



Analisis Performa dan Efisiensi VGG16, ResNet50, dan MobileNetV2 pada Klasifikasi Citra Jajanan Tradisional

Ni Luh Widi Rahayu^{1*}, Ni Kadek Bumi Krismentari², I Kayan Herdiana²

¹ Program Studi Sistem Informasi, Universitas Bali Internasional Muhammadiyah, Denpasar, Indonesia

² Program Studi Bisnis Digital, Institut Bisnis dan Teknologi Indonesia, Denpasar, Indonesia

Email: ^{1*}luhwidirahayu@gmail.com, ²kadekbumi@instiki.ac.id, ³herdikayan@instiki.ac.id

Email Penulis Korespondensi: luhwidirahayu@gmail.com

Abstrak—Penelitian sebelumnya mengenai klasifikasi citra jajanan tradisional telah menunjukkan bahwa arsitektur VGG16 mampu menghasilkan performa yang baik. Namun, penelitian tersebut masih berfokus pada satu arsitektur dan belum memberikan perbandingan yang memadai antara performa klasifikasi dan efisiensi komputasi beberapa model. Kondisi tersebut menimbulkan kebutuhan untuk mengevaluasi arsitektur lain agar dapat diperoleh model yang tidak hanya akurat, tetapi juga sesuai dengan kebutuhan implementasi pada platform yang berbeda. Penelitian ini bertujuan membandingkan performa dan efisiensi tiga arsitektur transfer learning, yaitu VGG16, ResNet50, dan MobileNetV2, dalam mengklasifikasikan citra makanan tradisional. Dataset terdiri atas 2.445 citra yang terbagi ke dalam sepuluh kelas, yaitu batun bedil, bubur injin, jaje lukis, jaje piling, jaje wajik, kaliadrem, klepon, laklak, ongol-ongol, dan pisang rai. Dataset dibagi dengan rasio 70:15:15 menjadi 1.706 citra latih, 363 citra validasi, dan 376 citra uji. Seluruh model dilatih menggunakan konfigurasi yang sama, yaitu ukuran input 224×224 piksel, batch size 64, optimizer Adam, learning rate 0,0001, categorical crossentropy, augmentasi data, dan 10 epoch. Evaluasi dilakukan menggunakan accuracy, precision, recall, F1-score, training time, model size, dan inference time per image. Hasil pengujian menunjukkan bahwa ResNet50 memperoleh performa klasifikasi terbaik dengan accuracy 84,04%, precision 85,15%, recall 85,16%, dan F1-score 84,89%. MobileNetV2 menghasilkan accuracy 82,45% dengan ukuran model paling kecil, yaitu 12,98 MB, sedangkan VGG16 memperoleh accuracy 73,67%. Temuan ini menunjukkan bahwa ResNet50 lebih sesuai untuk sistem yang mengutamakan performa klasifikasi, sedangkan MobileNetV2 lebih tepat untuk aplikasi yang membutuhkan model ringan. Kontribusi penelitian ini terletak pada evaluasi komparatif performa dan efisiensi tiga arsitektur dalam kondisi eksperimen yang sama sehingga menghasilkan rekomendasi model berdasarkan kebutuhan implementasi.

Kata Kunci: Klasifikasi Citra; Transfer Learning; VGG16; ResNet50; MobileNetV2

Abstract—Previous research on traditional food image classification demonstrated that the VGG16 architecture could achieve good classification performance. However, the study focused on a single architecture and did not provide a comprehensive comparison between predictive performance and computational efficiency across multiple models. This limitation highlights the need to evaluate other architectures in order to identify a model that is not only accurate but also suitable for different deployment environments. This study aims to compare the performance and efficiency of three transfer learning architectures, namely VGG16, ResNet50, and MobileNetV2, for traditional food image classification. The dataset consisted of 2,445 images grouped into ten classes: batun bedil, bubur injin, jaje lukis, jaje piling, jaje wajik, kaliadrem, klepon, laklak, ongol-ongol, and pisang rai. The dataset was divided using a 70:15:15 ratio into 1,706 training images, 363 validation images, and 376 testing images. All models were trained under the same configuration, using an input size of 224×224 pixels, batch size of 64, Adam optimizer, learning rate of 0.0001, categorical crossentropy loss, data augmentation, and 10 epochs. Model evaluation was conducted using accuracy, precision, recall, F1-score, training time, model size, and inference time per image. The results showed that ResNet50 achieved the best classification performance, with an accuracy of 84.04%, precision of 85.15%, recall of 85.16%, and F1-score of 84.89%. MobileNetV2 achieved an accuracy of 82.45% with the smallest model size of 12.98 MB, while VGG16 obtained an accuracy of 73.67%. These findings indicate that ResNet50 is more suitable for systems that prioritize classification performance, whereas MobileNetV2 is more appropriate for applications requiring a lightweight model. The contribution of this study lies in the comparative evaluation of performance and efficiency across three architectures under the same experimental setting, resulting in practical model recommendations based on deployment requirements.

Keywords: Image Classification; Transfer Learning; VGG16; ResNet50; MobileNetV2

1. PENDAHULUAN

Makanan tradisional tidak hanya berfungsi sebagai produk konsumsi, tetapi juga menjadi bagian dari pengetahuan lokal, praktik sosial, dan identitas budaya suatu masyarakat (Najla et al., 2025), (Felicia Lie et al., 2026). Di dalamnya terdapat keterkaitan antara bahan pangan, teknik pengolahan, kebiasaan masyarakat, serta nilai-nilai yang diwariskan dari satu generasi ke generasi berikutnya (Fransiska et al., 2026). Dalam konteks Bali, jajanan tradisional seperti batun bedil, bubur injin, jaje lukis, jaje piling, jaje wajik, kaliadrem, klepon, laklak, ongol-ongol, dan pisang rai masih dijumpai dalam kehidupan masyarakat. Jajanan tersebut dikonsumsi dalam kehidupan sehari-hari, diperdagangkan di pasar tradisional, serta digunakan dalam berbagai kegiatan adat dan keagamaan. Dengan demikian, keberadaan jajanan tradisional Bali tidak hanya berkaitan dengan aspek kuliner, tetapi juga dengan keberlanjutan pengetahuan budaya dan identitas lokal masyarakat Bali (Rahayu et al., 2024). Perubahan pola konsumsi, perkembangan industri makanan modern, serta meningkatnya pengaruh budaya kuliner global dapat memengaruhi cara masyarakat, khususnya generasi muda, mengenal makanan tradisional (Satato et al., 2025). Generasi muda saat ini lebih banyak memperoleh informasi melalui media digital, baik melalui mesin pencari, media sosial, aplikasi, maupun berbagai platform berbasis visual. Kondisi tersebut memberikan peluang sekaligus tantangan bagi keberadaan kuliner lokal (Erlanda & Krisnadi, 2025). Apabila pengetahuan mengenai jajanan tradisional tidak ikut hadir dalam ruang digital (Widjanarko et al., 2023), maka



akses generasi muda terhadap informasi mengenai jenis, bentuk, nama, serta karakteristik jajanan tradisional dapat semakin terbatas. Di sisi lain, dokumentasi digital mengenai jajanan tradisional Bali masih belum banyak tersedia secara terstruktur dan mudah diakses (Bagaskara et al., 2025). Sebagian besar pengetahuan mengenai jajanan tradisional masih diperoleh melalui pengalaman langsung, kebiasaan keluarga, pengetahuan lisan, pasar tradisional, atau interaksi dengan masyarakat local (Harsana et al., 2023), (Kertopati et al., 2025). Bentuk pewarisan seperti ini memiliki nilai sosial yang penting, tetapi juga menghadapi keterbatasan dalam hal jangkauan, konsistensi, dan keberlanjutan, (Romdona et al., 2026). Pengetahuan yang hanya bergantung pada transmisi lisan dapat terputus apabila tidak didukung oleh dokumentasi yang memadai (Amini et al., 2025). Oleh karena itu, dibutuhkan pendekatan yang mampu mendukung proses pengenalan, dokumentasi, dan preservasi pengetahuan kuliner lokal dalam bentuk digital.

Perkembangan teknologi informasi memberikan peluang untuk menjawab kebutuhan tersebut. Salah satu pendekatan yang dapat digunakan adalah klasifikasi citra berbasis deep learning. Melalui pendekatan computer vision, citra makanan dapat dianalisis berdasarkan karakteristik visual seperti bentuk, warna, tekstur, pola permukaan, dan cara penyajian (D. Liu et al., 2025). Sistem klasifikasi citra dapat membantu mengenali jenis makanan secara otomatis berdasarkan gambar yang diberikan (Thiodorus et al., 2021). Dalam konteks pelestarian kuliner, teknologi ini dapat dikembangkan lebih lanjut menjadi katalog digital, media edukasi, sistem informasi kuliner, atau aplikasi pengenalan makanan tradisional. Namun, klasifikasi citra makanan merupakan tugas yang tidak sederhana (Tahir & Loo, 2021), makanan tradisional sering memiliki bentuk yang tidak seragam karena proses pembuatannya masih banyak dilakukan secara manual (Kolanowski et al., 2025). Selain itu, beberapa jenis jajanan menggunakan bahan yang serupa, seperti tepung beras, ketan, kelapa, gula merah, atau daun pisang. Kesamaan bahan tersebut dapat menghasilkan warna, tekstur, dan tampilan permukaan yang hampir sama. Variasi sudut pengambilan gambar, pencahayaan, latar belakang, ukuran objek, dan penyajian juga dapat memengaruhi tampilan citra. Kondisi tersebut menyebabkan model klasifikasi perlu mampu menangkap perbedaan visual yang cukup halus agar setiap kelas dapat dibedakan dengan baik (L. Liu & Xiao, 2025).

