



Deteksi Dini Karies dan Karang Gigi Berbasis Android Menggunakan CNN MobileNetV2 dan Transfer Learning

Dafa Fadhilah Matondang^{1*}, Mahardika Abdi Prawira Tanjung²

^{1*,2}Program Studi Sistem Informasi, Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara, Medan, Indonesia

Email: ^{1*}dafafadhilah112@gmail.com, ²mahardikaabdiprawira@umsu.ac.id

Abstrak

Karies dan karang gigi merupakan gangguan kesehatan gigi yang umum dan dapat berkembang menjadi infeksi serta kerusakan gigi yang lebih parah apabila terlambat ditangani, sementara akses terhadap pemeriksaan rutin masih terbatas. Penelitian ini bertujuan mengembangkan sistem klasifikasi citra gigi sebagai alat bantu deteksi dini karies, karang gigi, dan gigi normal menggunakan Convolutional Neural Network (CNN) arsitektur MobileNetV2 dengan pendekatan *transfer learning* yang diimplementasikan pada aplikasi Android. *Dataset* Kaggle berjumlah 3.539 citra; pemeriksaan *hashing* MD5 menemukan 467 citra duplikat pada kelas karang gigi sehingga *dataset* akhir berjumlah 2.829 citra yang dibagi dengan rasio 70:15:15. Citra diubah menjadi 224×224 piksel dan data latih diaugmentasi. Pelatihan dilakukan dalam dua tahap, yaitu pelatihan *classifier* dengan *base model* dibekukan dan *fine-tuning* sekitar 30 lapisan terakhir. Evaluasi pada 425 citra uji menghasilkan akurasi 98,59% dengan *precision*, *recall*, dan *F1-score* yang seimbang pada ketiga kelas, dengan enam citra salah diklasifikasikan, sebagian besar antara karies dan karang gigi. Model diintegrasikan melalui FastAPI ke aplikasi Android berbasis Flutter. Pengujian pada citra di luar *dataset* menunjukkan *confidence* yang lebih rendah akibat *domain shift*. Sistem ini berpotensi menjadi alat bantu klasifikasi awal, bukan pengganti diagnosis tenaga medis, dan deduplikasi data perlu dilakukan sebelum pembagian *dataset*.

Kata Kunci: Android, CNN, karies dan karang gigi, MobileNetV2, transfer learning.

PENDAHULUAN

Kesehatan gigi dan mulut masih sering kurang mendapat perhatian, padahal dampaknya dapat memengaruhi kualitas hidup seseorang dalam jangka panjang (World Health Organization, 2025). Karies gigi dan karang gigi merupakan dua gangguan yang paling umum. Karies adalah kerusakan jaringan keras gigi akibat aktivitas bakteri yang menghasilkan asam dari sisa makanan (The Dental Association of Thailand, 2023), sedangkan karang gigi terbentuk dari plak yang mengalami mineralisasi atau pengerasan pada permukaan gigi (Warreth, 2023). Apabila tidak ditangani sejak dini, kedua kondisi tersebut dapat menimbulkan nyeri, bau mulut, peradangan gusi, infeksi, hingga kerusakan gigi yang lebih serius.

Deteksi dini penting karena penanganan pada tahap awal umumnya lebih mudah dan efektif dibandingkan setelah penyakit berkembang lebih parah (GBD 2021 Oral Disorders Collaborators, 2025). Namun, pemeriksaan gigi secara langsung bergantung pada kunjungan ke layanan kesehatan, sedangkan tidak semua individu melakukannya secara rutin. Keterbatasan akses layanan, biaya perawatan, kesibukan sehari-hari, dan rendahnya kesadaran masyarakat menyebabkan banyak kasus baru diketahui setelah muncul keluhan yang mengganggu (Hasan dkk., 2024; Kementerian Kesehatan Republik Indonesia, 2022).

Perkembangan internet dan perangkat seluler membuka peluang untuk menghadirkan solusi berbasis teknologi yang lebih mudah diakses (Erku dkk., 2023). Pengolahan citra digital dapat dimanfaatkan untuk membantu identifikasi awal kondisi gigi melalui foto (Perez-Meana dan Nakano-Miyatake, 2025). Pengolahan citra digital adalah pengolahan gambar menggunakan komputer agar informasi visual di dalamnya dapat dianalisis secara sistematis (Gonzalez dan Woods, 2021). Citra tersusun atas piksel yang merepresentasikan warna atau intensitas cahaya, sehingga citra gigi memuat informasi seperti warna, tekstur, bentuk, dan pola permukaan yang dapat dipelajari komputer. Secara visual, karies dapat tampak sebagai perubahan warna, bercak gelap, atau lubang pada struktur gigi, sedangkan karang gigi tampak sebagai lapisan atau endapan yang menempel pada permukaan gigi, terutama di sekitar garis gusi (Warreth, 2023). Perbedaan visual ini menjadi dasar klasifikasi citra, yaitu pengelompokan gambar ke dalam kelas tertentu berdasarkan ciri visualnya. Berbeda dengan *object detection*, klasifikasi citra hanya menghasilkan label kelas untuk keseluruhan gambar tanpa menunjukkan lokasi kelainan. Hasil klasifikasi

juga dipengaruhi kualitas gambar, pencahayaan, sudut pengambilan, dan fokus kamera sehingga hanya dapat dipahami sebagai bantuan berbasis citra, bukan diagnosis medis.

