BAB II
TINJAUAN LITERATUR

2.1 Penelitian Terkait

Penelitian mengenai deteksi citra palsu maupun citra yang dihasilkan oleh teknologi Artificial Intelligence telah banyak dilakukan dalam beberapa tahun terakhir. Salah satu metode yang paling sering digunakan adalah Convolutional Neural Network (CNN) karena kemampuannya dalam mengenali pola dan karakteristik visual pada citra. Dalam penelitian yang berjudul *Deteksi Deepfake dalam Citra Menggunakan Convolutional Neural Network (CNN)* mengembangkan model CNN untuk membedakan citra asli dan citra deepfake. Hasil penelitian menunjukkan bahwa CNN mampu mengidentifikasi pola-pola tertentu yang muncul pada citra hasil manipulasi digital sehingga dapat dimanfaatkan sebagai metode yang efektif dalam proses deteksi keaslian gambar secara otomatis (Gulo dkk., 2025).

Penelitian yang berjudul *Sistem Deteksi Gambar Deepfake Menggunakan CNN DenseNet-121 dengan Watermarking Least Significant Bit (LSB)*. Pada penelitian tersebut, arsitektur DenseNet-121 digunakan untuk mengklasifikasikan gambar asli dan gambar deepfake. Setelah melalui tahapan preprocessing dan augmentasi data, model yang dibangun mampu menunjukkan performa yang baik dalam mengenali perbedaan antara gambar asli dan gambar hasil manipulasi AI. Hasil tersebut memperlihatkan bahwa pendekatan CNN masih menjadi salah satu metode yang efektif untuk mendeteksi karakteristik visual yang terdapat pada citra deepfake (Bagus, 2025).

Penelitian yang berjudul *Penerapan Convolutional Neural Network dan Capsule Networks dalam Mendeteksi Deepfake* membahas perbandingan beberapa arsitektur CNN, yaitu ResNet50, VGG19, dan Xception yang dikombinasikan dengan Capsule Networks untuk mendeteksi citra deepfake. Penelitian tersebut menggunakan dataset yang terdiri dari 6.000 citra asli dan 4.000 citra deepfake yang kemudian dilatih dan diuji untuk mengetahui kemampuan model dalam membedakan kedua jenis citra tersebut. Hasil penelitian menunjukkan bahwa kombinasi CNN dan Capsule Networks mampu mendeteksi citra deepfake dengan performa yang baik serta meningkatkan kemampuan model dalam mengenali karakteristik visual yang terdapat pada gambar hasil manipulasi digital. Temuan tersebut menunjukkan bahwa metode berbasis CNN masih menjadi pendekatan yang efektif dalam mendeteksi citra deepfake dan dapat dikembangkan lebih lanjut untuk mendeteksi gambar hasil generatif Artificial Intelligence (Donni Suharyanto dkk., 2024).

Penelitian serupa berjudul *Identifikasi Citra untuk Membedakan Uang Asli dan Palsu Menggunakan Algoritma Convolutional Neural Network (CNN)*. Penelitian tersebut memanfaatkan CNN untuk mengenali perbedaan karakteristik visual antara uang asli dan uang palsu. Hasil yang diperoleh menunjukkan bahwa CNN mampu melakukan proses klasifikasi dengan tingkat akurasi yang baik. Meskipun objek yang digunakan berbeda dengan penelitian ini, hasil tersebut menunjukkan bahwa CNN memiliki kemampuan yang kuat dalam membedakan citra asli dan citra palsu berdasarkan fitur-fitur visual yang terkandung di dalamnya (Harsani Prihastuti, Maulana Muhammad, 2024).

Selain CNN, beberapa penelitian juga memanfaatkan arsitektur MobileNetV2 karena memiliki ukuran model yang lebih ringan namun tetap mampu menghasilkan performa klasifikasi yang baik. Salah satunya penelitian yang berjudul *Penerapan Transfer Learning MobileNetV2 pada Klasifikasi Citra Jenis Buah-Buahan*. Penelitian tersebut menerapkan metode transfer learning menggunakan arsitektur MobileNetV2 untuk mengklasifikasikan berbagai jenis buah berdasarkan citra digital. Hasil penelitian menunjukkan bahwa MobileNetV2 mampu memberikan performa klasifikasi yang baik berdasarkan nilai accuracy, precision, recall, dan F1-score yang diperoleh selama pengujian. Selain menghasilkan tingkat akurasi yang tinggi, penggunaan MobileNetV2 juga mampu mengurangi kebutuhan komputasi dibandingkan proses pelatihan model dari awal (from scratch). Temuan tersebut menunjukkan bahwa MobileNetV2 merupakan arsitektur yang efisien dan efektif untuk diterapkan pada berbagai permasalahan klasifikasi citra, termasuk dalam proses identifikasi gambar manusia hasil generasi Artificial Intelligence (Muslikh, 2025).

Tabel 2.1 Tinjauan Pustaka

Peneliti	Metode	Jenis Citra	Fokus Penelitian	kelebihan	keterbatasan
Steven Adventino Gulo dkk. (2025)	Convolutional Neural Network (CNN)	Citra Deepfake	Mendeteksi perbedaan antara citra asli dan citra deepfake	Akurat mengenali pola manipulasi	Hanya pada citra deepfake

Fatwa Akbar Jiwani dan Bagus S.W. Poetro (2025)	CNN DenseNet-121 + Watermarking Least Significant Bit (LSB)	Citra Deepfake	Klasifikasi gambar asli dan gambar deepfake	Akurasi klasifikasi tinggi	Model cukup kompleks
Donni Suharyanto dkk. (2024)	CNN (ResNet50, VGG19, Xception) + Capsule Networks	Citra Deepfake	Perbandingan beberapa arsitektur CNN dalam mendeteksi deepfake	Performa deteksi meningkat	Membutuhkan komputasi tinggi
Harsani Prihastuti dan Maulana Muhammad (2024)	Convolutional Neural Network (CNN)	Citra Uang Asli dan Palsu	Identifikasi perbedaan antara uang asli dan uang palsu	Klasifikasi akurat	Objek bukan citra manusia
Muslikh (2025)	Transfer Learning	Citra Buah	Klasifikasi jenis buah menggunakan	Ringan dan efisien	Belum diterapkan pada citra AI