Salah satu metode deep learning yang banyak digunakan dalam klasifikasi citra adalah Convolutional Neural Network (CNN) (Chen et al., 2021). CNN mampu mempelajari fitur visual secara bertahap melalui lapisan konvolusi. Lapisan awal biasanya menangkap fitur sederhana, seperti garis, tepi, warna, dan pola dasar. Sementara itu, lapisan yang lebih dalam membentuk representasi yang lebih kompleks, seperti tekstur, bentuk, struktur, dan pola objek (Reza et al., 2026). Kemampuan tersebut membuat CNN banyak diterapkan pada berbagai bidang pengenalan citra, termasuk klasifikasi makanan (Chen et al., 2021). Dalam perkembangannya, pendekatan transfer learning banyak digunakan untuk mengatasi keterbatasan jumlah data. Transfer learning memungkinkan model yang telah dilatih pada dataset berskala besar, seperti ImageNet, untuk digunakan kembali pada tugas klasifikasi yang lebih spesifik (Plested et al., 2025). Melalui pendekatan ini, model tidak perlu mempelajari seluruh fitur visual dari awal, tetapi dapat memanfaatkan bobot dan representasi fitur yang telah diperoleh sebelumnya (Kumar, 2023). Strategi ini sangat berguna ketika dataset penelitian relatif terbatas, seperti pada dataset budaya lokal atau citra makanan tradisional. Selain mengurangi kebutuhan data dalam jumlah besar, transfer learning juga dapat mempercepat proses pelatihan dan mengurangi kebutuhan komputasi (Mufidatuzzainiya & Faisal, 2025).

Penelitian Rahayu et al., (2024) telah menunjukkan bahwa arsitektur VGG16 dapat digunakan untuk mengklasifikasikan citra jajanan tradisional Bali dengan performa yang baik melalui beberapa skenario pengaturan parameter, seperti dropout, batch size, jumlah epoch, dan learning rate. Temuan tersebut menjadi dasar penting bagi pengembangan sistem klasifikasi citra jajanan tradisional Bali. Namun, penelitian tersebut masih berfokus pada optimasi satu arsitektur, sehingga belum dapat menjelaskan apakah performa yang diperoleh terutama berasal dari karakteristik struktural VGG16 atau dari konfigurasi pelatihan yang digunakan. Evaluasinya juga lebih menitikberatkan pada performa klasifikasi dan belum membahas efisiensi komputasi, seperti ukuran model, waktu pelatihan, dan waktu inferensi, serta belum menghasilkan rekomendasi model berdasarkan target implementasi, misalnya untuk sistem berbasis server atau aplikasi mobile. Padahal, dalam implementasi nyata, pemilihan model tidak hanya ditentukan oleh tingkat akurasi, tetapi juga oleh kebutuhan penyimpanan, kecepatan prediksi, dan sumber daya komputasi. Oleh karena itu, penelitian ini membandingkan tiga arsitektur dengan karakter struktural yang berbeda, yaitu VGG16 sebagai baseline, ResNet50 sebagai jaringan residual yang lebih dalam (Puspitasari et al., 2025), dan MobileNetV2 sebagai model ringan untuk lingkungan dengan sumber daya terbatas (Wijaya & Christioko, 2025). VGG16 menggunakan susunan lapisan konvolusi yang berurutan (Linkon et al., 2021), ResNet50 memanfaatkan residual connection untuk mempertahankan informasi dari lapisan sebelumnya (Muriagista & Kurniadi, 2025), sedangkan MobileNetV2 menggunakan inverted residual, linear bottleneck, dan depthwise separable convolution untuk mengurangi kompleksitas model (Alfan & Kumalasari, 2026). Arsitektur lain seperti EfficientNet tidak dimasukkan karena penelitian ini secara khusus difokuskan pada perbandingan tiga representasi arsitektur tersebut. EfficientNet menggunakan pendekatan compound scaling yang mengatur kedalaman, lebar, dan resolusi jaringan secara bersamaan (Witjaksono et al., 2025), sehingga akan memperluas ruang eksperimen di luar fokus utama penelitian dan lebih tepat dipertimbangkan pada penelitian lanjutan.

Berdasarkan kondisi tersebut, kesenjangan penelitian tidak hanya terletak pada belum adanya perbandingan langsung antara VGG16, ResNet50, dan MobileNetV2 pada dataset jajanan tradisional Bali dengan konfigurasi pelatihan yang sama, tetapi juga pada belum tersedianya evaluasi terpadu antara performa prediksi, efisiensi komputasi, dan kebutuhan implementasi. Penelitian terdahulu belum dapat menunjukkan model yang paling sesuai ketika akurasi, ukuran model, waktu pelatihan, dan waktu inferensi dipertimbangkan secara bersamaan. Kebaruan penelitian ini

terletak pada evaluasi komparatif tiga arsitektur transfer learning yang mewakili karakter struktural berbeda, yaitu baseline, residual network, dan lightweight network, menggunakan dataset serta skenario eksperimen yang sama. Evaluasi dilakukan melalui accuracy, precision, recall, F1-score, training time, model size, dan inference time per image sehingga rekomendasi model tidak hanya didasarkan pada tingkat akurasi, tetapi juga pada kesesuaiannya terhadap kebutuhan implementasi berbasis server maupun mobile.

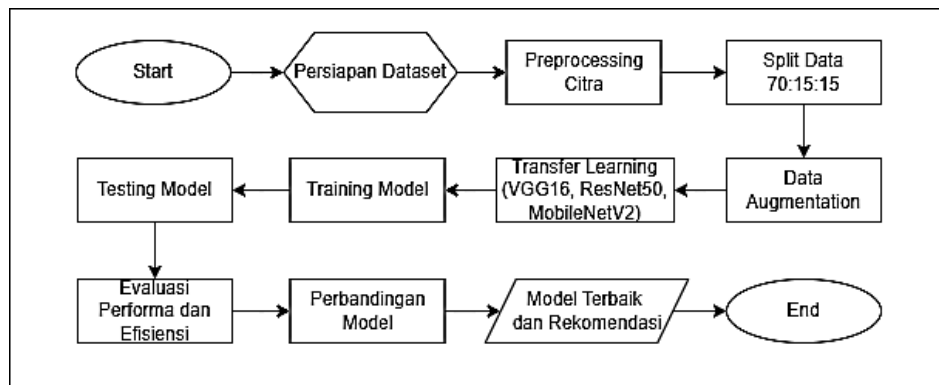
Tujuan penelitian ini adalah menganalisis dan membandingkan performa serta efisiensi arsitektur VGG16, ResNet50, dan MobileNetV2 dalam mengklasifikasikan citra jajanan tradisional Bali. Penelitian ini diharapkan dapat menghasilkan rekomendasi model yang lebih tepat berdasarkan dua kebutuhan utama, yaitu performa klasifikasi dan efisiensi implementasi. Hasil penelitian diharapkan dapat menjadi dasar bagi pengembangan sistem preservasi kuliner digital, seperti katalog jajanan tradisional, media edukasi berbasis citra, aplikasi pengenalan makanan, maupun sistem informasi kuliner lokal. Dengan demikian, penelitian ini tidak hanya memberikan kontribusi pada bidang klasifikasi citra, tetapi juga mendukung upaya dokumentasi dan pelestarian pengetahuan kuliner tradisional Bali melalui pemanfaatan teknologi digital.

2. METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan eksperimen komputasional dengan metode kuantitatif. Eksperimen dilakukan untuk membandingkan performa dan efisiensi tiga arsitektur *Convolutional Neural Network* berbasis *transfer learning*, yaitu VGG16, ResNet50, dan MobileNetV2. Ketiga model dilatih dan diuji menggunakan dataset, pembagian data, ukuran input, optimizer, *learning rate*, *batch size*, jumlah *epoch*, dan metrik evaluasi yang sama. Penggunaan konfigurasi yang seragam bertujuan agar perbedaan hasil lebih mencerminkan karakteristik masing-masing arsitektur, bukan perbedaan skenario pelatihan. Objek penelitian berupa dataset citra jajanan tradisional Bali yang terdiri atas sepuluh kelas, yaitu batun bedil, bubur injin, jaje lukis, jaje piling, jaje wajik, kaliadrem, klepon, laklak, ongol-ongol, dan pisang rai. Penelitian tidak hanya menilai kemampuan klasifikasi, tetapi juga mempertimbangkan efisiensi model. Performa klasifikasi diukur menggunakan *accuracy*, *precision*, *recall*, dan *F1-score*, sedangkan efisiensi model diukur berdasarkan *training time*, *model size*, dan *inference time per image*.

2.1 Tahapan Penelitian

Tahapan penelitian meliputi persiapan dataset, preprocessing, pembagian dataset, augmentasi data, penerapan tiga arsitektur transfer learning, pelatihan model, pengujian model, evaluasi performa dan efisiensi, perbandingan hasil, serta penentuan rekomendasi model. Alur tersebut ditunjukkan pada Gambar 1.



Gambar 1. Alur Tahapan Penelitian

Pada Gambar 1 memperlihatkan bahwa proses penelitian dimulai dengan menyiapkan dataset citra jajanan tradisional dan mengelompokkannya berdasarkan label kelas. Selanjutnya, citra melalui tahap preprocessing citra menjadi 224 x 224 piksel dan dibagi menjadi data latih, data validasi, serta data uji dengan rasio 70:15:15. Augmentasi hanya diterapkan pada data latih yang bertujuan untuk menambah variasi citra dan mengurangi risiko overfitting. Dataset kemudian digunakan secara terpisah untuk melatih VGG16, ResNet50, dan MobileNetV2 dengan konfigurasi yang sama. Hasil pengujian ketiga model selanjutnya dibandingkan berdasarkan performa klasifikasi dan efisiensi komputasi untuk menghasilkan rekomendasi model.

2.2 Dataset Citra Jajanan Bali

Dataset dalam penelitian ini terdiri atas citra jajanan tradisional Bali yang dikelompokkan ke dalam sepuluh kelas, yaitu batun_bedil, bubur_injin, jaje_lukis, jaje_piling, jaje_wajik, kaliadrem, klepon, laklak, ongol_ongol, dan pisang rai. Citra disusun ke dalam folder berdasarkan label kelas agar proses pembacaan data dan label dapat dilakukan secara otomatis pada saat proses pelatihan model. Adapun dataset citra jajanan khas Bali yang digunakan dapat dilihat pada Gambar 2.

