grad-CAM (heatmap)	Foto ponsel (mAP 88,17%; Akurasi 92,25%)

Berdasarkan Tabel 4 penelitian ini lebih unggul dibandingkan penelitian sebelumnya karena menggunakan pendekatan hybrid YOLOv8 dan ResNet-18 yang tidak hanya melakukan klasifikasi tetapi juga melokalisasi area gigi yang relevan sebelum proses klasifikasi, sehingga meningkatkan ketepatan prediksi. Selain itu, penelitian ini mampu mendeteksi 7 kelas kondisi gigi, lebih banyak dibandingkan penelitian Oroh [10] yang hanya mencakup 2 kelas. Dari sisi implementasi, sistem menggunakan foto gigi dari kamera ponsel yang lebih mudah diakses dibandingkan citra radiografi panoramik yang memerlukan peralatan khusus. Penelitian ini juga memiliki keunggulan dalam interpretabilitas melalui penggunaan Grad-CAM yang menampilkan area citra yang menjadi dasar keputusan model, sehingga hasil prediksi lebih transparan dan mudah dipahami. Keunggulan tersebut didukung oleh performa yang tinggi dengan mAP YOLOv8 sebesar 88,17% dan akurasi klasifikasi 92,25%, menjadikan sistem lebih komprehensif, praktis, dan andal untuk deteksi penyakit gigi.

Dari sisi implementasi aplikasi web, sistem berhasil dioperasikan sebagai platform skrining awal yang fungsional. Pengujian black-box terhadap seluruh fitur antarmuka (unggah gambar, validasi format, deteksi gambar bukan gigi, tampilan hasil diagnosis, unduh laporan PDF, dan responsivitas mobile) menunjukkan bahwa seluruh skenario berjalan sesuai ekspektasi tanpa error. Secara end-to-end, waktu pemrosesan total pipeline mulai dari penerimaan gambar di server, inferensi YOLOv8 (54,1 ms), pemotongan ROI, klasifikasi ResNet-18, hingga generasi heatmap Grad-CAM dan pengiriman respons JSON ke frontend tercatat dalam rentang 2-4 detik pada GPU NVIDIA T4 via Google Colab. Latensi ini mencakup overhead jaringan dan rendering antarmuka, namun tetap berada dalam ambang batas yang dapat diterima untuk aplikasi web skrining klinis. Mayoritas waktu dikonsumsi oleh proses forward pass ResNet-18 dan komputasi gradien Grad-CAM, bukan oleh komponen deteksi YOLOv8 yang relatif ringan.



Gambar 6. Tampilan Antarmuka Aplikasi

4. KESIMPULAN

Penelitian ini telah berhasil mengembangkan dan mengevaluasi sistem deteksi penyakit gigi berbasis deep learning yang mengintegrasikan arsitektur YOLOv8 dan ResNet-18 dalam sebuah pipeline ensemble. Berdasarkan rangkaian eksperimen yang telah dilakukan, diperoleh beberapa kesimpulan. Pertama, model YOLOv8 terbukti mampu melokalisasi area penyakit gigi dengan nilai mAP@50 sebesar 88,17%, menunjukkan bahwa pendekatan object detection efektif untuk mengisolasi Region of Interest dari citra fotografi intraoral yang mengandung berbagai noise visual. Kedua, arsitektur ResNet-18 mencapai akurasi klasifikasi 92,25% dan F1-Score 92,37% pada data uji, dengan gap overfitting yang sangat minim (1,53%) yang membuktikan kemampuan generalisasi model yang baik. Ketiga, meskipun performa global cukup tinggi, masih terdapat tantangan pada klasifikasi kelas Calculus dan Gingivitis yang sering tertukar akibat kemiripan visual dan koeksistensi klinis kedua kondisi tersebut. Keempat, validasi statistik menggunakan K-Fold Cross Validation (k=5) mengonfirmasi stabilitas model: ResNet-18 mencapai akurasi rata-rata 98,67% ± 0,45% dan F1-Score rata-rata 98,67% lintas fold, sedangkan YOLOv8 memperoleh mAP@50 rata-rata 39,43% ± 1,95%. Perbedaan antara hasil K-Fold dan hold-out evaluation disebabkan oleh perbedaan skenario evaluasi yang bersifat komplementer—K-Fold mengukur stabilitas lintas distribusi data, sedangkan hold-out mengukur performa model produksi terbaik—keduanya saling melengkapi dan mengonfirmasi reliabilitas sistem.