Dalam bidang ini, Convolutional Neural Network (CNN) banyak digunakan karena mampu mengenali pola visual secara otomatis melalui ekstraksi fitur (Zhao dkk., 2024). Lapisan awal mempelajari fitur sederhana seperti garis dan tepi, sedangkan lapisan yang lebih dalam mempelajari fitur yang lebih kompleks seperti bentuk dan tekstur. Komponen utamanya meliputi *convolution layer* yang menghasilkan *feature map* menggunakan filter, fungsi aktivasi ReLU yang menambahkan sifat non-linear, *pooling layer* yang mengurangi ukuran *feature map*, serta *fully connected layer* dan *output layer* yang menentukan kelas prediksi beserta nilai *confidence*. MobileNetV2 adalah arsitektur CNN yang dikembangkan untuk perangkat seluler dan perangkat dengan sumber daya komputasi terbatas, dengan tujuan memaksimalkan akurasi sekaligus meminimalkan jumlah parameter dan operasi komputasi (Sandler dkk., 2018). Inovasi utamanya adalah blok *inverted residual with linear bottleneck* yang memanfaatkan *depthwise separable convolution*, yaitu memisahkan konvolusi menjadi *depthwise convolution* untuk menyaring setiap kanal dan *pointwise convolution* (1×1) untuk menggabungkan kembali fitur (Zhao dkk., 2024). Dimensi data diekspansi pada lapisan tengah kemudian dikompresi secara linear sehingga informasi penting tidak hilang selama ekstraksi pola.

Transfer learning adalah teknik *deep learning* yang memanfaatkan model yang telah dilatih sebelumnya untuk menyelesaikan tugas baru (Iman dkk., 2023). Teknik ini bermanfaat ketika data terbatas karena model pra-latih telah memiliki kemampuan dasar mengenali tepi, warna, tekstur, dan bentuk. Penerapannya umumnya dilakukan melalui dua pendekatan, yaitu *feature extraction*, yang menggunakan model pra-latih sebagai pengekstraksi fitur dengan lapisan klasifikasi baru, dan *fine-tuning*, yang melatih kembali sebagian lapisan model pra-latih agar fitur yang telah dipelajari menyesuaikan dengan karakteristik data baru tanpa menghilangkan kemampuan dasarnya.

Penelitian terdahulu menunjukkan bahwa CNN efektif untuk mendeteksi karies gigi (Lee dkk., 2018), bahwa foto yang diambil menggunakan *smartphone* dapat dimanfaatkan untuk deteksi karies berbasis *deep learning* (Thanh dkk., 2022), dan bahwa arsitektur MobileNetV2 memiliki performa tinggi pada klasifikasi karies gigi (Wahyuningsih dkk., 2024). Beberapa penelitian terdahulu yang relevan dengan klasifikasi dan deteksi kondisi gigi berbasis citra dirangkum pada Tabel 1.

Tabel 1. Penelitian terdahulu yang relevan

Peneliti (tahun)	Metode dan objek	Hasil utama
Lee dkk. (2018)	CNN untuk deteksi dan diagnosis karies gigi	AUC 0,917 (premolar), 0,890 (molar), dan 0,845 (gabungan premolar dan molar)
Thanh dkk. (2022)	<i>Deep learning</i> untuk deteksi karies pada foto intraoral yang diambil dengan <i>smartphone</i>	Klasifikasi cavitated dan non-cavitated caries: Faster R-CNN 87,4% dan YOLOv3 83,4%
Park dkk. (2023)	CNN untuk klasifikasi penyakit periodontal pada citra gigi berwarna	Akurasi model yang diusulkan 74,54%, lebih tinggi daripada ResNet152 (63,09%)
Makarim dkk. (2024)	Klasifikasi kondisi gigi pada citra intraoral	YOLOv5 82%, MobileNetV2 80%, dan IONet 70%
Wahyuningsih dkk. (2024)	Perbandingan EfficientNet-B0, MobileNetV2, ResNet-50, dan InceptionV3 untuk klasifikasi karies gigi	ResNet-50 0,99; EfficientNet-B0 dan MobileNetV2 0,98; InceptionV3 0,97 (akurasi)

Sumber: Diolah dari penelitian-penelitian yang dirujuk.

Secara umum, penelitian-penelitian tersebut menunjukkan bahwa CNN dan *transfer learning* efektif untuk analisis citra gigi, bahwa foto yang diambil dengan *smartphone* dapat dimanfaatkan, dan bahwa MobileNetV2 memiliki performa yang tinggi serta efisien untuk klasifikasi karies gigi. Perbedaan *dataset*, jenis citra, jumlah kelas, dan tugas klasifikasi menyebabkan nilai performa antarpemelitian tidak dapat dibandingkan secara langsung.

Meskipun demikian, hasil penelitian yang ditelaah dalam studi ini terutama dilaporkan dalam bentuk performa model, sedangkan integrasi model ke dalam aplikasi seluler yang dapat langsung digunakan oleh pengguna awam, serta pemeriksaan validitas *dataset*, khususnya keberadaan citra duplikat yang dapat menimbulkan kebocoran data (*data leakage*), belum banyak diuraikan secara bersamaan. Penelitian ini berupaya mengisi celah tersebut dengan mengembangkan klasifikasi citra gigi ke dalam tiga