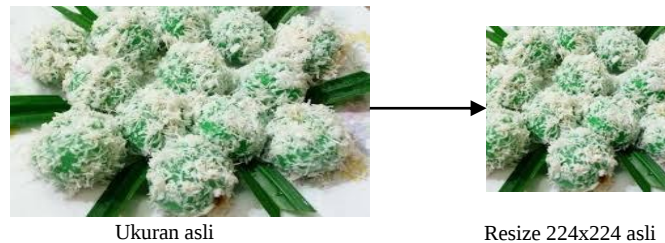
Berdasarkan Gambar 2, dataset citra jajanan khas Bali yang digunakan dalam penelitian ini sebanyak 2.445 citra yang terdiri dari batun bedil 189 citra, bubur injin 194 citra, jaje Lukis 347, jaje piling 184, jaje wajik 240, kaliadrem 168, klepon 358, laklak 206, ongol-ongol 307 dan pisang rai 252.



Gambar 2. Dataset Citra Jajanan khas Bali

2.3 Preprocessing Data

Pada tahap ini, preprocessing dilakukan untuk memastikan seluruh data memiliki format yang seragam sebelum masuk ke tahap pelatihan agar sesuai dengan standar input dari model VGG16, ResNet50 dan MobileNetV2 (Akmal et al., 2026). Adapun contoh resize dapat dilihat pada Gambar 3 berikut.



Gambar 3. Contoh resize image

Gambar 3 adalah salah satu contoh preprocessing yang dilakukan, yaitu citra klepon diubah ukurannya menjadi 224×224 piksel dan proses ini diterapkan pada seluruh dataset yang digunakan. Selain itu, label data dikonversi ke dalam format kategorikal karena penelitian ini menggunakan klasifikasi multi-kelas.

2.4 Split Dataset dan Augmentasi Data

Dataset dibagi menjadi data latih, validasi, dan uji dengan rasio 70:15:15. Hasil pembagian menghasilkan 1.706 citra untuk pelatihan, 363 citra untuk validasi, dan 376 citra untuk pengujian. Setiap kelas tetap terwakili pada seluruh subset sehingga proses pelatihan dan evaluasi dapat dilakukan secara proporsional. Distribusi dataset setelah proses pembagian ditunjukkan pada Tabel 1.

Tabel 1. Distribusi Dataset

Class	Training	Validation	Testing
batun_bedil	132	28	29
bubur_injin	135	29	30
jaje_lukis	242	52	53
jaje_piling	128	27	29
jaje_wajik	168	36	36
kaliadrem	117	25	26
klepon	250	53	55
laklak	144	30	32
ongol_ongol	214	46	47
pisang_rai	176	37	39
Total	1,706	363	376

Berdasarkan Tabel 1 jumlah citra pada setiap kelas tidak sepenuhnya sama. Kondisi ini mencerminkan ketersediaan data citra yang berbeda pada setiap jenis jajanan. Meskipun terdapat sedikit ketidakseimbangan jumlah data antar kelas, setiap kelas masih memiliki jumlah sampel yang cukup untuk proses pelatihan dan pengujian. Setelah dilakukan proses split data kemudian dilanjutkan dengan tahap augmentasi, dimana proses ini diterapkan hanya pada data latih untuk menambah variasi visual dan mengurangi risiko *overfitting*. Teknik yang digunakan meliputi *random horizontal flip*, *random rotation* sebesar 0,10, *random zoom* sebesar 0,10, *random translation* sebesar 0,10 pada arah horizontal dan vertikal, serta *random contrast* sebesar 0,10. Data validasi dan data uji tidak diberi augmentasi agar evaluasi dilakukan menggunakan citra yang tidak dimodifikasi.



2.5 Penerapan Arsitektur Transfer Learning

Transfer learning memungkinkan pengetahuan visual yang telah dipelajari model dari dataset besar digunakan kembali pada dataset baru dengan karakteristik yang lebih spesifik (Plested et al., 2025). Dalam penelitian ini, VGG16, ResNet50, dan MobileNetV2 menggunakan bobot awal ImageNet dengan parameter *include_top=False*. Lapisan klasifikasi bawaan dari masing-masing model dihapus dan diganti dengan *classification head* yang disesuaikan dengan sepuluh kelas keluaran. VGG16 merupakan arsitektur CNN yang menggunakan susunan lapisan konvolusi berukuran 3×3 dan dikembangkan untuk mengevaluasi pengaruh kedalaman jaringan terhadap performa pengenalan citra (Linkon et al., 2021). Arsitektur ini memiliki struktur yang seragam dan menggunakan 16 lapisan berbobot, sehingga sering digunakan sebagai model dasar atau baseline dalam penelitian klasifikasi citra. Dalam penelitian ini, VGG16 digunakan sebagai baseline karena telah diterapkan pada penelitian sebelumnya mengenai klasifikasi jajanan tradisional.

ResNet50 merupakan jaringan residual yang dirancang untuk mempermudah pelatihan jaringan yang lebih dalam. Arsitektur ini menggunakan residual connection atau skip connection untuk meneruskan informasi dari lapisan sebelumnya ke lapisan berikutnya (Puspitasari et al., 2025). Mekanisme residual membantu proses optimasi jaringan yang dalam dan mengurangi masalah degradasi performa ketika jumlah lapisan bertambah. ResNet50 digunakan dalam penelitian ini untuk menguji kemampuan jaringan yang lebih dalam dalam menangkap perbedaan visual antarjenis makanan tradisional. Sedangkan, MobileNetV2 merupakan arsitektur ringan yang dirancang untuk aplikasi mobile dan lingkungan dengan sumber daya komputasi terbatas (Wijaya & Christioko, 2025). Arsitektur ini menggunakan inverted residual, linear bottleneck, dan depthwise convolution untuk mengurangi jumlah operasi serta parameter tanpa menghilangkan kemampuan representasi fitur secara berlebihan. MobileNetV2 dipilih untuk mengevaluasi apakah model berukuran lebih kecil tetap mampu memberikan performa klasifikasi yang kompetitif.

Pada ketiga arsitektur, *classification head* baru terdiri atas *Global Average Pooling*, *Dense layer* sebanyak 256 neuron dengan aktivasi *ReLU*, *Dropout* sebesar 0,2, serta lapisan keluaran sebanyak sepuluh neuron dengan aktivasi *softmax*. Seluruh lapisan pada *convolutional base* dibekukan sehingga pembaruan bobot selama pelatihan hanya dilakukan pada *classification head*. Penelitian ini tidak melakukan tahap *fine-tuning* terhadap lapisan konvolusional.

2.6 Training Model

Tahap training model dilakukan dengan melatih tiga arsitektur transfer learning, yaitu VGG16, ResNet50, dan MobileNetV2, menggunakan data latih yang telah melalui proses preprocessing dan augmentasi. Proses ini bertujuan agar model mampu mempelajari karakteristik visual jajanan tradisional Bali, seperti bentuk, warna, tekstur, dan pola penyajian dari setiap kelas (R. A. Pratama et al., 2025). Adapun skenario pelatihan model dapat dilihat pada tabel 2 berikut:

Tabel 2. Skenario Pelatihan

Komponen	Skenario
Arsitektur	VGG16, ResNet50, MobileNetV2
Pendekatan	Transfer learning
Bobot awal	ImageNet
Ukuran masukan	224×224 piksel
Pembagian dataset	70% latih, 15% validasi, 15% uji
Random seed	42
Batch size	64
Epoch maksimum	10
Optimizer	Adam
Learning rate	0,0001
Loss function	Categorical crossentropy
Classification head	Global Average Pooling, Dense 256, Dropout 0,2, Softmax
Strategi pelatihan	Convolutional base dibekukan
Callback	Early stopping dan model checkpoint

Berdasarkan Tabel 2 ketiga model dilatih menggunakan konfigurasi yang sama, yaitu ukuran input 224×224 piksel, *batch size* 64, *optimizer* Adam, *learning rate* 0,0001, *loss function* *categorical crossentropy*, dan 10 *epoch*. Pada tahap awal pelatihan, *convolutional base* dari model *pretrained* dibekukan sehingga proses pembelajaran difokuskan pada *classification head* baru yang terdiri atas *global average pooling layer*, *dense layer*, *dropout layer*, dan *softmax output layer*. Penggunaan konfigurasi yang sama bertujuan agar hasil perbandingan antar model dapat dianalisis secara adil. Data validasi digunakan selama proses *training* untuk memantau kemampuan generalisasi model pada setiap *epoch*. Selain itu, *early stopping* dan model *checkpoint* diterapkan untuk menjaga proses pelatihan tetap optimal dan menyimpan model dengan performa validasi terbaik (Arumsari et al., 2024). Setelah tahap *training* selesai, model yang telah dilatih kemudian diuji menggunakan data testing untuk memperoleh *nilai accuracy*, *precision*, *recall*, *F1-score*, serta indikator efisiensi seperti *training time*, model size, dan *inference time per image*.



2.7 Testing dan Evaluasi Model

Pengujian dilakukan menggunakan 376 citra uji yang tidak digunakan selama proses pelatihan maupun validasi. Setiap citra dimasukkan ke dalam model untuk menghasilkan probabilitas prediksi pada sepuluh kelas. Kelas dengan nilai probabilitas tertinggi ditetapkan sebagai hasil prediksi. Performa klasifikasi dievaluasi menggunakan accuracy, precision, recall, F1-score, dan *confusion matrix*. Accuracy menunjukkan proporsi prediksi yang benar terhadap seluruh data uji. Precision menunjukkan ketepatan model ketika menetapkan suatu citra ke dalam kelas tertentu, sedangkan recall menunjukkan kemampuan model mengenali seluruh citra aktual pada kelas tersebut (Nurhidayat & Kania Evita Dewi, 2023). F1-score merupakan rata-rata harmonik antara precision dan recall sehingga dapat digunakan untuk menilai keseimbangan keduanya. Nilai precision, recall, dan F1-score dihitung menggunakan pendekatan *macro average* agar setiap kelas memperoleh bobot yang sama. *Confusion matrix* digunakan untuk mengetahui jumlah prediksi benar serta pola kesalahan klasifikasi pada setiap kelas (A. Y. Pratama & Ghazi, 2025).