Untuk pengembangan lebih lanjut, penelitian ini merekomendasikan beberapa arah. Implementasi fitur human-in-the-loop perlu dipertimbangkan agar pengguna dapat mengoreksi bounding box secara manual sebelum proses klasifikasi berjalan. Strategi cropping juga perlu disempurnakan dengan penambahan margin piksel dinamis agar konteks jaringan di sekitar lesi turut tersertakan dalam area potongan. Di samping itu, variasi dataset dapat ditingkatkan dengan menambahkan gambar yang diambil pada kondisi pencahayaan ekstrem dan sudut pengambilan yang asimetris. Eksplorasi arsitektur single-stage yang mampu mendeteksi sekaligus mengklasifikasikan dalam satu model juga layak dipertimbangkan untuk menekan latensi dan risiko cascading error antar komponen pipeline.

5. REFERENCES

- [1] K. Q. Ayun dan A. C. Adisasmita, "Perbandingan Faktor Risiko Gangguan Kesehatan Gigi dan Mulut Pada Usia 20-59 Tahun Berdasarkan Wilayah di Indonesia: Analisis Survei Kesehatan Indonesia (SKI) Tahun 2023," *J. Epidemiol. Kesehat. Indones.*, vol. 9, no. 3, 2025, doi: 10.7454/epidkes.v9i3.1154.
- [2] K. P. Sukma, "Identifikasi Kerentanan Karies Gigi Berdasarkan Aktivitas Mikroflora Normal Pada Sampel Saliva," *J. Kesehat. Tambusai*, vol. 6, no. 1, hal. 1656-1663, 2025, doi: 10.31004/jkt.v6i1.42082.
- [3] M. syarif Wicaksono, M. Arsanti, dan H. A. Hartanto, "Indonesia," vol. 7, no. 2, hal. 1-10, 2024.
- [4] A. Y. Riyadi, W. Soewondo, dan E. Riyanti, "Prevalensi anomali gigi pada Down syndrome : scoping review," *Padjadjaran J. Dent. Res. Students*, vol. 9, no. 1, hal. 15-30, 2025, doi: 10.24198/pjdrs.v9i1.57122.
- [5] R. Indraswary *et al.*, "EDUKASI STRUKTUR ANATOMI DAN TAHAPAN ERUPSI GIGI PADA MASYARAKAT KALIGAWA SEMARANG: PENDEKATAN PENYULUHAN PENDAHULUAN Kesadaran masyarakat Indonesia akan kesehatan gigi dan mulut masih tergolong rendah [1][2]. Penyakit seperti karies gigi dan penyaki," vol. 8, no. 3, hal. 305-313, 2025.
- [6] F. Nuraeni, Y. Septiana, dan Y. Siti Wahyuni, "Implementasi Metode Inferensi