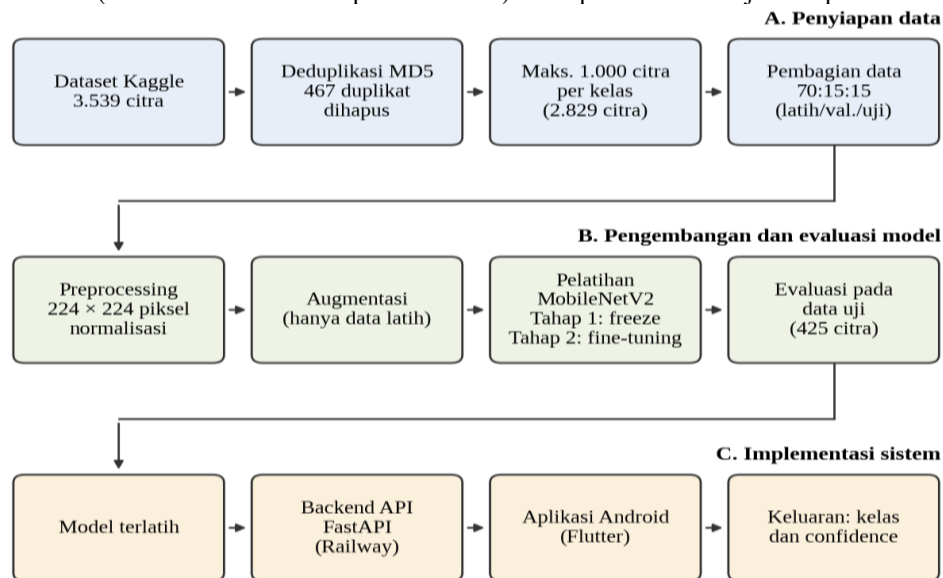
kelas, yaitu karies, karang gigi, dan gigi normal, yang dilengkapi pemeriksaan duplikasi citra sebelum pembagian *dataset* dan diimplementasikan pada aplikasi Android.

MobileNetV2 dipilih karena dirancang dengan struktur jaringan yang ringan dan efisien untuk perangkat dengan sumber daya terbatas (Sandler dkk., 2018; Zhao dkk., 2024), sedangkan *transfer learning* memanfaatkan bobot hasil pelatihan pada *dataset* berskala besar sehingga membantu mengatasi keterbatasan jumlah data pelatihan (Iman dkk., 2023). Tujuan penelitian ini adalah merancang dan membangun model klasifikasi citra gigi ke dalam tiga kelas menggunakan CNN MobileNetV2 berbasis *transfer learning*, mengevaluasi performa model berdasarkan *accuracy*, *precision*, *recall*, *F1-score*, dan *confusion matrix*, serta mengimplementasikan model tersebut pada aplikasi Android melalui sistem berbasis API. Sistem yang dikembangkan diposisikan sebagai alat bantu klasifikasi awal berbasis citra, bukan sebagai alat diagnosis medis.

METODE

Desain Penelitian

Penelitian ini merupakan penelitian terapan (applied research) dengan pendekatan kuantitatif dan eksperimental, yang dilengkapi pendekatan rekayasa perangkat lunak. Pendekatan kuantitatif digunakan karena pengujian model dilakukan melalui metrik numerik, sedangkan pendekatan eksperimental diwujudkan dalam tahapan *training*, *validation*, dan *testing* yang sistematis pada *dataset* citra gigi. Sistem terdiri atas dua bagian utama, yaitu pengembangan model *machine learning* (pengumpulan *dataset*, *preprocessing*, pelatihan CNN MobileNetV2 dengan *transfer learning*, dan evaluasi) serta pengembangan aplikasi Android (*backend* FastAPI dan aplikasi Flutter). Alur penelitian ditunjukkan pada Gambar 1.



Gambar 1. Alur penelitian dan pengembangan sistem

Dataset dan Pembagian Data

Dataset awal diperoleh dari Kaggle sebanyak 3.539 citra yang terbagi dalam tiga kelas, yaitu 1.088 citra karies, 1.296 citra karang gigi, dan 1.155 citra normal. Data tidak diambil langsung dari pasien oleh peneliti. Label mengikuti pengelompokan citra pada folder kelas dari sumber *dataset*, dan citra yang tidak menampilkan area gigi secara jelas atau berkualitas sangat rendah tidak digunakan. Duplikasi citra diperiksa menggunakan *hashing* MD5, yaitu membandingkan nilai *hash* setiap berkas untuk menemukan citra dengan konten identik meskipun nama berkasnya berbeda. Jumlah citra pada kelas karies dan normal kemudian dibatasi maksimum 1.000 citra per kelas agar distribusi antarkelas relatif seimbang.

Dataset dibagi per kelas menjadi data latih, data validasi, dan data uji dengan rasio 70:15:15, yang dilakukan setelah deduplikasi dan sebelum augmentasi sehingga tidak ada citra yang sama pada lebih dari satu subset. Rasio ini dipilih karena kelas karang gigi hanya memiliki 829 citra unik; dengan porsi 10% untuk validasi dan pengujian, kelas tersebut hanya memiliki sekitar 83 citra pada setiap subset sehingga satu kesalahan prediksi saja akan mengubah akurasi kelas secara cukup besar. Data validasi digunakan untuk memantau pelatihan dan memilih bobot terbaik, sedangkan data uji tidak dilibatkan dalam pelatihan maupun pemilihan model dan hanya digunakan setelah seluruh proses pelatihan selesai.

Preprocessing dan Augmentasi

Preprocessing menyeragamkan ukuran, skala warna, dan format citra sebelum dimasukkan ke model (Xu dkk., 2023). Tahap yang umum adalah *resizing*, misalnya ke 224 × 224 piksel untuk

MobileNetV2, dan *rescaling* nilai piksel dari rentang 0–255 ke rentang yang lebih kecil agar komputasi lebih stabil. Augmentasi data, seperti rotasi, *flip* horizontal, pergeseran, dan *zoom*, menambah variasi data latih agar model mampu mengenali objek pada berbagai kondisi pengambilan gambar tanpa mengubah makna citra (Xu dkk., 2023).