Selain performa prediksi, penelitian ini mengevaluasi efisiensi model menggunakan *training time*, *model size*, dan *inference time per image*. *Training time* dihitung sejak pelatihan dimulai hingga selesai. *Model size* dihitung berdasarkan ukuran file model yang telah disimpan dalam satuan megabita. Sementara itu, *inference time per image* dihitung dengan membagi total waktu prediksi terhadap 376 citra uji dengan jumlah citra yang diuji. Seluruh model diuji pada lingkungan komputasi yang sama. Hasil evaluasi kemudian dibandingkan untuk mengetahui model dengan performa klasifikasi terbaik dan model yang paling efisien. Model dengan *accuracy* dan *F1-score* tertinggi dipertimbangkan untuk sistem yang mengutamakan performa klasifikasi, sedangkan model dengan ukuran file lebih kecil dan waktu inferensi rendah dipertimbangkan untuk aplikasi mobile atau sistem dengan sumber daya terbatas.

3. HASIL DAN PEMBAHASAN

3.1 Implementasi Algoritma

Penelitian ini mengimplementasikan tiga arsitektur transfer learning, yaitu VGG16, ResNet50, dan MobileNetV2. Ketiga model menggunakan dataset, pembagian data, konfigurasi pelatihan, dan metrik evaluasi yang sama. Pendekatan ini digunakan agar perbedaan hasil lebih merepresentasikan karakteristik masing-masing arsitektur dan bukan disebabkan oleh perbedaan skenario eksperimen. Proses klasifikasi dimulai dengan memasukkan citra jajan tradisional yang telah diubah ukurannya menjadi 224×224 piksel. Dataset kemudian dibagi menjadi data latih, validasi, dan uji dengan rasio 70:15:15 dan pembagian tersebut menghasilkan 1.706 citra latih, 363 citra validasi, dan 376 citra uji. Augmentasi diterapkan hanya pada data latih melalui transformasi berupa pembalikan horizontal, rotasi, zoom, translasi, dan penyesuaian kontras. Data validasi dan data uji tidak diberi augmentasi agar proses evaluasi menggunakan citra asli yang tidak dimodifikasi. Setiap citra diteruskan ke convolutional base dari VGG16, ResNet50, atau MobileNetV2 kemudian lapisan konvolusi mengekstraksi fitur visual secara bertahap. Pada lapisan awal, model menangkap fitur sederhana seperti garis, tepi, arah, perubahan warna, dan pola permukaan. Pada lapisan yang lebih dalam, fitur tersebut dikombinasikan menjadi representasi yang lebih kompleks, seperti bentuk jajan, tekstur, taburan, struktur permukaan, dan pola penyajian.

Feature map yang dihasilkan oleh convolutional base kemudian diringkas melalui Global Average Pooling menjadi vektor fitur. Vektor tersebut diproses oleh Dense layer sebanyak 256 neuron dengan aktivasi ReLU untuk mempelajari hubungan antar fitur. Dropout sebesar 0,2 digunakan untuk mengurangi ketergantungan model terhadap neuron tertentu selama pelatihan. Lapisan keluaran menggunakan sepuluh neuron dengan aktivasi softmax untuk menghasilkan nilai probabilitas bagi sepuluh kelas jajan. Kelas dengan probabilitas tertinggi ditetapkan sebagai hasil prediksi. Selama pelatihan, hasil prediksi dibandingkan dengan label aktual menggunakan fungsi kerugian categorical crossentropy. Optimizer Adam dengan learning rate 0,0001 digunakan untuk memperbarui bobot pada classification head. Seluruh lapisan pada convolutional base dibekukan sehingga pembaruan bobot hanya dilakukan pada bagian klasifikasi yang baru ditambahkan serta ketiga model dilatih selama maksimum 10 epoch dengan batch size 64.

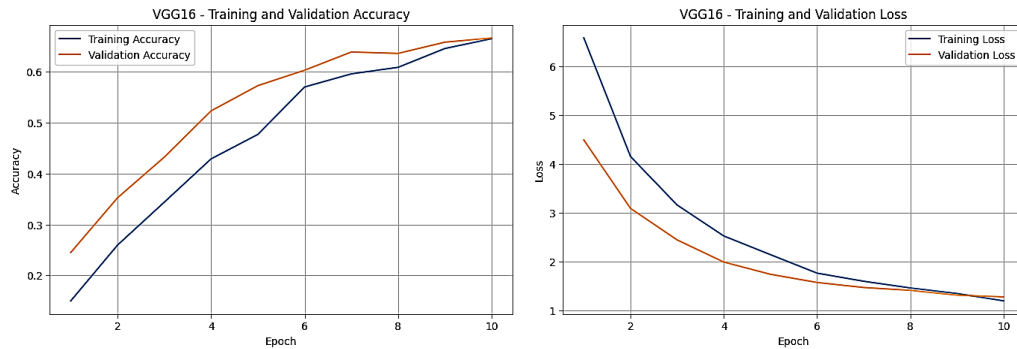
Setelah proses pelatihan selesai, setiap model diuji menggunakan 376 citra uji. Hasil prediksi dibandingkan dengan label sebenarnya untuk memperoleh accuracy, macro precision, macro recall, macro F1-score, dan confusion matrix. Selain performa klasifikasi, penelitian juga mengukur training time, model size, dan inference time per image. Dengan tahapan tersebut, masalah penelitian diselesaikan melalui proses yang terstruktur, mulai dari pengolahan citra, ekstraksi fitur, pembelajaran pola visual, klasifikasi, hingga evaluasi performa dan efisiensi.

3.2 Performa Pelatihan dan Validasi VGG16

Hasil training dan validasi model VGG16 mengalami peningkatan secara bertahap selama 10 epoch, training accuracy VGG16 meningkat dari 0,1501 pada epoch pertama menjadi 0,6653 pada epoch kesepuluh. Nilai validation accuracy juga meningkat dari 0,2452 menjadi 0,6667. Pada saat yang sama, nilai training loss dan validation loss mengalami penurunan secara bertahap. Pola tersebut menunjukkan bahwa VGG16 mampu mempelajari karakteristik visual pada dataset, tetapi laju pembelajarannya lebih lambat dibandingkan ResNet50 dan MobileNetV2, perubahan nilai *accuracy* dan *loss* selama proses pelatihan VGG16 ditunjukkan pada Gambar 4.

Berdasarkan Gambar 4, nilai training accuracy dan validation accuracy pada akhir pelatihan berada pada tingkat yang hampir sama. Kondisi ini menunjukkan bahwa tidak terdapat kesenjangan generalisasi yang besar. Namun, kurva

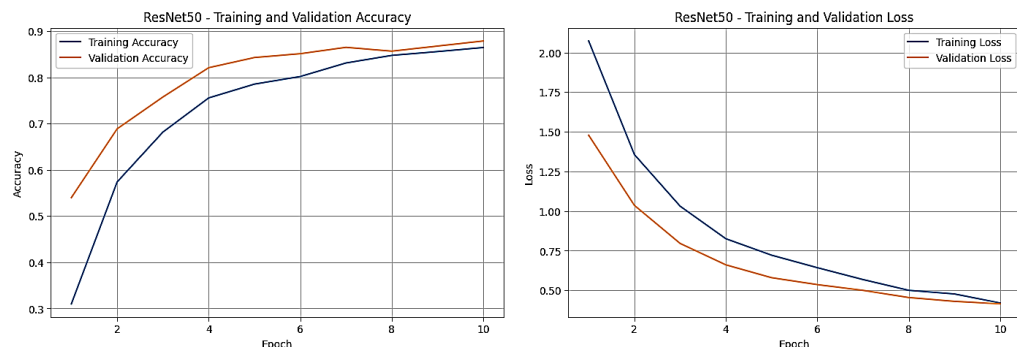
yang masih meningkat hingga epoch terakhir mengindikasikan bahwa VGG16 belum mencapai kondisi konvergen. Dengan kata lain, hasil VGG16 pada penelitian ini lebih menggambarkan performa pada skenario pelatihan singkat daripada kemampuan maksimal arsitektur tersebut. VGG16 memiliki struktur berurutan dengan jumlah parameter yang relatif besar. Dalam penelitian ini, seluruh lapisan konvolusional dibekukan sehingga model hanya memperbarui bobot pada classification head. Strategi tersebut membatasi kemampuan model untuk menyesuaikan fitur tingkat tinggi dengan karakteristik lokal dataset. Kondisi ini menjadi penting karena citra jajanan tradisional tidak hanya dibedakan melalui bentuk umum, tetapi juga melalui perbedaan tekstur, taburan, warna, pola permukaan, dan cara penyajian yang relatif halus.



Gambar 4. Training Validation Accuracy and Loss of VGG16

3.3 Performa Pelatihan dan Validasi ResNet50

Hasil pelatihan pada model ResNet50 menunjukkan performa pembelajaran yang lebih kuat dibandingkan dengan VGG16. Performa training dan validasi ResNet50 ditunjukkan pada Gambar 5.