	MobileNet V2		MobileNet V2		
Penelitian yang Diusulkan	CNN dengan Arsitektur MobileNetV2	Citra Manusia (Asli dan Hasil Gemini AI)	Deteksi gambar manusia asli dan gambar hasil generatif Gemini A	Ringan, efisien, dan akurat	Terbatas pada citra Gemini AI

Berdasarkan hasil kajian terhadap penelitian-penelitian terdahulu, dapat diketahui bahwa metode Convolutional Neural Network (CNN) dan berbagai arsitektur Deep Learning telah menunjukkan kemampuan yang baik dalam mendeteksi citra palsu, citra deepfake, maupun citra hasil manipulasi digital. Selain itu, arsitektur MobileNetV2 juga terbukti mampu memberikan performa klasifikasi yang baik dengan kebutuhan komputasi yang lebih ringan dibandingkan beberapa model Deep Learning lainnya. Meskipun demikian, sebagian besar penelitian yang telah dilakukan masih berfokus pada deteksi deepfake atau citra hasil generatif AI secara umum, sedangkan penelitian yang secara khusus membahas deteksi gambar manusia hasil generasi Gemini AI masih relatif terbatas. Oleh karena itu, penelitian ini dilakukan untuk membangun sistem yang mampu mengklasifikasikan gambar manusia asli dan gambar manusia hasil generasi Gemini AI menggunakan metode CNN dengan arsitektur MobileNetV2. Penelitian ini diharapkan dapat menjadi

alternatif solusi yang efektif dan efisien dalam mendukung proses verifikasi keaslian konten visual serta memberikan kontribusi pada pengembangan teknologi forensik digital di era perkembangan Artificial Intelligence yang semakin pesat.

2.2 Kecerdasan Buatan (Artificial Intelligence)

Kecerdasan buatan atau artificial intelligence (AI) merupakan cabang ilmu komputer yang berfokus pada pengembangan sistem yang mampu meniru kemampuan intelektual manusia, seperti proses belajar, penalaran, pengenalan pola, serta pengambilan keputusan. Melalui pemanfaatan algoritma dan analisis data, teknologi ini memungkinkan komputer menjalankan berbagai tugas secara otomatis sehingga dapat meningkatkan efektivitas dan ketepatan dalam berbagai bidang (Tresnawati *et al.*, 2022).

Perkembangan Artificial Intelligence semakin pesat seiring dengan meningkatnya kemampuan komputasi, tersedianya data dalam jumlah besar, serta kemajuan algoritma pembelajaran. Faktor-faktor tersebut mendorong penerapan AI di berbagai sektor, seperti pendidikan, kesehatan, industri, keamanan, dan *computer vision*. Pemanfaatan AI mampu meningkatkan efisiensi kerja, mempercepat proses pengolahan dan analisis data, serta menghasilkan pengambilan keputusan yang lebih tepat dan akurat (Natasya *et al.*, 2023).

Perkembangan teknologi telah mendorong Artificial Intelligence menjadi semakin maju dengan lahirnya berbagai cabang, seperti *Machine Learning*, *Deep Learning*, dan *Generative Artificial Intelligence (Generative AI)*. Kemajuan tersebut memungkinkan sistem AI tidak hanya melakukan proses klasifikasi dan prediksi, tetapi juga menghasilkan berbagai jenis konten baru, seperti teks, gambar,

audio, maupun video. Meskipun memberikan banyak manfaat dalam berbagai aspek kehidupan, pemanfaatan AI juga menghadirkan tantangan baru, terutama dalam menjaga keaslian, keandalan, dan keamanan informasi digital (Zahara *et al.*, 2023)

2.3 Generative Artificial Intelligence (Generative AI)

Generative Artificial Intelligence (*Generative AI*) merupakan salah satu perkembangan dari teknologi kecerdasan buatan yang mampu menghasilkan berbagai jenis konten baru berdasarkan pola yang dipelajari dari data pelatihan. Berbeda dengan sistem AI konvensional yang berfokus pada proses klasifikasi, prediksi, atau pengenalan pola, *Generative AI* dapat menciptakan konten digital berupa teks, gambar, audio, video, hingga kode program dengan karakteristik yang menyerupai data asli. Kemampuan tersebut didukung oleh penerapan algoritma *Deep Learning*, peningkatan kemampuan komputasi, serta ketersediaan data dalam jumlah besar sehingga menghasilkan konten dengan tingkat realisme yang tinggi. Perkembangan ini mendorong pemanfaatan *Generative AI* pada berbagai bidang, seperti pendidikan, industri kreatif, pengembangan perangkat lunak, desain visual, pemasaran digital, dan produksi media karena mampu meningkatkan efisiensi sekaligus mendukung proses kreatif dalam menghasilkan konten digital (Shiddiq, 2024)

Meskipun menawarkan berbagai manfaat, perkembangan *Generative AI* juga menimbulkan tantangan dalam menjaga keaslian informasi digital. Kemampuan sistem dalam menghasilkan gambar yang sangat realistis berpotensi dimanfaatkan untuk membuat konten palsu, memanipulasi identitas, menyebarkan

informasi yang menyesatkan, hingga melakukan berbagai bentuk penyalahgunaan di ruang digital. Kondisi tersebut menunjukkan pentingnya pengembangan metode yang mampu membedakan konten asli dengan konten hasil generasi AI guna meningkatkan kepercayaan terhadap informasi digital (Nahla, 2024).