Pada penelitian ini, seluruh citra diubah ukurannya menjadi 224×224 piksel dengan tiga kanal warna RGB sesuai masukan MobileNetV2, dikonversi menjadi larik numerik, dan nilai pikselnya dinormalisasi dari rentang 0–255. Augmentasi berupa rotasi, *zoom*, dan *flip* horizontal diterapkan secara dinamis selama pelatihan menggunakan ImageDataGenerator pada TensorFlow/Keras (Medavarapu, 2024), hanya pada data latih. Data validasi dan data uji hanya melalui *resize* dan normalisasi agar mencerminkan kondisi citra yang sebenarnya.

Arsitektur Model dan Proses Pelatihan

MobileNetV2 dengan bobot pra-latih ImageNet digunakan sebagai *base model* untuk mengekstraksi fitur visual. *Classification head* bawaan tidak digunakan karena dirancang untuk klasifikasi ImageNet dengan banyak kelas. Sebagai gantinya ditambahkan Global Average Pooling yang meringkas hasil ekstraksi fitur menjadi vektor, lapisan Dense yang mempelajari hubungan antara fitur dan kelas, Dropout untuk mengurangi risiko *overfitting*, serta *output layer* yang menghasilkan prediksi untuk tiga kelas, yaitu karies, karang gigi, dan normal.

Pelatihan dilakukan dalam dua tahap (Tabel 2). Pada tahap pertama, seluruh lapisan *base model* dibekukan dan hanya lapisan *classifier* tambahan yang dilatih. Pada tahap kedua (*fine-tuning*), sekitar 30 lapisan terakhir *base model* dibuka dan dilatih kembali dengan learning rate yang lebih kecil. Tiga *callback* memanfaatkan data validasi, yaitu EarlyStopping yang menghentikan pelatihan apabila *val_loss* tidak membaik selama lima epoch berturut-turut, ModelCheckpoint yang menyimpan bobot dengan *val_accuracy* terbaik, dan ReduceLROnPlateau yang menurunkan learning rate ketika performa stagnan.

Tabel 2. Konfigurasi pelatihan model

Parameter	Tahap 1 (pelatihan <i>classifier</i>)	Tahap 2 (<i>fine-tuning</i>)
Ukuran masukan	$224 \times 224 \times 3$	$224 \times 224 \times 3$
Lapisan yang dilatih	Global Average Pooling, Dense, Dropout (<i>base model</i> dibekukan)	Sekitar 30 lapisan terakhir <i>base model</i> dibuka
Optimizer	Adam	Adam
Fungsi <i>loss</i>	Categorical crossentropy	Categorical crossentropy
Learning rate	1e-3	1e-5
Epoch maksimum	20	15
<i>Callback</i>	EarlyStopping (<i>val_loss</i> , 5 epoch berturut-turut), ModelCheckpoint (<i>val_accuracy</i> terbaik), ReduceLROnPlateau	

Sumber: Hasil penelitian.

Implementasi Sistem

Model terlatih disajikan melalui *backend* RESTful API berbasis FastAPI, yaitu framework Python untuk membangun API secara cepat dan efisien (Azhari, 2022), dengan pertukaran data dalam format JSON (Iqbal, 2023). *Backend* dijalankan menggunakan Uvicorn, menerima citra dari aplikasi, melakukan *preprocessing* berupa *resize* dan normalisasi, menjalankan prediksi model MobileNetV2, dan mengembalikan hasil klasifikasi beserta nilai *confidence*. Aplikasi Android dikembangkan menggunakan Flutter dengan fitur utama pemilihan citra gigi dari galeri perangkat, pengiriman citra ke API, dan penampilan hasil klasifikasi. Prediksi dijalankan di sisi server sehingga aplikasi tidak menjalankan model secara langsung pada perangkat pengguna.

Evaluasi Model

Kinerja model klasifikasi umumnya dievaluasi menggunakan *confusion matrix* beserta metrik turunannya (Carrington dkk., 2023). *Confusion matrix* memetakan prediksi model terhadap label sebenarnya ke dalam true positive (TP), true negative (TN), false positive (FP), dan false negative (FN) (Luque dkk., 2023). Berdasarkan keempat nilai tersebut dihitung empat metrik utama, yaitu *accuracy*, *precision*, *recall*, dan *F1-score*, sebagaimana dinyatakan pada Persamaan (1) sampai (4).

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN}) \quad (1)$$

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP}) \quad (2)$$

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN}) \quad (3)$$

$$\text{F1-score} = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}) \quad (4)$$

Accuracy menyatakan rasio prediksi yang benar terhadap seluruh data uji. *Precision* menyatakan rasio prediksi positif yang benar terhadap seluruh prediksi positif, sedangkan *recall* menyatakan rasio data positif yang berhasil dikenali terhadap seluruh data yang sebenarnya positif. *F1-score* merupakan rata-rata

harmonik *precision* dan *recall*. Pada klasifikasi tiga kelas, *precision*, *recall*, dan *F1-score* dihitung untuk setiap kelas dengan memperlakukan kelas yang dievaluasi sebagai kelas positif, kemudian dirangkum dalam macro average dan weighted average.