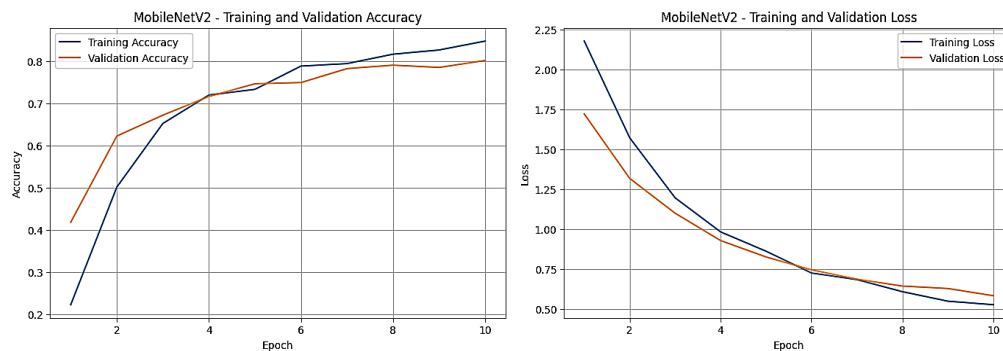
Gambar 5. Training Validation Accuracy and Loss of ResNet50

Berdasarkan Gambar 5, training accuracy ResNet50 meningkat dari 0,3101 pada epoch pertama menjadi 0,8646 pada epoch kesepuluh. Nilai validation accuracy mencapai 0,8788. Nilai tersebut merupakan performa validasi tertinggi di antara ketiga model. Kurva training loss dan validation loss juga menunjukkan penurunan yang konsisten. Nilai validation accuracy yang sedikit lebih tinggi daripada training accuracy dapat dipengaruhi oleh augmentasi data. Citra latih telah mengalami transformasi sehingga memiliki variasi visual yang lebih tinggi dan menjadi lebih sulit dikenali dibandingkan data validasi. Oleh karena itu, performa validasi yang sedikit lebih tinggi tidak secara langsung menunjukkan adanya masalah pada proses pelatihan. Keunggulan ResNet50 tidak hanya disebabkan oleh jumlah lapisannya yang lebih dalam, tetapi juga oleh keberadaan residual connection. Mekanisme ini memungkinkan informasi dari lapisan sebelumnya diteruskan langsung ke lapisan berikutnya. Dengan demikian, fitur awal seperti tepi, tekstur, dan pola warna tetap dapat dipertahankan ketika jaringan membentuk representasi yang lebih kompleks. Karakteristik tersebut sesuai dengan tantangan pada dataset jajanan tradisional. Beberapa kelas memiliki bentuk, warna, dan tekstur yang saling menyerupai. Pada kondisi tersebut, model perlu mempertahankan fitur tingkat rendah sekaligus menggabungkannya dengan fitur tingkat tinggi. ResNet50 lebih mampu melakukan proses tersebut karena jalur residual membantu mempertahankan informasi penting selama proses ekstraksi fitur.

3.4 Performa Pelatihan dan Validasi MobileNetV2

Pada model MobileNetV2 juga menghasilkan performa pembelajaran yang kompetitif dibandingkan dua model sebelumnya. Performa training dan validasi MobileNetV2 ditunjukkan pada Gambar 6. Berdasarkan Gambar 6, training accuracy MobileNetV2 meningkat dari 0,2233 menjadi 0,8476. Nilai validation accuracy mencapai 0,8017. Nilai training loss dan validation loss juga menurun selama proses pelatihan. Hasil tersebut menunjukkan bahwa MobileNetV2 mampu mempelajari pola visual dengan baik meskipun memiliki ukuran arsitektur yang lebih kecil. MobileNetV2 menggunakan inverted residual, linear bottleneck, dan depthwise convolution untuk mengurangi jumlah

parameter dan operasi komputasi. Struktur ini memungkinkan model tetap memperoleh performa yang kompetitif dengan ukuran file yang jauh lebih kecil. Namun, kapasitas representasinya lebih terbatas dibandingkan ResNet50. Pada kelas dengan karakteristik visual yang cukup khas, MobileNetV2 mampu memberikan hasil yang sangat baik. Sebaliknya, ketika beberapa kelas memiliki warna, tekstur, atau pola penyajian yang hampir sama, model cenderung menghasilkan tumpang tindih prediksi. Dengan demikian, MobileNetV2 memberikan kompromi antara kemampuan klasifikasi dan efisiensi model.



Gambar 6. Training Validation Accuracy and Loss of MobileNetV2

3.5 Hasil Pengujian Model

Pengujian dilakukan menggunakan 376 citra yang tidak digunakan pada tahap pelatihan maupun validasi. Tahap ini bertujuan untuk mengetahui kemampuan setiap model dalam mengenali citra baru. Ketiga arsitektur dibandingkan menggunakan metrik klasifikasi dan indikator efisiensi, yaitu accuracy, precision, recall, F1-score, training time, model size, dan inference time per image. Perbandingan lengkap model ditunjukkan pada Gambar 7.

Model	Accuracy	Precision	Recall	F1-Score	Training Time (seconds)	Model Size (MB)	Inference Time/Image (seconds)
VGG16	0.736702	0.756293	0.748050	0.746975	201.430339	57.773004	0.035346
ResNet50	0.840426	0.851486	0.851632	0.848898	206.879360	96.662491	0.023669
MobileNetV2	0.824468	0.857080	0.829422	0.836897	196.249075	12.977555	0.027273

Gambar 7. Perbandingan Performa Model

Berdasarkan Gambar 7, ResNet50 memperoleh performa terbaik dengan accuracy 84,04%, macro precision 85,15%, macro recall 85,16%, dan macro F1-score 84,89%. Nilai precision, recall, dan F1-score yang relatif seimbang menunjukkan bahwa model tidak hanya akurat secara keseluruhan, tetapi juga cukup konsisten dalam mengenali berbagai kelas. MobileNetV2 memperoleh accuracy 82,45%, hanya terpaut 1,59 poin persentase dari ResNet50. Model ini menghasilkan macro precision tertinggi, yaitu 85,71%, sedangkan macro recall dan macro F1-score masing-masing sebesar 82,94% dan 83,69%. Nilai precision yang tinggi menunjukkan bahwa prediksi MobileNetV2 cenderung tepat ketika menetapkan suatu citra ke kelas tertentu. Namun, nilai recall yang lebih rendah menunjukkan masih terdapat citra aktual yang tidak terdeteksi secara optimal. VGG16 memperoleh accuracy 73,67%, macro precision 75,63%, macro recall 74,81%, dan macro F1-score 74,70%. Performa tersebut lebih rendah dibandingkan ResNet50 dan MobileNetV2. Selisih accuracy antara VGG16 dan ResNet50 mencapai 10,37 poin persentase. Hasil ini menunjukkan bahwa pada konfigurasi pelatihan yang sama dan tanpa fine-tuning, VGG16 belum mampu menghasilkan representasi fitur sebaik dua model lainnya. Dari sisi waktu pelatihan, MobileNetV2 menjadi model tercepat dengan durasi 196,25 detik. VGG16 membutuhkan 201,40 detik, sedangkan ResNet50 membutuhkan 206,87 detik. Selisih waktu pelatihan tidak terlalu besar karena seluruh convolutional base dibekukan.

Perbedaan yang paling mencolok terlihat pada ukuran model. MobileNetV2 hanya memiliki ukuran 12,98 MB, sedangkan VGG16 berukuran 57,77 MB dan ResNet50 mencapai 96,66 MB. Ukuran MobileNetV2 sekitar 77,5% lebih kecil daripada VGG16 dan sekitar 86,6% lebih kecil daripada ResNet50. Hasil ini memperlihatkan keunggulan MobileNetV2 untuk sistem dengan keterbatasan penyimpanan. ResNet50 mencatat inference time tercepat, yaitu 0,0237 detik per citra, meskipun memiliki ukuran model paling besar. MobileNetV2 memperoleh waktu inferensi 0,0273 detik, sedangkan VGG16 membutuhkan 0,0354 detik. Hasil ini menunjukkan bahwa ukuran model tidak selalu berbanding lurus dengan waktu inferensi. Kecepatan prediksi juga dipengaruhi oleh struktur operasi, optimalisasi perangkat lunak, dan lingkungan komputasi.

3.6 Analisis Performa per Kelas

3.6.1 VGG16

Evaluasi per kelas pada VGG16 ini memberikan gambaran yang lebih rinci mengenai kemampuan model dalam mengenali setiap kategori jajanan. Analisis per kelas diperlukan karena accuracy keseluruhan belum cukup untuk



menjelaskan apakah model bekerja secara merata pada seluruh kelas. Hasil evaluasi VGG16 dapat dilihat pada gambar 8.

```
Accuracy : 0.7367021276595744
Precision: 0.756292875475376
Recall   : 0.7480495984133563
F1-score : 0.7469753141810112
Total inference time: 13.29002070426941
Inference time/image: 0.035345799745397366
```

	precision	recall	f1-score	support
batun_bedil	0.79	0.79	0.79	29
bubur_injin	0.87	0.90	0.89	30
jaje_lukis	0.74	0.60	0.67	53
jaje_piling	0.74	0.69	0.71	29
jaje_wajik	0.76	0.72	0.74	36
kaliadrem	0.95	0.77	0.85	26
klepon	0.70	0.82	0.76	55
laklak	0.62	0.88	0.73	32
ongol_ongol	0.74	0.62	0.67	47
pisang_rai	0.63	0.69	0.66	39
accuracy			0.74	376
macro avg	0.76	0.75	0.75	376
weighted avg	0.75	0.74	0.74	376

Gambar 8. Hasil Evaluasi Model VGG16

Berdasarkan Gambar 8, VGG16 menunjukkan performa terbaik pada bubur injin dengan *F1-score* 0,89. Karakteristik warna gelap, tekstur bubur, dan bentuk penyajian yang relatif konsisten kemungkinan membuat kelas tersebut lebih mudah dikenali. Kaliadrem memiliki *precision* tinggi sebesar 0,95, tetapi *recall* hanya 0,77. Hal ini menunjukkan bahwa prediksi kaliadrem cenderung tepat, tetapi model masih gagal mengenali sebagian citra kaliadrem aktual. Variasi sudut pengambilan gambar, bentuk, dan warna permukaan dapat menjadi penyebab sebagian citra tidak terdeteksi. Kelas dengan performa terendah adalah pisang rai, jaje lukis, dan ongol-ongol. Ketiga kelas tersebut memiliki *F1-score* masing-masing 0,66, 0,67, dan 0,67. Rendahnya performa menunjukkan bahwa VGG16 belum mampu membentuk fitur yang cukup diskriminatif. Pisang rai dan ongol-ongol dapat memiliki warna pucat serta lapisan kelapa yang serupa, sedangkan jaje lukis memiliki variasi bentuk dan warna yang cukup tinggi. Selain itu, laklak memiliki *recall* 0,88 tetapi *precision* hanya 0,62. Artinya, sebagian besar citra laklak aktual dapat dikenali, tetapi sejumlah citra dari kelas lain juga diprediksi sebagai laklak. Pola ini menunjukkan bahwa model terlalu mengandalkan fitur dominan seperti warna hijau, bentuk pipih, atau tekstur permukaan, tanpa cukup mempertimbangkan struktur objek secara keseluruhan.