2.4 Gemini Artificial Intelligence

Gemini Artificial Intelligence (Gemini AI) merupakan model kecerdasan buatan multimodal yang dikembangkan oleh Google DeepMind dengan kemampuan untuk memahami, mengolah, serta menghasilkan berbagai jenis data dalam satu sistem terpadu, meliputi teks, gambar, audio, video, dan kode program. Model ini dibangun menggunakan teknologi *Deep Learning* dengan arsitektur *Transformer* sehingga mampu memproses hubungan antar berbagai jenis data dan menghasilkan keluaran yang lebih relevan serta kontekstual dibandingkan model AI generatif sebelumnya. Kemampuan tersebut menjadikan Gemini AI sebagai salah satu model kecerdasan buatan yang dapat dimanfaatkan dalam berbagai kebutuhan, mulai dari analisis informasi hingga pembuatan konten digital (Nazarius *et al.*, 2024).

Salah satu fitur utama Gemini AI adalah kemampuannya menghasilkan gambar digital berdasarkan deskripsi atau *prompt* yang diberikan oleh pengguna. Kemampuan ini banyak dimanfaatkan pada berbagai bidang, seperti pendidikan, industri kreatif, pemasaran digital, desain visual, dan pembuatan media, karena mampu menghasilkan gambar dengan kualitas yang realistis dalam waktu yang relatif singkat. Walaupun terkadang masih terdapat karakteristik tertentu, seperti tekstur kulit yang terlalu halus, pencahayaan kurang alami, atau ketidaksesuaian

pada beberapa bagian gambar, kualitas hasil generatif AI terus berkembang sehingga semakin sulit dibedakan secara visual. Di sisi lain, perkembangan tersebut juga meningkatkan potensi penyalahgunaan, seperti pembuatan gambar hasil AI (*AI-generated image*), manipulasi visual, serta penyebaran konten digital yang sulit dibedakan dari gambar asli (Afiani dan Mahenra, 2025).

Pada penelitian ini, Gemini AI digunakan sebagai sumber untuk menghasilkan dataset berupa gambar manusia hasil generasi AI yang selanjutnya dibandingkan dengan gambar manusia asli. Kedua kategori gambar tersebut kemudian diproses menggunakan metode Convolutional Neural Network (CNN) dengan arsitektur MobileNetV2 untuk mempelajari karakteristik visual masing-masing kategori sehingga model mampu mengklasifikasikan gambar hasil generasi AI dan gambar asli secara lebih akurat.

2.5 Citra Digital

Citra digital merupakan representasi visual dari suatu objek atau peristiwa yang disimpan dalam bentuk data digital sehingga dapat diproses oleh komputer. Berbeda dengan gambar analog, gambar digital tersusun atas kumpulan piksel (*pixel*) yang membentuk matriks dua dimensi. Setiap piksel menyimpan informasi mengenai warna dan tingkat intensitas cahaya sehingga menghasilkan citra yang dapat dianalisis menggunakan berbagai teknik pengolahan citra digital. Citra digital adalah representasi visual suatu objek yang terdiri atas kumpulan piksel dengan nilai tertentu yang menggambarkan warna maupun tingkat kecerahan (Dijaya Rohman, 2023).

Pengolahan citra digital merupakan proses pengolahan gambar untuk meningkatkan kualitas visual maupun memperoleh informasi tertentu dari suatu objek (Ramadhanu dan Syahputra, 2022). Proses tersebut meliputi berbagai tahapan, seperti perubahan ukuran citra, normalisasi, pengurangan *noise*, segmentasi, hingga transformasi citra sebagai langkah awal sebelum dilakukan analisis lebih lanjut (Silvia, 2022).

Dalam penelitian ini, gambar digital digunakan sebagai dataset utama yang terdiri atas gambar manusia asli (*real image*) dan gambar manusia hasil Gemini Artificial Intelligence (*AI-generated image*). Sebelum dilakukan proses klasifikasi, seluruh gambar melalui tahap *preprocessing* berupa *resize*, normalisasi, dan *data augmentation*. Tahapan ini bertujuan menghasilkan dataset yang seragam sehingga model Convolutional Neural Network (CNN) dengan arsitektur MobileNetV2 dapat mengenali karakteristik visual setiap kategori gambar secara lebih optimal.

2.6 Pengolahan Citra Digital *(Image Processing)*

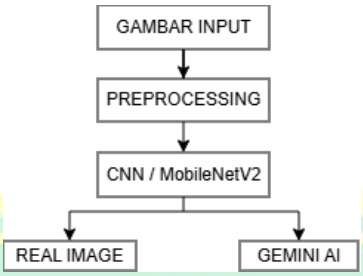
Pengolahan citra digital (*image processing*) merupakan salah satu cabang ilmu komputer yang membahas berbagai teknik untuk memperoleh, memperbaiki, memodifikasi, menganalisis, dan mengekstraksi informasi dari citra digital dengan bantuan komputer. Citra digital tersusun atas kumpulan piksel (*pixel*) yang membentuk matriks dua dimensi, di mana setiap piksel memiliki nilai intensitas yang merepresentasikan warna atau tingkat keabuan. Melalui proses pengolahan tersebut, informasi visual yang terdapat pada sebuah gambar dapat diolah sehingga lebih mudah dianalisis maupun dimanfaatkan sebagai masukan pada berbagai sistem berbasis computer (Ashillah *et al.*, 2025).