Evaluasi dilakukan pada data uji yang tidak digunakan selama pelatihan maupun validasi, menggunakan *accuracy*, *precision*, *recall*, dan *F1-score* pada setiap kelas, serta *confusion matrix* untuk melihat distribusi kesalahan prediksi (Persamaan 1–4). Akurasi minimal 80% pada data uji ditetapkan sebagai target performa sistem. Selain itu, sistem diuji menggunakan beberapa citra gigi di luar *dataset* yang diambil dengan kamera ponsel untuk mengamati kemampuan generalisasi pada kondisi yang mendekati penggunaan nyata, serta diuji integrasinya mulai dari pengunggahan citra hingga penampilan hasil pada aplikasi.

HASIL DAN PEMBAHASAN

Hasil Pengumpulan dan Pembersihan Data

Pemeriksaan *dataset* mentah menemukan dua permasalahan. Pertama, terdapat inkonsistensi penamaan folder kelas, yaitu dua folder yang tampak identik secara visual tetapi berbeda secara teknis karena tambahan spasi pada nama folder (misalnya “caries” dan “caries”), sehingga sebagian citra tidak terbaca dengan benar karena dianggap sebagai kelas yang berbeda. Kedua, ditemukan duplikasi citra pada kelas karang gigi: dari 1.296 citra terdapat 467 citra duplikat sehingga hanya 829 citra unik yang tersisa. Citra duplikat berpotensi menyebabkan kebocoran data (*data leakage*), yaitu citra yang identik muncul pada data latih dan data uji sehingga model dapat “menghafal” citra tersebut alih-alih mempelajari pola visual secara umum.

Perbaikan dilakukan dengan menyeragamkan nama folder kelas, mendeteksi dan menghapus citra duplikat menggunakan MD5 sebelum pembagian *dataset*, membagi *dataset* setelah deduplikasi, dan memverifikasi ulang menggunakan perhitungan *hash* bahwa tidak ada citra yang sama antarsubset. Setelah deduplikasi terdapat 3.072 citra unik, dan setelah kelas karies dan normal dibatasi masing-masing 1.000 citra diperoleh *dataset* akhir sebanyak 2.829 citra. Komposisi *dataset* ditampilkan pada Tabel 3. Hasil verifikasi menunjukkan tidak ada lagi citra yang sama pada data latih, data validasi, maupun data uji (*overlap* = 0 pada seluruh kombinasi), sehingga pelatihan dan evaluasi berlangsung pada kondisi data yang valid dan bebas dari kebocoran data.

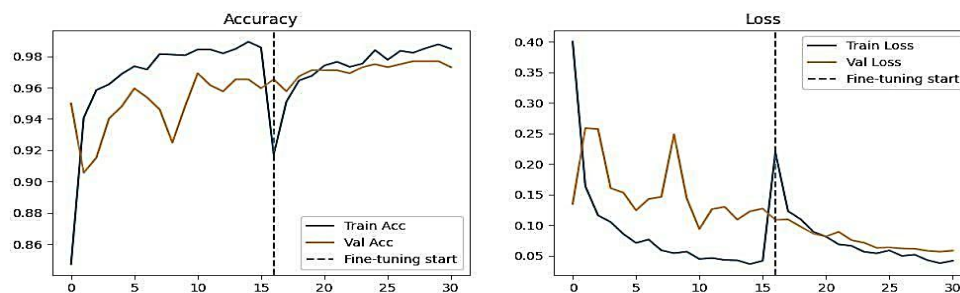
Tabel 3. Komposisi *dataset* sebelum dan sesudah deduplikasi serta pembagian data

Kelas	<i>Dataset</i> mentah	Duplikat dihapus	Digunakan	Data latih (70%)	Data validasi (15%)	Data uji (15%)
Karang gigi	1.296	467	829	580	124	125
Karies	1.088	0	1.000	700	150	150
Normal	1.155	0	1.000	700	150	150
Total	3.539	467	2.829	1.980	424	425

Sumber: Hasil penelitian.

Hasil Pelatihan Model

Gambar 2 menampilkan grafik *accuracy* dan *loss* selama pelatihan. Garis putus-putus menandai awal *fine-tuning*, yang pada grafik terjadi sekitar epoch ke-16, dan pelatihan tahap kedua berlangsung hingga sekitar epoch ke-30. Model menunjukkan tren peningkatan *accuracy* dan penurunan *loss* yang konsisten pada data latih maupun data validasi tanpa indikasi *overfitting* yang signifikan. Pada awal *fine-tuning* tampak penurunan *accuracy* dan kenaikan *loss* data latih secara sementara, yang kemudian pulih seiring penyesuaian bobot pada lapisan yang dibuka. Menjelang akhir pelatihan, berdasarkan pembacaan grafik, *accuracy* data latih dan validasi berada pada kisaran 0,97–0,99 dengan selisih yang kecil, sedangkan *loss* data validasi menurun hingga sekitar 0,06. Hal ini menunjukkan bahwa pelatihan dua tahap, yaitu *freeze* dilanjutkan *fine-tuning*, efektif membantu model mempelajari pola visual citra karies, karang gigi, dan normal.



Gambar 2. Grafik *accuracy* dan *loss* selama pelatihan (garis putus-putus menandai awal *fine-tuning*)

Hasil Evaluasi Model

Evaluasi pada 425 citra uji menghasilkan *accuracy* sebesar 98,59%. Sebanyak 419 citra diklasifikasikan dengan benar, yaitu 121 citra karang gigi, 149 citra karies, dan 149 citra normal, sedangkan 6 citra salah diklasifikasikan. Rincian per kelas pada Tabel 4 menunjukkan nilai *precision*, *recall*, dan *F1-score* yang tinggi dan seimbang, dengan *F1-score* 0,98 untuk karang gigi dan karies serta 0,99 untuk normal.