3.6.2 ResNet50

Evaluasi per kelas pada ResNet50 memberikan penjelasan yang lebih detail mengenai perilaku klasifikasi model pada sepuluh kelas jajanan. Model ResNet50 memberikan performa paling merata dibandingkan dengan model VGG16 dan MobileNetV2. Pada batun bedil memperoleh *precision*, *recall*, dan *F1-score* sebesar 0,93. Hal ini menunjukkan bahwa bentuk, warna, dan pola penyajian batun bedil cukup konsisten untuk dikenali. Sedangkan, bubur injin memperoleh *recall* 1,00 dan seluruh citra bubur injin pada data uji berhasil dikenali. Karakter warna yang lebih gelap dan tekstur bubur yang khas menjadi ciri visual yang kuat dibandingkan kelas lain. Kelas jaje wajik memiliki *precision* 0,97 dan *recall* 0,81. Untuk hasil klasifikasi per kelas pada model ResNet50 dapat dilihat pada gambar 9.

```
Accuracy : 0.8404255319148937
Precision: 0.8514856216580355
Recall   : 0.8516316682075136
F1-score : 0.8488981859922815
Total inference time: 8.899528980255127
Inference time/image: 0.02366896005387002
```

	precision	recall	f1-score	support
batun_bedil	0.93	0.93	0.93	29
bubur_injin	0.83	1.00	0.91	30
jaje_lukis	0.80	0.75	0.78	53
jaje_piling	0.82	0.79	0.81	29
jaje_wajik	0.97	0.81	0.88	36
kaliadrem	0.88	0.88	0.88	26
klepon	0.80	0.87	0.83	55
laklak	0.88	0.88	0.88	32
ongol_ongol	0.82	0.70	0.76	47
pisang_rai	0.78	0.90	0.83	39
accuracy			0.84	376
macro avg	0.85	0.85	0.85	376
weighted avg	0.84	0.84	0.84	376

Gambar 9. Hasil Evaluasi Model ResNet50

Berdasarkan Gambar 9, nilai precision yang sangat tinggi menunjukkan bahwa prediksi jaje wajik hampir selalu benar. Namun, recall yang lebih rendah menunjukkan masih terdapat citra jaje wajik aktual yang diklasifikasikan ke kelas lain. Kemungkinan penyebabnya adalah variasi bentuk potongan, warna gula, dan latar penyajian. Kelas dengan performa terendah adalah ongol-ongol dan jaje lukis. Meskipun demikian, F1-score keduanya masih lebih tinggi daripada hasil VGG16. Hal ini menunjukkan bahwa mekanisme residual membantu model mempertahankan fitur penting. Namun, variasi bentuk, tekstur lembut, warna yang berdekatan, dan pola permukaan yang tidak konsisten tetap menjadi tantangan.

3.6.3 MobileNetV2

Evaluasi per kelas pada MobileNetV2 sangat penting dilakukan, karena MobileNetV2 dirancang sebagai arsitektur ringan. Analisis ini membantu menilai apakah model yang lebih kecil tetap dapat memberikan hasil klasifikasi yang andal pada citra jajanan tradisional Bali. Untuk hasil evaluasi MobileNetV2 dapat dilihat pada gambar 10.

Accuracy :	0.824468085106383			
Precision:	0.857080156727754			
Recall :	0.8294222299909931			
F1-score :	0.8368965255935155			
Total inference time:	10.25479245185852			
Inference time/image:	0.02727338418047479			

Classification Report:				
	precision	recall	f1-score	support
batun_bedil	0.96	0.86	0.91	29
bubur_injin	0.83	1.00	0.91	30
jaje_lukis	0.80	0.70	0.75	53
jaje_piling	0.92	0.79	0.85	29
jaje_wajik	0.85	0.92	0.88	36
kaliadrem	1.00	0.85	0.92	26
klepon	0.67	0.95	0.78	55
laklak	0.85	0.72	0.78	32
ongol_ongol	0.90	0.74	0.81	47
pisang_rai	0.79	0.77	0.78	39
accuracy			0.82	376
macro avg	0.86	0.83	0.84	376
weighted avg	0.84	0.82	0.82	376

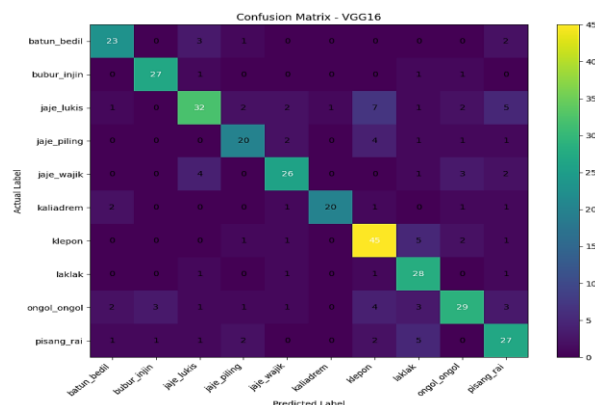
Gambar 10. Hasil Evaluasi Model MobileNetV2

Berdasarkan Gambar 10, MobileNetV2 memperoleh performa tertinggi pada kaliadrem dengan precision 1,00 dan F1-score 0,92. Hasil tersebut menunjukkan bahwa ciri visual kaliadrem cukup spesifik bagi model ringan. Kelas klepon memiliki recall sangat tinggi, yaitu 0,95, tetapi precision hanya 0,67. Hal ini berarti hampir seluruh citra klepon berhasil dikenali, tetapi sejumlah citra non-klepon juga diprediksi sebagai klepon. Warna hijau, bentuk membulat, dan taburan kelapa kemungkinan menjadi fitur dominan yang terlalu kuat. Jaje lukis kembali menjadi kelas dengan performa terendah. Variasi warna, bentuk, dan penyajian diduga membuat MobileNetV2 kesulitan membentuk batas kelas yang stabil. Struktur ringan yang mengutamakan efisiensi dapat mengurangi kemampuan model dalam menangkap kombinasi fitur yang sangat detail.

3.7 Analisis Confusion Matrix

3.7.1 Confusion Matrix VGG16

Confusion matrix digunakan untuk mengamati perilaku prediksi setiap model secara lebih rinci. Jika accuracy, precision, recall, dan F1-score memberikan ringkasan numerik, confusion matrix menunjukkan kelas mana yang berhasil dikenali dengan benar dan kelas mana yang sering tertukar. Analisis ini sangat berguna karena beberapa jajanan tradisional Bali memiliki karakteristik visual yang mirip. Confusion matrix VGG16 ditunjukkan pada Gambar 11.

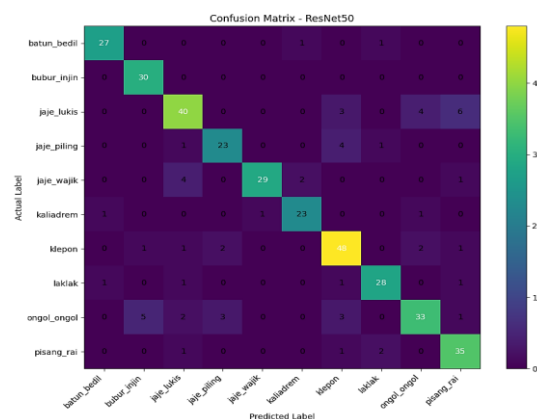


Gambar 11. Confusion Matrix VGG16

Berdasarkan Gambar 11, prediksi benar terkonsentrasi pada diagonal utama, tetapi kesalahan di luar diagonal masih cukup tersebar. Bubur injin dan kaliadrem relatif mudah dikenali karena memiliki karakter visual yang lebih khas. Sebaliknya, jaje lukis, ongol-ongol, dan pisang rai menunjukkan kesalahan yang lebih tinggi. Kesalahan tersebut tidak hanya menunjukkan rendahnya jumlah prediksi benar, tetapi juga menggambarkan kegagalan model dalam membedakan kombinasi fitur. Jaje lukis memiliki variasi bentuk dan warna yang cukup tinggi. Ongol-ongol dan pisang rai dapat berbagi warna terang, tekstur lembut, dan taburan kelapa. VGG16 kemungkinan lebih mengandalkan tekstur dan warna umum, tetapi belum cukup mampu mengintegrasikan fitur bentuk, susunan objek, dan pola penyajian.

3.7.2 Confusion Matrix ResNet50

Hasil confusion matrix dari ResNet50 menunjukkan pola klasifikasi yang lebih seimbang dibandingkan dengan VGG16. Matriks ini memperlihatkan hubungan antara label aktual dan label hasil prediksi, sehingga dapat digunakan untuk mengidentifikasi kelas yang berhasil dikenali dengan baik maupun kelas yang masih sering tertukar. Hasil confusion matrix ResNet50 dapat dilihat pada Gambar 12.

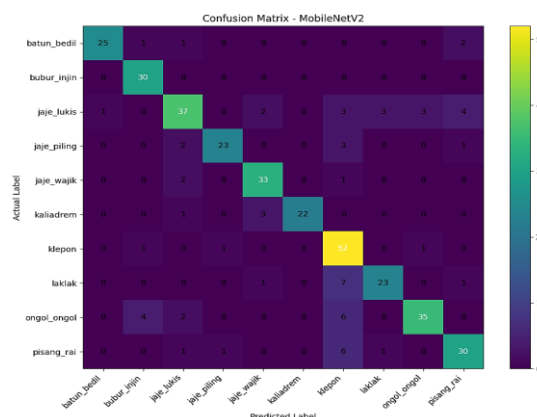


Gambar 12. Confusion Matrix ResNet50

Berdasarkan Gambar 12, hasil confusion matrix ResNet50 menunjukkan konsentrasi diagonal yang lebih kuat dibandingkan VGG16. Hal ini menunjukkan bahwa model mampu membentuk batas kelas yang lebih jelas. Batun bedil, bubur injin, kaliadrem, laklak, dan jaje wajik memiliki prediksi benar yang relatif tinggi. Kesalahan masih muncul pada jaje lukis dan ongol-ongol. Fitur yang kemungkinan sulit dibedakan meliputi tekstur lembut, warna yang berdekatan, permukaan yang tidak seragam, serta variasi latar belakang. Meskipun residual connection membantu mempertahankan informasi visual, fitur yang sangat dipengaruhi cara penyajian masih dapat menghasilkan tumpang tindih kelas.