- Certainty Factor Pada Sistem Pakar Diagnosa Penyakit Gigi," *J. Algoritma.*, vol. 22, no. 2, hal. 923–932, 2025, doi: 10.33364/algoritma/v.22-2.2124.
- [7] A. Jauhar, Istiadi, dan F. Marisa, "Sistem pakar identifikasi penyakit gigi berlubang menggunakan metode neural network dengan algoritma backpropagation," vol. 4, no. 1, hal. 48–58, 2025.
- [8] A. Ramdhani, D. Fatmasari, dan M. Djamil, "Oral Detect-Tri To Improve Dental and Oral Diseases Detection Behaviour in Houswives," *JDHT J. Dent. Hyg. Ther.*, vol. 4, no. 2, hal. 132–139, 2023, doi: 10.36082/jdht.v4i2.1276.
- [9] F. Izza, K. Khotimah, N. Sadilah Arif, dan N. Rifat Aqillah, "Literature Review: Sistem Cerdas untuk Klasifikasi Penyakit Gigi Berdasarkan Citra Rontgen Metode: Convolutional Neural Network (CNN) dan Decision Support System (DSS)," *J. Artif. Intel. dan Sist. Penunjang Keputusan*, vol. 2, no. 2, hal. 142–145, 2024.
- [10] E. dan C. L. Oroh, "Multidisciplinary Science Klasifikasi Penyakit Gigi Karies Dan Kalkulus Menggunakan," vol. 1, no. 4, hal. 757–765, 2023.
- [11] G. M. A. Sihotang dan J. Supardi, "Pengembangan Model CNN ResNet-18 untuk Klasifikasi Kondisi Gigi Berbasis Citra RGB sebagai Solusi Diagnostik Digital Development of CNN ResNet-18 Model for RGB Image-Based Dental Condition Classification as a Digital Diagnostic Solution," *J. Pendidik. dan Teknol. Indones.*, vol. 4, no. 12, hal. 18, 2024.
- [12] L. Mahdiyah, S. Oktamuliani, dan W. L. Putri, "Penerapan Algoritma Deep Learning YOLOv8 pada Platform Roboflow untuk Segmentasi Citra Panoramik," *J. Fis. Unand*, vol. 14, no. 3, hal. 228–234, 2025, doi: 10.25077/jfu.14.3.228-234.2025.
- [13] W. Sofiyah, B. S. Negara, M. Irsyad, I. Iskandar, dan F. Yanto, "Lung Disease Detection Using Gradient-Weighted Class Activation Mapping (Grad-CAM)," *J. Artif. Intell. Softw. Eng.*, vol. 5, no. 2, hal. 720–730, 2025, doi: 10.30811/jaise.v5i2.7041.
- [14] R. I. dan R. Rahmadewi, "DETEKSI ANATOMI GIGI MENGGUNAKAN ARSITEKTUR YOLOV8 (YOU ONLY LOOK ONCE) UNTUK PENGENALAN JENIS-JENIS GIGI," *J. Huk. dan Etika Kesehat.*, vol. 9, no. 3, hal. 4870–4877, 2024.
- [15] D. J. Marcelleno dan M. P. K. Putra, "Performance Evaluation of Yolov8 in Real-Time Vehicle Detection in Various Environmental Conditions," *J. Tek. Inform.*, vol. 6, no. 1, hal. 269–279, 2025, doi: 10.52436/1.jutif.2025.6.1.3916.
- [16] I. P. Harahap, "Deteksi Objek pada Citra Menggunakan Model YOLO," *J. Nas. Teknol. Inf. dan Apl.*, vol. 2, hal. 469–474, 2024.
- [17] F. L. Pakpahan, J. S. Sembiring, T. B. Abellista, dan E. Indra, "Integration of YOLOv8 and FastAPI for Early Detection of Nail Diseases," *Sinkron*, vol. 9, no. 2, hal. 978–986, 2025, doi: 10.33395/sinkron.v9i2.14796.
- [18] K. D. Dharmasaputra dan B. P. Hartato, "Evaluation of YOLOv8 and Centroid Tracking in Vehicle Detection , Classification , and Counting System," vol. 9, no. 5, hal. 2310–2319, 2025.
- [19] Maria Bestarina Laili, Raihan Alfariji, James Tri Septiono, Muhammad Farid Idlal, dan Egi Sunardi, "Implementasi Sistem Pendeteksi Buku dengan YOLOv8," *Epsil. J. Electr. Eng. Technol.*, vol. 23, no. 1, hal. 35–43, 2025, doi: 10.55893/epsilon.v23i1.130.