Secara umum, citra digital dibedakan menjadi tiga jenis, yaitu RGB (*color image*), grayscale, dan binary. Citra RGB merepresentasikan warna melalui kombinasi tiga komponen utama, yaitu merah (*Red*), hijau (*Green*), dan biru (*Blue*). Sementara itu, citra grayscale hanya menyimpan informasi tingkat keabuan tanpa unsur warna, sedangkan citra binary terdiri atas dua nilai piksel, yaitu hitam dan putih. Penggunaan jenis citra tersebut disesuaikan dengan kebutuhan aplikasi dan teknik pengolahan yang diterapkan. Pengolahan citra digital bertujuan untuk meningkatkan kualitas gambar, mengurangi gangguan (*noise*), melakukan segmentasi objek, mengekstraksi fitur, serta mendukung proses klasifikasi citra. Salah satu teknik yang banyak digunakan adalah normalisasi RGB, yaitu metode yang memanfaatkan proporsi nilai komponen merah, hijau, dan biru untuk memisahkan objek dari latar belakang sehingga proses identifikasi objek dapat dilakukan dengan lebih efektif (Fadjeri *et al.*, 2022).

2.7 Image Classification

Image Classification merupakan salah satu tugas dalam bidang computer vision yang bertujuan mengelompokkan sebuah citra ke dalam satu atau beberapa kelas berdasarkan karakteristik visual yang dipelajari oleh model. Ketika hanya terdapat dua kelas keluaran, proses ini disebut binary image classification. Penelitian ini pada dasarnya merupakan binary image classification, karena model diharapkan mampu menentukan apakah sebuah gambar manusia termasuk kelas Real (gambar asli) atau kelas AI-Generated (gambar hasil generasi Gemini AI), Secara umum, alur proses image classification dimulai dari citra masukan yang melalui tahap preprocessing, kemudian diproses oleh model CNN dengan arsitektur

MobileNetV2 untuk menghasilkan prediksi kelas berupa gambar Real atau AI-Generated. sebagaimana pendekatan serupa yang diterapkan pada klasifikasi wajah asli dan wajah hasil deepfake (Mu *et al.*, 2024).



Gambar 2. 1 Alur Proses Image Classification

2.8 Machine Learning

Machine Learning merupakan salah satu cabang dari Artificial Intelligence (AI) yang berfokus pada pengembangan sistem komputer agar mampu mempelajari pola dari data dan meningkatkan kinerjanya secara otomatis tanpa harus diprogram secara rinci untuk setiap tugas. Melalui proses pembelajaran dari data yang tersedia, model machine learning dapat mengenali pola, melakukan klasifikasi, membuat prediksi, serta mendukung proses pengambilan keputusan. Kemampuan model dalam menghasilkan prediksi yang akurat sangat dipengaruhi oleh kualitas dan jumlah data yang digunakan selama proses pelatihan (Faiza dan Andriani, 2022).

Dalam penerapannya, machine learning telah digunakan pada berbagai bidang, seperti pengolahan citra digital, pengenalan wajah, deteksi objek, sistem rekomendasi, analisis kesehatan, keamanan siber, hingga pengenalan suara. Pada bidang pengolahan citra, machine learning dimanfaatkan untuk mengenali karakteristik visual suatu objek berdasarkan fitur-fitur yang diperoleh dari gambar digital. Kemampuan tersebut menjadikan machine learning sebagai salah satu

teknologi utama dalam pengembangan sistem klasifikasi citra secara otomatis (Ilmawan *et al.*, 2025).

Dalam membangun model machine learning, dataset umumnya dibagi menjadi tiga bagian, yaitu:

1. *Training dataset*, merupakan kumpulan atau himpunan data yang digunakan untuk membangun model machine learning.
2. *Development* atau *validation dataset*, merupakan himpunan data untuk mengoptimasi saat melatih kinerja model machine learning.
3. *Testing dataset*, merupakan himpunan data untuk menguji hasil akhir model setelah selesai proses latihan dengan tidak memperkaitkan model dan manusia dalam proses dilaksanakan.

Pada penelitian ini, dataset dibagi menjadi 90% data training, 7% data validation, dan 3% data testing. Pembagian tersebut dilakukan untuk memberikan porsi data yang lebih besar pada proses pembelajaran model, sementara data validation digunakan untuk memantau proses pelatihan dan data testing dimanfaatkan untuk mengevaluasi performa akhir model dalam mendeteksi gambar asli dan gambar hasil Gemini AI.

Machine learning memiliki empat jenis klasifikasi berdasarkan tipe dari pengolahan data (Alia, 2023). Jenis klasifikasi machine learning, yaitu:

1. Supervised Learning

   Supervised Learning atau jenis pembelajaran terarah yang mesin diberikan dataset bersamaan dengan hasil jawaban yang benar sesuai pertanyaan yang ada pada datapoint. Dalam kasus supervised learning, model pembelajaran

memiliki kondisi instance yang merupakan kaitan antara input dengan desired output yang merupakan contoh hasil dari kemungkinan input yang didapat. Supervised learning umum digunakan untuk kondisi regression atau classification. Regression adalah kondisi mesin mendapatkan data yang bernilai real dan numerical untuk dapat memprediksi nilai output bersifat kontinu seperti harga saham dikemudian hari. Classification adalah kondisi mesin diberi data berlabel atau terkategori untuk dapat mengambil keputusan berdasarkan label atau kategori tersebut seperti mengklasifikasi buah

2. Unsupervised learning

   Unsupervised learning adalah kebalikan dari supervised learning. Mesin tidak diberi data secara berlabel namun secara otomatis dapat dibagi berdasarkan kemiripan atau tingkat similaritas dan struktur lain dari data tersebut. Tujuan digunakan unsupervised learning untuk dapat memodelkan dan mengelompokkan informasi yang tidak disotir berdasarkan persamaan atau pola tanpa melakukan pelatihan data sebelumnya. Unsupervised learning memiliki tiga jenis kategori, yaitu clustering, asosiation, dan dimensionality reduction.

3. Semi-supervised Learning

   Memiliki kesamaan dengan supervised learning dengan perbedaan terletak pada penggunaan data berlabel namun tidak untuk melatih algoritma

4. Reinforcement Learning

   Jenis Machine learning yang menggunakan data untuk kondisi label if dan else dalam mengambil keputusan terbaik dari hasil trail dan error berulang kali dengan kemungkinan mesin berinteraksi secara dinamis.