3.7.3 Confusion Matrix MobileNetV2

Hasil confusion matrix dari model MobileNetV2 memberikan hasil yang kompetitif, meskipun akurasi sedikit di bawah ResNet50. Confusion matrix MobileNetV2 ditunjukkan pada Gambar 13.



Gambar 13. Confusion Matrix MobileNetV2

Berdasarkan Gambar 13, confusion matrix MobileNetV2 menunjukkan prediksi yang cukup baik pada batun bedil, bubur injin, jaje wajik, kaliadrem, dan klepon. Namun, model cenderung mengklasifikasikan beberapa citra non-klepon sebagai klepon. Hal ini sesuai dengan recall klepon yang tinggi dan precision yang lebih rendah. Pola tersebut menunjukkan bahwa MobileNetV2 kemungkinan memberi bobot besar pada warna hijau, bentuk membulat, dan taburan kelapa. Fitur tersebut efektif untuk mendeteksi klepon, tetapi kurang diskriminatif ketika kelas lain memiliki karakteristik serupa. Kesalahan ini menunjukkan bahwa model ringan dapat menangkap fitur dominan, tetapi lebih



terbatas dalam menggabungkan fitur detail seperti struktur permukaan, proporsi objek, dan pola penyajian. Secara keseluruhan, analisis confusion matrix menunjukkan bahwa kesalahan klasifikasi tidak terjadi secara acak. Kesalahan terkonsentrasi pada kelas dengan variasi intra-kelas tinggi dan kemiripan antar-kelas yang kuat. Warna, tekstur, taburan kelapa, bentuk, dan latar belakang menjadi fitur yang paling memengaruhi proses klasifikasi.

3.8 Analisis Struktural Keunggulan ResNet50

Keunggulan ResNet50 pada penelitian ini dapat dijelaskan melalui hubungan antara struktur arsitektur dan karakteristik dataset. Dataset jajanan tradisional memiliki sejumlah kelas yang berbeda secara halus. Perbedaanannya tidak selalu terlihat dari bentuk utama, tetapi dari kombinasi warna, tekstur, permukaan, taburan, dan penyajian. ResNet50 menggunakan residual block yang memungkinkan informasi dari lapisan awal dipertahankan melalui jalur pintas. Dengan mekanisme ini, fitur sederhana seperti tepi dan pola warna tetap dapat digunakan bersama fitur tingkat tinggi seperti bentuk dan tekstur kompleks. Kemampuan menggabungkan fitur multitingkat tersebut relevan untuk membedakan jajanan yang memiliki tampilan hampir serupa.

VGG16 tidak memiliki jalur residual dan menggunakan aliran konvolusi berurutan. Dalam kondisi convolutional base dibekukan, representasi fitur menjadi kurang fleksibel untuk menyesuaikan diri dengan karakter lokal dataset. MobileNetV2 memiliki struktur yang efisien, tetapi pengurangan operasi dan parameter juga membatasi kapasitas representasi pada kelas dengan perbedaan yang sangat halus. Dengan demikian, ResNet50 unggul bukan semata-mata karena jaringan lebih dalam, tetapi karena mekanisme residual memungkinkan informasi visual penting tetap dipertahankan dan digabungkan sepanjang jaringan.

3.9 Pembahasan

Penelitian sebelumnya mengenai klasifikasi jajanan tradisional Bali menggunakan VGG16 dan memperoleh accuracy 97,5% dengan konfigurasi dropout 20%, batch size 64, serta pelatihan hingga 1.000 epoch (Rahayu et al., 2024). Nilai tersebut lebih tinggi dibandingkan hasil VGG16 sebesar 73,67% pada penelitian ini. Namun, kedua penelitian memiliki tujuan yang berbeda. Penelitian terdahulu berfokus pada optimasi satu arsitektur, sedangkan penelitian ini membandingkan tiga arsitektur menggunakan konfigurasi yang sama. Penelitian ini hanya menggunakan 10 epoch dan tidak melakukan fine-tuning. Kurva VGG16 juga menunjukkan bahwa model masih mengalami peningkatan hingga epoch terakhir. Oleh karena itu, hasil penelitian ini lebih tepat digunakan untuk membandingkan performa relatif antararsitektur, bukan untuk menunjukkan performa maksimal setiap model.

Hasil penelitian ini juga sejalan dengan studi komparatif lain yang menunjukkan bahwa kinerja arsitektur CNN dipengaruhi oleh karakteristik dataset dan tujuan implementasi. ResNet50 sering menunjukkan performa yang kuat pada tugas yang membutuhkan representasi fitur dalam, sedangkan MobileNetV2 lebih banyak digunakan ketika efisiensi ukuran model menjadi pertimbangan utama. Perbedaan hasil antarpelatihan menegaskan bahwa pemilihan arsitektur tidak dapat hanya didasarkan pada satu nilai accuracy. Jumlah epoch, strategi fine-tuning, ukuran dataset, distribusi kelas, kualitas citra, dan lingkungan komputasi juga memengaruhi hasil akhir.

Hasil penelitian menghasilkan dua rekomendasi utama. ResNet50 lebih sesuai digunakan pada sistem yang memprioritaskan performa klasifikasi, seperti katalog digital berbasis server, aplikasi web, atau layanan pengenalan citra dengan sumber daya komputasi yang memadai. Model ini memperoleh accuracy dan F1-score tertinggi serta waktu inferensi tercepat pada lingkungan pengujian. MobileNetV2 lebih sesuai digunakan ketika ukuran model dan efisiensi penyimpanan menjadi prioritas. Dengan ukuran 12,98 MB, model ini jauh lebih ringan daripada VGG16 dan ResNet50, sementara selisih accuracy terhadap ResNet50 hanya 1,59 poin persentase. Karakteristik tersebut membuat MobileNetV2 lebih potensial untuk aplikasi mobile atau media edukasi ringan. VGG16 tetap bermanfaat sebagai baseline, tetapi membutuhkan pelatihan lebih panjang atau fine-tuning agar dapat mencapai performa yang lebih kompetitif. Hasil ini menunjukkan bahwa pemilihan model perlu disesuaikan dengan kebutuhan platform, bukan hanya berdasarkan nilai akurasi.

Penelitian ini memiliki beberapa keterbatasan. Pertama, setiap model hanya dilatih selama 10 epoch tanpa tahap fine-tuning pada lapisan konvolusional. Kedua, eksperimen dilakukan satu kali untuk setiap arsitektur sehingga belum menghasilkan rata-rata dan standar deviasi dari beberapa pengulangan. Ketiga, distribusi jumlah citra pada setiap kelas belum sepenuhnya seimbang. Keempat, pengukuran waktu inferensi dilakukan pada satu lingkungan komputasi dan belum diuji langsung pada perangkat mobile atau sistem produksi. Kelima, variasi latar belakang, pencahayaan, sudut pengambilan gambar, dan cara penyajian belum sepenuhnya dikendalikan. Penelitian selanjutnya dapat melakukan fine-tuning, pengulangan eksperimen, optimasi hiperparameter, penyeimbangan kelas, serta pengujian langsung pada perangkat target. Penggunaan visualisasi fitur seperti Grad-CAM juga dapat dipertimbangkan untuk mengidentifikasi area citra yang paling memengaruhi keputusan model.

4. KESIMPULAN

Penelitian ini menunjukkan bahwa perbandingan VGG16, ResNet50, dan MobileNetV2 pada konfigurasi pelatihan yang sama menghasilkan perbedaan yang jelas antara kemampuan klasifikasi dan efisiensi implementasi. ResNet50 memberikan hasil paling konsisten dalam mengenali sepuluh kelas jajanan tradisional dengan accuracy 84,04% dan F1-score 84,89%, yang menunjukkan bahwa mekanisme residual lebih efektif dalam mempertahankan serta



menggabungkan fitur visual halus seperti tekstur, bentuk, warna, dan pola penyajian. MobileNetV2 menghasilkan performa yang mendekati ResNet50 dengan accuracy 82,45% dan ukuran model hanya 12,98 MB, sehingga lebih sesuai untuk sistem yang membutuhkan penyimpanan kecil dan implementasi ringan, sedangkan VGG16 memperoleh performa terendah pada skenario 10 epoch tanpa fine-tuning. Hasil analisis per kelas dan confusion matrix juga menunjukkan bahwa kesalahan klasifikasi terutama terjadi pada kelas dengan kemiripan warna, tekstur, taburan kelapa, serta variasi bentuk dan latar belakang yang tinggi. Penelitian ini masih terbatas pada satu kali eksperimen untuk setiap model, jumlah epoch yang singkat, convolutional base yang seluruhnya dibekukan, distribusi kelas yang belum sepenuhnya seimbang, serta pengujian yang belum dilakukan langsung pada perangkat mobile atau lingkungan produksi. Oleh karena itu, penelitian selanjutnya perlu melakukan pengulangan eksperimen untuk memperoleh rata-rata dan standar deviasi, membuka sebagian lapisan akhir untuk fine-tuning, menerapkan class weighting atau penyeimbangan data pada kelas yang kurang terwakili, menggunakan Grad-CAM untuk memverifikasi area visual yang memengaruhi prediksi, serta menguji MobileNetV2 dan ResNet50 secara langsung pada perangkat target untuk membandingkan latensi, penggunaan memori, dan stabilitas prediksi dalam kondisi nyata