Tabel 4. Hasil classification report pada data uji

Kelas	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>	<i>Support</i>
Karang gigi	0,99	0,97	0,98	125
Karies	0,97	0,99	0,98	150
Normal	0,99	0,99	0,99	150
<i>Accuracy</i>			0,99	425
<i>Macro average</i>	0,99	0,98	0,99	425
<i>Weighted average</i>	0,99	0,99	0,99	425

Sumber: Hasil penelitian.

Distribusi kesalahan prediksi ditunjukkan oleh *confusion matrix* pada Tabel 5. Sebanyak 3 citra karang gigi diprediksi sebagai karies dan 1 citra karang gigi diprediksi sebagai normal, sedangkan 1 citra karies diprediksi sebagai karang gigi dan 1 citra normal diprediksi sebagai karies. Dengan demikian, 4 dari 6 kesalahan melibatkan kelas karies dan karang gigi. *Recall* terendah dimiliki kelas karang gigi (0,97), sedangkan *precision* terendah dimiliki kelas karies (0,97) karena kelas ini menerima 3 prediksi keliru dari karang gigi dan 1 dari normal.

Tabel 5. *Confusion matrix* pada data uji (jumlah citra)

Kelas aktual	Kelas prediksi			Jumlah
	Karang gigi	Karies	Normal	
Karang gigi	121	3	1	125
Karies	1	149	0	150
Normal	0	1	149	150

Sumber: Hasil penelitian.

Hasil Pengujian pada Citra di Luar Dataset

Model juga diuji menggunakan beberapa citra gigi yang diambil langsung dengan kamera ponsel di luar *dataset* penelitian. Model tetap memberikan prediksi yang sesuai pada sebagian besar citra uji, tetapi dengan *confidence* yang cenderung lebih rendah dan bervariasi dibandingkan hasil pada data uji dari *dataset*. Pada salah satu citra uji, model memprediksi kelas normal dengan *confidence* yang berbeda antarpercobaan, bergantung pada sudut, jarak, dan pencahayaan pengambilan gambar. Kondisi ini mengindikasikan adanya *domain shift* antara citra pelatihan, yang sebagian besar berupa citra *close-up* dengan kondisi terkontrol, dan citra yang diambil secara kasual oleh pengguna. Selain itu, karena model bersifat *closed-set* yang hanya mengenal tiga kelas, model belum mampu menolak citra non-gigi.

Hasil Implementasi Backend API dan Aplikasi Android

Backend FastAPI dijalankan menggunakan Uvicorn dengan satu *endpoint* utama, yaitu */predict*, yang menerima citra dalam format multipart/form-data dan mengembalikan JSON berisi kelas prediksi, nilai *confidence*, serta probabilitas untuk setiap kelas. Pengujian *endpoint* menggunakan dokumentasi interaktif Swagger UI dan permintaan HTTP langsung menunjukkan bahwa API mampu menerima citra, menjalankan klasifikasi, dan mengembalikan respons sesuai rancangan. *Backend* kemudian *di-deploy* pada layanan *hosting* Railway agar dapat diakses publik melalui internet dan diintegrasikan dengan aplikasi Android berbasis Flutter. Pengguna memilih citra gigi dari galeri, citra dikirim ke API, dan aplikasi menampilkan kelas kondisi gigi beserta nilai *confidence* (Gambar 3). Antarmuka juga memuat keterangan bahwa hasil yang ditampilkan bukan diagnosis medis resmi. Pengujian integrasi menunjukkan bahwa aplikasi Android, *backend* API, dan model klasifikasi dapat berjalan dan saling terhubung dengan baik.



Gambar 3. Contoh tampilan hasil klasifikasi pada aplikasi Android

Pembahasan

Accuracy sebesar 98,59% dengan *precision*, *recall*, dan *F1-score* yang seimbang menunjukkan bahwa kombinasi MobileNetV2 dan strategi *transfer learning* dua tahap mampu membedakan ketiga kelas dengan baik pada *dataset* yang digunakan, serta melampaui target akurasi minimal 80% yang ditetapkan pada tahap perancangan. Hasil ini sejalan dengan teori bahwa bobot pra-latih ImageNet telah menyediakan fitur visual umum seperti tepi, tekstur, dan bentuk (Iman dkk., 2023), sehingga *fine-tuning* pada sekitar 30 lapisan terakhir cukup untuk menyesuaikan fitur tersebut dengan pola khas karies dan karang gigi. Kurva pelatihan yang stabil dan selisih kecil antara performa data latih dan validasi mengindikasikan bahwa risiko *overfitting* dapat dikendalikan melalui augmentasi, Dropout, dan penggunaan *callback*.

Nilai tersebut berada pada rentang yang sama dengan laporan Wahyuningsih dkk. (2024), yang memperoleh akurasi 0,98 untuk MobileNetV2 pada klasifikasi karies gigi, dan lebih tinggi daripada akurasi MobileNetV2 sebesar 80% pada klasifikasi kondisi gigi dari citra intraoral (Makarim dkk., 2024) maupun akurasi 83,4–87,4% pada deteksi karies dari foto *smartphone* (Thanh dkk., 2022). Perbandingan ini perlu ditafsirkan secara hati-hati karena *dataset*, jenis citra, jumlah kelas, dan tugas klasifikasinya berbeda. Selain itu, performa tinggi pada penelitian ini diperoleh pada data uji yang berasal dari sumber *dataset* yang sama dengan data latih. Hasil pengujian pada citra di luar *dataset* memperkuat kehati-hatian tersebut, karena *accuracy* yang tinggi pada data uji belum tentu berlaku pada citra yang diambil pengguna dengan kondisi berbeda (*domain shift*).