Pada penelitian ini, metode yang digunakan termasuk dalam kategori supervised learning, karena proses pelatihan dilakukan menggunakan dataset yang telah diberi label, yaitu gambar manusia asli (*non-Gemini*) dan gambar hasil generatif Gemini AI. Dengan mempelajari karakteristik dari kedua kelas tersebut, model diharapkan mampu mengklasifikasikan gambar baru ke dalam kategori yang sesuai.

2.9 Deep Learning

Deep Learning merupakan salah satu metode dalam *Machine Learning* yang memanfaatkan jaringan saraf tiruan (*Artificial Neural Network*) dengan banyak lapisan tersembunyi (*hidden layer*) untuk mempelajari pola dan karakteristik data secara otomatis. Berbeda dengan metode pembelajaran mesin konvensional yang masih bergantung pada proses ekstraksi fitur secara manual, *Deep Learning* mampu mempelajari representasi fitur langsung dari data masukan selama proses pelatihan (*training*). Kemampuan tersebut memungkinkan model mengidentifikasi hubungan yang kompleks antar data sehingga mampu memberikan hasil yang lebih akurat, terutama dalam pengolahan data tidak terstruktur seperti gambar, suara, maupun teks (Mienye, 2024).

Prinsip utama *Deep Learning* terletak pada proses pembelajaran bertingkat (*hierarchical learning*), di mana setiap lapisan jaringan memiliki fungsi untuk

mempelajari karakteristik data secara berurutan. Lapisan awal umumnya mengenali fitur dasar, seperti garis, warna, dan tepi objek. Informasi tersebut kemudian diproses pada lapisan berikutnya untuk membentuk fitur yang lebih kompleks, seperti tekstur, bentuk, hingga karakteristik suatu objek secara menyeluruh. Proses pembelajaran yang dilakukan secara bertahap tersebut membuat *Deep Learning* mampu mengenali pola visual dengan lebih efektif dibandingkan metode pembelajaran mesin tradisional (Khoei, 2023).

Perkembangan teknologi *Deep Learning* telah memperluas pemanfaatannya di berbagai bidang, khususnya pada *computer vision*. Hal ini disebabkan oleh kemampuannya dalam mempelajari pola dan karakteristik data secara otomatis melalui jaringan saraf tiruan (*Artificial Neural Network*). Kemampuan tersebut membuat *Deep Learning* banyak diterapkan dalam berbagai penelitian, seperti klasifikasi citra, deteksi objek, pengenalan pola, serta pengenalan wajah, karena mampu memberikan tingkat akurasi yang lebih baik dibandingkan metode konvensional. Selain itu, meningkatnya kebutuhan akan sistem klasifikasi citra yang lebih efektif dan efisien turut mendorong perkembangan serta penerapan *Deep Learning* pada berbagai aplikasi berbasis kecerdasan buatan (Naufal dan Kusuma, 2023).

Pada bidang *computer vision*, *Deep Learning* mampu mempelajari informasi visual yang terdapat pada setiap piksel gambar untuk mengenali karakteristik suatu objek secara otomatis. Model akan mengidentifikasi berbagai fitur, seperti bentuk, tekstur, warna, dan pola, tanpa memerlukan proses ekstraksi fitur secara manual. Kemampuan tersebut menjadikan *Deep Learning* sebagai salah

satu pendekatan yang banyak digunakan dalam penelitian klasifikasi citra karena mampu menghasilkan performa yang tinggi pada berbagai jenis dataset (Harsani Prihastuti, Maulana Muhammad, 2024).

Salah satu arsitektur yang paling banyak digunakan dalam implementasi *Deep Learning* untuk klasifikasi citra adalah Convolutional Neural Network (CNN). Arsitektur ini dirancang untuk mengekstraksi fitur secara otomatis melalui lapisan konvolusi (*convolution layer*) dan *pooling*, sehingga mampu mengenali pola visual, tekstur, maupun bentuk objek dengan tingkat akurasi yang baik. Seiring perkembangan teknologi, CNN menjadi dasar bagi berbagai arsitektur modern, salah satunya MobileNetV2, yang dikembangkan untuk menghasilkan model dengan jumlah parameter lebih sedikit dan kebutuhan komputasi yang lebih ringan tanpa mengurangi kemampuan dalam melakukan klasifikasi citra (Said, 2024).

Gambar 2.2 Kaitan Kecerdasan Buatan, *Machine Learning*, dan *Deep Learning* (Alia, 2023)

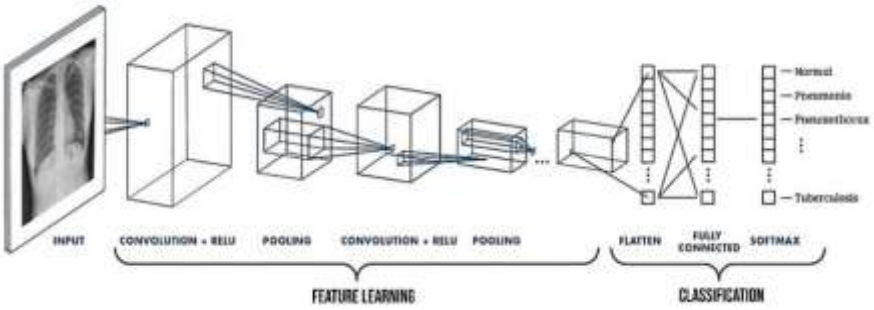
2.10 Convolutional Neural Network

Convolutional Neural Network (CNN) merupakan salah satu arsitektur *Deep Learning* yang banyak digunakan dalam pengolahan citra digital. CNN dirancang untuk mengenali dan mempelajari karakteristik visual suatu gambar secara otomatis melalui proses ekstraksi fitur, sehingga tidak lagi bergantung pada teknik ekstraksi fitur secara manual seperti pada metode konvensional. Dengan kemampuan tersebut, CNN dapat mendeteksi berbagai informasi visual, seperti tepi, garis, tekstur, bentuk, serta pola pada suatu objek, yang kemudian dimanfaatkan untuk meningkatkan ketepatan dalam proses klasifikasi citra (Ibnul *et al.*, 2022).