REFERENCES

- Akmal, L., Yadi, I. Z., Kunang, Y. N., & Sari, F. (2026). Pengembangan Aplikasi Presensi Berbasis Deep Learning. *Jurnal Ilmu Komputer Dan Sistem Informasi (JIRSI)*, 5(2), 219–233. <https://doi.org/10.70340/jirsi.v5i2.354>
- Alfan, D. S., & Kumalasari, I. (2026). Implementasi Model Deep Learning MobileNetV2 untuk Klasifikasi Citra Melanoma Berbasis Web. *Journal of Information System Research (JOSH)*, 7(3), 670–680. <https://doi.org/10.47065/josh.v7i3.8848>
- Amini, A., Nurul, F., & Zalmi, H. (2025). Strategi Preservasi Pengetahuan Tradisional Minangkabau Melalui Digitalisasi Tradisi Lisan Di Perpustakaan Perguruan Tinggi. *Jurnal Ilmu Perpustakaan Dan Informasi Islam*, 5(1), 76–86. <https://doi.org/10.37108/almaarif.v5i1.2316>
- Arumsari, D., Kharisma, & Ulfi Saidata Aesyi. (2024). Sistem Chatbot Layanan Informasi Mahasiswa Menggunakan Algoritma Long Short-Term Memory. *Indonesian Journal on Data Science*, 2(2), 77–86. <https://doi.org/10.30989/ijds.v2i2.1489>
- Bagaskara, M., Putra, D., Bratayadnya, P. A., & Raharjo, A. (2025). Jajanan Khas Bali Dalam Fotografi Komersial. *Retina Jurnal Fotografi*, 5(1), 118–127. <https://doi.org/10.59997/rjf.v5i1.5270>
- Chen, L., Li, S., Bai, Q., Yang, J., Jiang, S., & Miao, Y. (2021). Review of Image Classification Algorithms Based on Convolutional Neural Networks. *Remote Sensing*, 13(22). <https://doi.org/10.3390/rs13224712>
- Erlanda, E., & Krisnadi, A. R. (2025). Pengaruh Perubahan Prefrensi Makanan Dan Dampak Tren Kuliner Terhadap Minat Beli Makanan Tradisional Generasi Z Di Jakarta. *Pendas : Jurnal Ilmiah Pendidikan Dasar*, 10(September), 230–244. <https://doi.org/10.23969/jp.v10i03.30741>
- Felicia Lie, A., Susanto, J. E., Sutarman, M. O., Simon, S., & Tiatri, S. (2026). Peran Makanan Tradisional Dalam Pembentukan Identitas Budaya: Suatu Tinjauan Sistematis. *CENDEKIA: Jurnal Ilmu Pengetahuan*, 6(1), 80–89. <https://doi.org/10.51878/cendekia.v6i1.8195>
- Fransiska, M., Wiradnyani, N. K., & Suryanto, I. W. (2026). Analisis Popularitas dan Profitabilitas Jajanan Khas Bali pada Toko JASTIP di Wilayah Badung dengan Pendekatan Menu Engineering. *Jurnal Ekonomi, Manajemen, Akuntansi Dan Keuangan*, 7(1), 1–14. <https://doi.org/10.53697/emak.v7i1.3479>
- Harsana, M., Rinawati, W., & Fauziah, A. (2023). Inventarisasi makanan tradisional dalam menunjang pengembangan wisata kuliner. *JPPI (Jurnal Penelitian Pendidikan Indonesia)*, 9(1), 81–86. <https://doi.org/10.29210/020221974Contents>
- Kertopati, B. W., Yuspi, L., & Fajar, W. N. (2025). JURNAL LOCUS : Penelitian & Pengabdian Nilai-Nilai Kearifan Lokal Kuliner (Studi Fenomenologi Transformasi Makanan Tradisional Gembus , Makna , serta Eksistensinya sebagai Sumber Ekonomi Masyarakat). *JURNAL LOCUS: Penelitian & Pengabdian*, 4(8), 7723–7738. <https://doi.org/10.58344/locus.v4i8.4142>
- Kolanowski, W., Nair, S. S., Varghese, A., & Trz, M. (2025). Post-Certification Quality Analysis of Traditional Indian Fried Snacks. *Applied Sciences*, 1–16. <https://doi.org/10.3390/app15137404>
- Kumar, R. (2023). Transfer Learning in Computer Vision: Technique and applications. *Tuijin Jishu/Journal of Propulsion Technology*, 44, 81–83. <https://doi.org/10.52783/tjjpt.v44.i1.2210>
- Linkon, A. H. M., Labib, M. M., Hasan, T., Hossain, M., & Jannat, M.-E.-. (2021). Deep learning in prostate cancer diagnosis and Gleason grading in histopathology images: An extensive study. *Informatics in Medicine Unlocked*, 24, 100582. <https://doi.org/10.1016/j.imu.2021.100582>
- Liu, D., Zuo, E., Wang, D., He, L., Dong, L., & Lu, X. (2025). Deep Learning in Food Image Recognition : A Comprehensive Review. *Applied Sciences*, 1–18. <https://doi.org/10.3390/app15147626>
- Liu, L., & Xiao, Z. (2025). Feature-Enhanced TResNet for Fine-Grained Food Image Classification. *ArXiv*. <https://doi.org/10.48550/arXiv.2507.12828>
- Mufidatuzzainiya, A., & Faisal, M. (2025). Penggunaan Teknik Transfer Learning pada Metode CNN untuk Pengenalan Tanaman Bunga. *JISKA (Jurnal Informatika Sunan Kalijaga)*, 10(2), 195–206. <https://doi.org/10.14421/jiska.2025.10.2.195-206>
- Muriagista, A., & Kurniadi, D. (2025). Implementasi Arsitektur Resnet50 Pada Klasifikasi Motif Batik Indonesia



- Menggunakan Metode Convolutional Neural Network (CNN). *Jurnal Rekayasa Sistem Informasi Dan Teknologi*, 2(4). <https://doi.org/10.70248/jrsit.v2i4.2375>
- Najla, A., Adibah, N., Fitria, R., Yan, A., Husna, N., Salsabila, Z. N., & Aulia, S. (2025). Nasionalisme Melalui Pelestarian Makanan Tradisional Indonesia. *Jurnal Pendidikan Non Formal*, 3(2), 1–8. <https://doi.org/10.47134/jpn.v3i2.2287>
- Nurhidayat, R., & Kania Evita Dewi. (2023). Penerapan Algoritma K-Nearest Neighbor Dan Fitur Ekstraksi N-Gram Dalam Analisis Sentimen Berbasis Aspek. *KOMPUTA : Jurnal Ilmiah Komputer Dan Informatika*, 12(1), 91–100. <https://doi.org/10.34010/komputa.v12i1.9458>
- Plested, J., Phiri, M., & Gedeon, T. (2025). Deep transfer learning for image classification : a survey. *ArXiv*. <https://doi.org/10.48550/arXiv.2205.09904>
- Pratama, A. Y., & Ghazi, W. (2025). Prediksi Potensi Kinerja Calon Karyawan Customer Service Call Center Menggunakan Model Machine Learning Berbasis Data Rekrutmen. *Building of Informatics, Technology and Science (BITS)*, 7(1), 201–212. <https://doi.org/10.47065/bits.v7i1.7285>
- Pratama, R. A., Octariadi, B. C., & Alkadri, S. P. A. (2025). Penerapan Learning Vector Quantization Dalam Pengolahan Citra Digital Untuk Deteksi Penyakit Kulit. *Jurnal Computer Science and Information Technology (CoSciTech)*, 6(2), 148–157. <https://doi.org/10.37859/coscitech.v6i2.9270>
- Puspitasari, A., Sava, D., & Roliawati, D. (2025). Indonesian Journal on Data Science Penerapan ResNet-50 CNN untuk Optimalisasi Klasifikasi pada Data Fashion. *Indonesian Journal on Data Science*, 3(1), 1–12. <https://doi.org/10.55382/jurnalpustakadata.v5i2.1117>
- Rahayu, N. L. W., Gunantara, N., & Sudarma, M. (2024). Klasifikasi Jajanan Khas Bali Untuk Preservasi Pengetahuan Kuliner Lokal. *SINTECH JOURNAL*, 7(1), 1–14. <https://doi.org/10.31598>
- Reza, H., Biswas, N., Shaikh, S., Rabbani, G., & Chowdhury, M. E. H. (2026). A comprehensive review of convolutional neural networks : foundations , enhancements and applications. *Neural Computing and Applications*. <https://doi.org/10.1007/s00521-025-11827-w>
- Romdona, H., Fitri, A., Qolyubi, A. H., & Azmul, M. (2026). Analisis Adaptasi Pedagang Kue Tradisional Terhadap Produk Makanan Modern Di Pasar Mingguan Manglayang. 1837–1845. <https://doi.org/10.61104/alz.v4i3.5531>
- Satato, Y. R., Prabowo, B. A., & Pramudya, A. L. (2025). Pola Konsumsi dan Persepsi Generasi Z terhadap Ketahanan Pangan : Pengembangan Consumer-Centric Framework. 6(3), 997–1004. <https://doi.org/10.47065/jbe.v6i3.8352>
- Tahir, G. A., & Loo, C. K. (2021). A Comprehensive Survey of Image-Based Food Recognition and Volume Estimation Methods for Dietary Assessment. *Healthcare (Basel, Switzerland)*, 9(12). <https://doi.org/10.3390/healthcare9121676>
- Thiodorus, G., Prasetya, A., Afrizal, L., & Yudistira, N. (2021). Klasifikasi citra makanan / nonmakanan menggunakan metode Transfer Learning dengan model Residual Network Classification of food / non-food images using Transfer Learning method with Residual Network model. *Teknologi: Jurnal Ilmiah Sistem Informasi*, 11(2), 74–83. <http://doi.org/10.26594/teknologi.v11i2.2402>
- Widjanarko, W., Lusiana, Y., Marhaeni, D. P., & Widodo, B. (2023). *The Utilization of Social Media as Traditional Culinary Documentation in Strengthening Local Tourism : A Study on an Instagram Account of @ Dinporabudpar _ banyumas*. Atlantis Press SARL. <https://doi.org/10.2991/978-2-38476-028-2>
- Wijaya, N. P., & Christioko, B. V. (2025). Implementasi Arsitektur MobileNetV2 dengan Metode Transfer Learning untuk Identifikasi Objek Wisata Religi. *Building of Informatics, Technology and Science (BITS)*, 7(3). <https://doi.org/10.47065/bits.v7i3.8447>
- Witjaksono, J., Pusadan, M. Y., Anshori, Y., & Ardiansyah, R. (2025). Klasifikasi jenis batik bomba menggunakan convolutional neural network dengan arsitektur efficient-net b2 (batik bomba sulawesi tengah). *JIPi (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, 10(3), 2134–2147. <https://doi.org/10.29100/jipi.v10i3.6191>