Kesalahan terbanyak terjadi antara kelas karies dan karang gigi (4 dari 6 kesalahan). Hal ini dapat dijelaskan oleh karakteristik visual keduanya yang relatif serupa pada permukaan gigi, misalnya perubahan warna kecokelatan atau gelap, sehingga pola yang dipelajari CNN dapat tumpang tindih. *Recall* karang gigi yang paling rendah (0,97) diduga berkaitan dengan jumlah data latih kelas ini yang lebih sedikit setelah deduplikasi (580 citra dibandingkan 700 citra pada dua kelas lainnya). Dugaan ini belum diuji secara eksperimental sehingga perlu dibuktikan pada penelitian lanjutan.

Temuan penting lainnya adalah bahwa 467 dari 1.296 citra karang gigi (sekitar 36%) merupakan duplikat. Apabila tidak dihilangkan sebelum pembagian *dataset*, citra identik dapat muncul pada data latih dan data uji sehingga evaluasi tidak lagi mengukur kemampuan generalisasi model. Dengan deduplikasi sebelum pembagian *dataset* dan verifikasi *overlap* = 0, nilai evaluasi yang dilaporkan lebih dapat dipercaya sebagai gambaran kinerja model pada data baru dari sumber yang sama.

Kelebihan sistem yang dikembangkan adalah, pertama, MobileNetV2 bersifat ringan dan efisien sehingga sesuai untuk layanan prediksi yang dihubungkan dengan perangkat seluler (Sandler dkk., 2018); kedua, alur kerja terintegrasi dari model hingga aplikasi sehingga pengguna awam cukup mengunggah foto dari galeri; dan ketiga, pemisahan data validasi dan data uji serta verifikasi bebas *overlap* menjaga objektivitas evaluasi. Di sisi lain, penelitian ini memiliki beberapa keterbatasan. *Dataset* berasal dari satu sumber publik dan labelnya belum divalidasi oleh dokter gigi, sehingga keandalan *ground truth* belum

dapat dipastikan, dan jumlah citra karang gigi lebih sedikit daripada kelas lainnya. Evaluasi menggunakan satu kali pembagian data tanpa *k-fold cross-validation* dan belum membandingkan MobileNetV2 dengan arsitektur CNN lain. Pengujian pada citra di luar *dataset* dilaporkan secara kualitatif dan menunjukkan *domain shift*. Model bersifat *closed-set* sehingga tidak dapat menolak citra non-gigi, aplikasi hanya menerima citra dari galeri dan memerlukan koneksi internet karena prediksi dilakukan di server, serta cakupan penelitian terbatas pada citra 2D (bukan radiografi). Waktu respons sistem juga belum dilaporkan secara kuantitatif.

Secara terapan, sistem ini dapat berperan sebagai alat bantu klasifikasi awal yang praktis dan sarana untuk meningkatkan kesadaran masyarakat terhadap kesehatan gigi dan mulut, tetapi hasilnya tidak boleh dipandang sebagai diagnosis medis dan tetap memerlukan pemeriksaan oleh tenaga medis. Secara teoretis, penelitian ini menunjukkan bahwa MobileNetV2 dengan *transfer learning* dua tahap layak digunakan untuk klasifikasi citra gigi tiga kelas dan bahwa deduplikasi merupakan tahap penting dalam menyusun *dataset* citra agar evaluasi model valid.

KESIMPULAN

Penelitian ini berhasil merancang dan membangun model klasifikasi citra gigi berbasis CNN MobileNetV2 dengan pendekatan *transfer learning* dan *fine-tuning* yang mampu mengklasifikasikan citra ke dalam tiga kelas, yaitu karies, karang gigi, dan normal. Model mencapai *accuracy* 98,59% pada 425 citra uji dengan *precision*, *recall*, dan *F1-score* yang seimbang pada ketiga kelas. Kualitas *dataset*, khususnya keberadaan citra duplikat, terbukti memengaruhi validitas evaluasi sehingga deduplikasi perlu dilakukan sebelum pembagian *dataset*. Model juga berhasil diimplementasikan dalam sistem terpadu berupa *backend* FastAPI yang *di-deploy* pada Railway dan aplikasi Android berbasis Flutter yang menampilkan hasil klasifikasi beserta nilai *confidence*. Sistem ini diposisikan sebagai alat bantu klasifikasi awal, bukan alat diagnosis medis.

Penelitian selanjutnya disarankan untuk menambahkan variasi citra yang menyerupai pengambilan gambar oleh pengguna secara nyata, seperti foto ponsel dengan berbagai sudut, jarak, dan pencahayaan, guna mengurangi *domain shift*. Pengumpulan *dataset* sebaiknya disertai validasi yang lebih ketat, termasuk pemeriksaan duplikasi sejak tahap awal, dan label divalidasi oleh dokter gigi profesional untuk meningkatkan keandalan *ground truth*. Perbandingan dengan arsitektur CNN lain serta penerapan *k-fold cross-validation* dapat memperkuat validitas evaluasi, sedangkan pengembangan aplikasi dapat menambahkan fitur pengambilan gambar langsung melalui kamera.

UCAPAN TERIMA KASIH

Artikel ini disusun berdasarkan skripsi penulis pada Program Studi Sistem Informasi, Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara. Penulis mengucapkan terima kasih kepada dosen pembimbing dan pihak fakultas yang telah memberikan bimbingan dan dukungan selama penelitian.