Salah satu keunggulan CNN adalah kemampuannya dalam mempelajari fitur secara bertingkat (*hierarchical feature learning*), sehingga banyak dimanfaatkan pada berbagai aplikasi *computer vision*. Pada lapisan awal, model akan mengenali fitur-fitur sederhana, seperti garis, tepi, dan sudut, kemudian informasi tersebut diteruskan ke lapisan berikutnya untuk membentuk fitur yang lebih kompleks, seperti tekstur, bentuk, hingga karakteristik objek secara keseluruhan. Mekanisme pembelajaran tersebut memungkinkan CNN menghasilkan tingkat akurasi yang tinggi pada berbagai penerapan, antara lain klasifikasi citra, deteksi objek, segmentasi citra, dan pengenalan wajah (Wita dan Liliana, 2022).

2.10.1 Arsitektur Convolutional Neural Network (CNN)

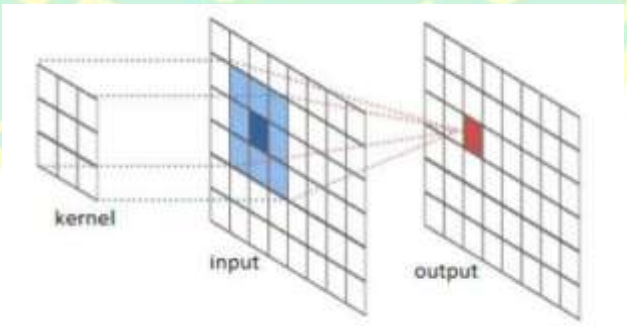
Terdapat tiga lapisan utama dalam proses pengolahan citra, yaitu *Convolution Layer, Pooling Layer* dan *Fully Connected Layer* (Wita dan Liliana, 2022).



Gambar 2.3 Arsitektur CNN (Gunawan dan Setiawan, 2022)

1. *Convolution Layer*

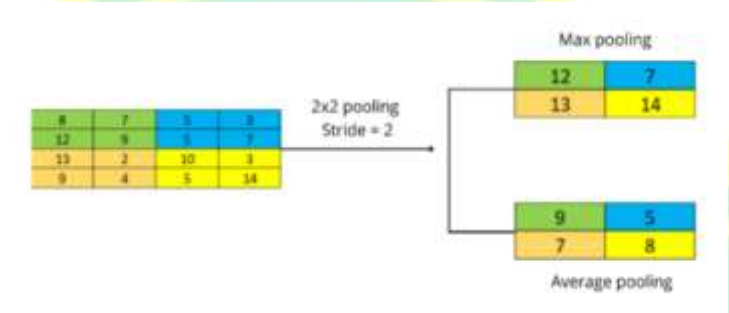
Convolution Layer merupakan lapisan utama pada *Convolutional Neural Network (CNN)* yang berperan dalam melakukan proses ekstraksi fitur dari citra masukan. Lapisan ini terdiri atas sejumlah *filter* atau *kernel* yang digeser secara sistematis pada seluruh bagian gambar untuk mendeteksi berbagai karakteristik visual, seperti garis, tepi, tekstur, bentuk, maupun pola tertentu. Selama proses konvolusi, setiap *kernel* akan mengolah nilai piksel pada area tertentu sehingga menghasilkan keluaran berupa feature map yang berisi informasi penting dari gambar (Ramadhani dan Satria, 2023).



Gambar 2.4 Convolutional layer (Maulana *et al.*, 2023)

2. *Pooling Layer*

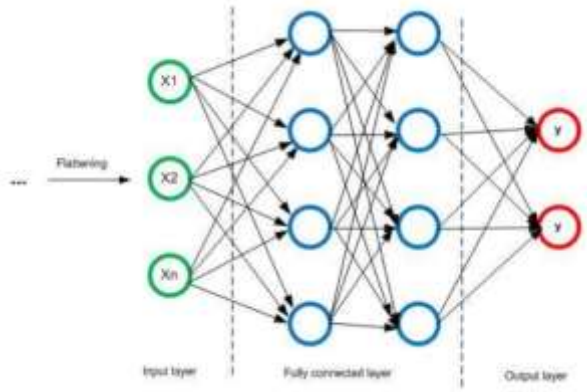
Pooling Layer berfungsi untuk mengurangi dimensi *feature map* hasil konvolusi sehingga proses komputasi menjadi lebih efisien. Terdapat dua jenis *pooling*, yaitu Max Pooling dan Average Pooling. Pada Max Pooling, sistem memilih nilai maksimum dari setiap area *feature map* untuk mempertahankan fitur-fitur penting yang akan digunakan pada proses klasifikasi selanjutnya (Husna *et al.*, 2022).



Gambar 2.5 Contoh Pooling Layer (Syarul Ramadhani, 2023)

3. *Fully Connected Layer*

Fully Connected Layer merupakan lapisan pada CNN yang menghubungkan seluruh neuron dari lapisan sebelumnya untuk melakukan proses klasifikasi. Lapisan ini umumnya ditempatkan pada bagian akhir arsitektur CNN sebagai tahap pengambilan keputusan berdasarkan fitur-fitur yang telah diekstraksi. Pada proses klasifikasi, *Fully Connected Layer* biasanya menggunakan fungsi aktivasi Softmax pada *Output Layer* untuk menghitung probabilitas setiap kelas, sehingga model dapat menentukan hasil klasifikasi dengan memilih kelas yang memiliki nilai probabilitas tertinggi (Duta Pratama, Rendra Gustriansyah, 2024).

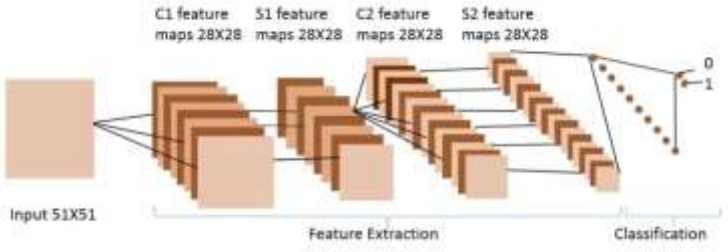


Gambar 2.6 Proses Fully Connected (Herlambang, 2024)

2.10.2 Feature Extraction

Feature extraction merupakan tahapan penting dalam Convolutional Neural Network (CNN) yang berfungsi untuk mengekstraksi berbagai karakteristik visual dari suatu citra secara otomatis. Berbeda dengan metode pengolahan citra konvensional yang mengharuskan ekstraksi fitur dilakukan secara manual, CNN mampu mempelajari dan mengenali fitur-fitur penting secara langsung melalui proses pelatihan. Fitur yang dipelajari meliputi tepi (*edges*), garis, tekstur, bentuk, pola, hingga karakteristik objek yang lebih kompleks (Syakuroh *et al.*, 2025).

Proses feature extraction dilakukan pada Convolution Layer dengan memanfaatkan filter atau *kernel* yang bergerak di seluruh bagian citra. Setiap filter akan menghasilkan feature map yang berisi representasi informasi penting dari gambar. Pada lapisan awal, CNN umumnya mengenali fitur-fitur dasar, seperti garis dan tepi, kemudian pada lapisan berikutnya fitur tersebut dikombinasikan menjadi representasi yang lebih kompleks, seperti tekstur, bentuk objek, dan pola visual. Mekanisme ini dikenal sebagai hierarchical feature learning, yaitu kemampuan CNN dalam mempelajari fitur secara bertingkat dari tingkat sederhana hingga kompleks.



Gambar 2.7 Proses Feature Extraction (Gomede, 2024)

Pada penelitian ini, proses feature extraction diterapkan menggunakan arsitektur MobileNetV2 melalui pendekatan transfer learning. Model memanfaatkan bobot (*weights*) yang telah dilatih sebelumnya pada dataset ImageNet untuk mengekstraksi berbagai karakteristik visual dari gambar manusia, seperti tekstur, warna, bentuk, dan pola objek. Karakteristik tersebut selanjutnya digunakan sebagai dasar dalam proses klasifikasi untuk membedakan gambar manusia asli (*real image*) dan gambar hasil generatif Gemini AI (*AI-generated image*).

2.10.3 Classification Layer

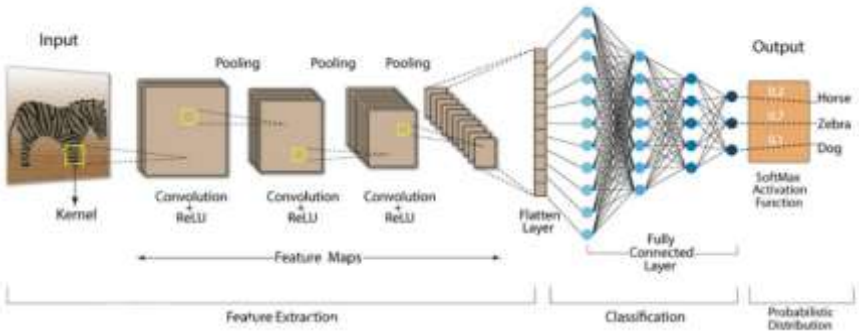
Classification layer merupakan bagian akhir dari arsitektur Convolutional Neural Network (CNN) yang berfungsi untuk mengklasifikasikan citra berdasarkan fitur-fitur yang telah diperoleh melalui proses feature extraction. Informasi yang tersimpan pada feature map kemudian diteruskan ke lapisan klasifikasi untuk menghasilkan prediksi sesuai dengan karakteristik gambar yang telah dipelajari oleh model. Pada CNN, proses klasifikasi umumnya dilakukan melalui Fully Connected Layer, yaitu lapisan yang menghubungkan seluruh neuron dari lapisan sebelumnya untuk mengolah hasil ekstraksi fitur menjadi informasi yang siap digunakan dalam proses pengambilan keputusan. Selanjutnya, Output Layer

menerapkan fungsi aktivasi Softmax untuk menghitung probabilitas setiap kelas, sehingga model dapat menentukan hasil klasifikasi berdasarkan kelas yang memiliki nilai probabilitas tertinggi

2.10.4 Cara Kerja Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) melakukan proses klasifikasi citra melalui serangkaian lapisan (*layer*) yang saling terhubung. Setiap lapisan memiliki peran tertentu, mulai dari mengekstraksi karakteristik visual hingga menghasilkan keputusan klasifikasi. Dengan mekanisme tersebut, CNN dapat mempelajari pola dan fitur pada gambar secara otomatis tanpa memerlukan proses ekstraksi fitur secara manual, sehingga banyak dimanfaatkan dalam berbagai aplikasi *computer vision* (Szeliski, 2022).

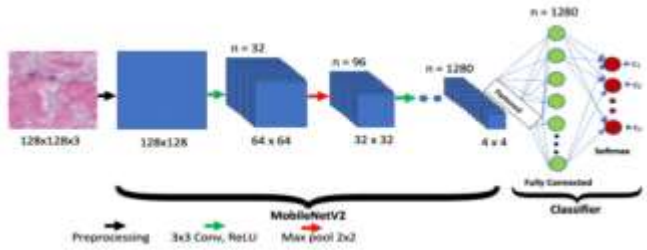
Proses kerja CNN dimulai ketika citra dimasukkan ke Input Layer sebagai data masukan. Selanjutnya, gambar diproses pada Convolution Layer untuk mengekstraksi berbagai fitur penting, seperti garis, tepi, tekstur, dan bentuk objek. Fitur yang dihasilkan kemudian diteruskan ke Pooling Layer untuk mengurangi ukuran data sekaligus mempertahankan informasi yang relevan. Setelah itu, hasil pemrosesan diteruskan ke Fully Connected Layer dan Output Layer untuk menentukan kelas gambar berdasarkan pola yang telah dipelajari selama proses pelatihan (Said, 2024).