DAFTAR PUSTAKA

- Carrington, A. M., Manuel, D. G., Fieguth, P. W., Ramsay, T., Osmani, V., Wernly, B., ... Holzinger, A. (2023). Deep ROC analysis and AUC as balanced average accuracy, for improved classifier selection, audit and explanation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1), 329–341. <https://doi.org/10.1109/TPAMI.2022.3145392>
- Daniel Erku, R. K. A. E. W. F. N. A. Z. dan Y. A. (2023). Digital health interventions to improve access to and quality of primary health care services: A scoping review. *International Journal of Environmental Research and Public Health*, 20(19), 6854. <https://doi.org/10.3390/ijerph20196854>
- GBD 2021 Oral Disorders Collaborators. (2025). Trends in the global, regional, and national burden of oral conditions from 1990 to 2021: A systematic analysis for the Global Burden of Disease Study 2021. *The Lancet*, 405(10482), 897–910. [https://doi.org/10.1016/S0140-6736\(24\)02811-3](https://doi.org/10.1016/S0140-6736(24)02811-3)
- Gonzalez, R. C., & Woods, R. E. (2021). *Digital image processing* (4th ed.). Pearson.
- Hasan, F., Yuliana, L. T., Budi, H. S., Ramasamy, R., Ambiya, Z. I., & Ghaisani, A. M. (2024). Prevalence of dental caries among children in Indonesia: A systematic review and meta-analysis of observational studies. *Heliyon*, 10(11), e32102. <https://doi.org/10.1016/j.heliyon.2024.e32102>
- Iman, M., Arabnia, H. R., & Rasheed, K. (2023). A review of deep transfer learning and recent advancements. *Technologies*, 11(2), 40. <https://doi.org/10.3390/technologies11020040>
- Iqbal, F. N. (2023). A brief introduction to application programming interface (API). <https://doi.org/10.5281/zenodo.10198423>

- Kementerian Kesehatan Republik Indonesia. (2022). Profil kesehatan Indonesia 2021. https://kemkes.go.id/app_asset/file_content_download/Profil-KesehatanIndonesia-2021.pdf
- Lee, J. H., Kim, D. H., Jeong, S. N., & Choi, S. H. (2018). Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm. *Journal of Dentistry*, 77, 106–111. <https://doi.org/10.1016/j.jdent.2018.07.015>
- Luque, A., Carrasco, A., Martín, A., & de las Heras, A. (2019). The impact of class imbalance in classification performance metrics based on the binary confusion matrix. *Pattern Recognition*, 91, 216–231. <https://doi.org/10.1016/j.patcog.2019.02.023>
- Makarim, A., Karlita, T., Sigit, R., Bayu Dewantara, B. S., & Brahmana, A. (2024). Deep learning models for dental conditions classification using intraoral images. *JOIV: International Journal on Informatics Visualization*, 8(3), 1099–1108. <https://doi.org/10.62527/joiv.8.3.1914>
- Medavarapu, S. S. (2024). Advancements in deep learning: A review of Keras and TensorFlow frameworks. *Journal of Scientific and Engineering Research*, 11(5), 282–286. <https://jsaer.com/download/vol-11-iss-5-2024/JSAER2024-11-5-282-286.pdf>
- Nguyen, T. T. T., Tran, V. T., Nguyen, T. T., Le, V. N., & Nguyen, T. T. H. (2022). Deep learning application in dental caries detection using intraoral photos taken by smartphones. *Scientific Reports*, 12, 19114. <https://doi.org/10.1038/s41598-022-23329-7>
- Park, S., Erkinov, H., Hasan, M. A. M., Nam, S.-H., Kim, Y.-R., Shin, J., & Chang, W.-D. (2023). Periodontal disease classification with color teeth images using convolutional neural networks. *Electronics*, 12(7), 1518. <https://doi.org/10.3390/electronics12071518>
- Perez-Meana, H., & Nakano-Miyatake, M. (2025). Digital image processing: Technologies and applications. *Applied Sciences*, 15(23), 12709. <https://doi.org/10.3390/app152312709>
- Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C. (2018). MobileNetV2: Inverted residuals and linear bottlenecks. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 4510–4520. <https://doi.org/10.1109/CVPR.2018.00474>
- The Dental Association of Thailand. (2023). Guideline on caries risk assessment and management. *Journal of the Dental Association of Thailand*. <https://doi.org/10.14456/jdat.2023.24>
- Wahyuningsih, W., Nugraha, G. S., & Dwiyanaputra, R. (2024). Classification of dental caries disease in tooth images using a comparison of EfficientNet-B0, MobileNetV2, ResNet-50, InceptionV3 architectures. *Jurnal Teknik Informatika (JUTIF)*, 5(4), 177–185. <https://doi.org/10.52436/1.jutif.2024.5.4.2187>
- Warreth, A. (2023). Dental caries and its management. *International Journal of Dentistry*, 2023, 1–15. <https://doi.org/10.1155/2023/9365845>
- World Health Organization. (2025). Oral health. Retrieved from <https://www.who.int/news-room/fact-sheets/detail/oral-health>
- Xu, M., Yoon, S., Fuentes, A., & Park, D. S. (2023). A comprehensive survey of image augmentation techniques for deep learning. *Pattern Recognition*, 137, 109347. <https://doi.org/10.1016/j.patcog.2023.109347>
- Zhao, X., Wang, L., Zhang, Y., Han, X., Deveci, M., & Parmar, M. (2024). A review of convolutional neural networks in computer vision. *Artificial Intelligence Review*, 57, 99. <https://doi.org/10.1007/s10462-024-10721-6>