Gambar 2.8 Alur Kerja *Convolutional Neural Network* (Munir, 2024)

2.11 Arsitektur MobileNetV2

MobileNetV2 merupakan salah satu arsitektur *Deep Learning* yang dikembangkan oleh Google sebagai pengembangan dari MobileNet generasi sebelumnya. Arsitektur ini dirancang agar memiliki jumlah parameter yang lebih sedikit dan kebutuhan komputasi yang lebih efisien, sehingga dapat diimplementasikan pada perangkat dengan kapasitas sumber daya yang terbatas tanpa mengurangi kinerja model secara signifikan. Untuk mencapai efisiensi tersebut, MobileNetV2 mengadopsi tiga komponen utama, yaitu Depthwise Separable Convolution, Inverted Residual, dan Linear Bottleneck, yang berperan dalam mengoptimalkan proses komputasi sekaligus mempertahankan akurasi klasifikasi (Maulana *et al.*, 2024).

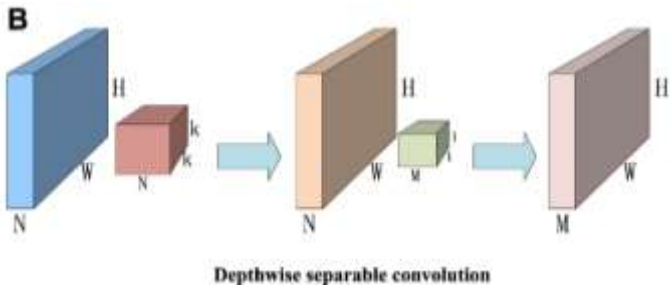


Gambar 2.9 Arsitektur MobileNetV2 (Yong, 2021)

MobileNetV2 dikembangkan dengan mengadopsi tiga komponen utama yang membedakannya dari arsitektur CNN konvensional, yaitu:

2.11.1 Depthwise Separable Convolution

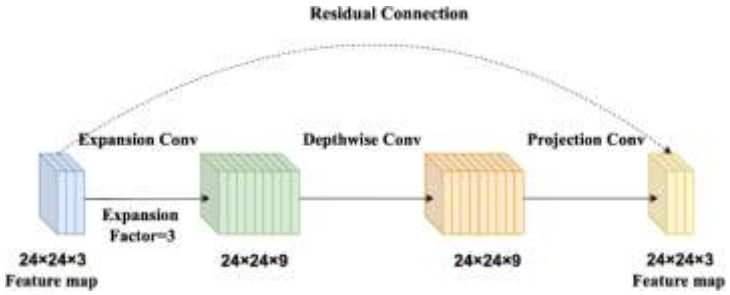
Depthwise Separable Convolution merupakan metode konvolusi yang diterapkan pada arsitektur MobileNetV2 untuk meningkatkan efisiensi komputasi dengan mengurangi jumlah parameter yang digunakan. Teknik ini bekerja melalui dua tahapan, yaitu Depthwise Convolution yang melakukan ekstraksi fitur pada setiap kanal gambar secara terpisah, kemudian dilanjutkan dengan Pointwise Convolution yang menggabungkan hasil ekstraksi menggunakan *kernel* berukuran 1×1. Melalui mekanisme tersebut, MobileNetV2 mampu melakukan proses ekstraksi fitur secara lebih efisien tanpa mengurangi kemampuan model dalam melakukan klasifikasi (Baay *et al.*, 2021).



Gambar 2.10 Proses Depthwise Separable Convolution (Yun et al., 2022)

2.11.2 Inverted Residual

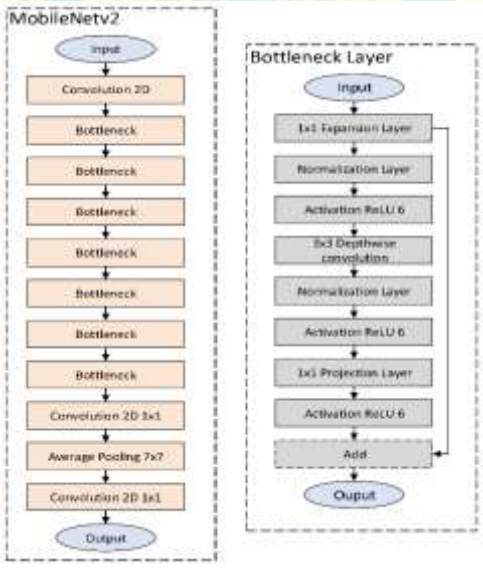
Inverted Residual merupakan salah satu komponen penting dalam arsitektur MobileNetV2 yang dirancang untuk meningkatkan efisiensi proses komputasi sekaligus menjaga informasi penting selama proses ekstraksi fitur. Berbeda dengan *residual block* pada ResNet yang menerapkan *shortcut connection* pada lapisan dengan jumlah kanal yang lebih besar, MobileNetV2 menghubungkan *shortcut* langsung pada bottleneck layer yang memiliki dimensi lebih kecil. Pendekatan ini didasarkan pada konsep bahwa informasi utama dipertahankan pada lapisan *bottleneck*, sedangkan lapisan ekspansi digunakan sebagai media untuk melakukan transformasi nonlinier. Dengan mekanisme tersebut, MobileNetV2 mampu mengoptimalkan penggunaan memori tanpa mengurangi performa model dalam melakukan proses klasifikasi (Sandler *et al.*, n.d.).



Gambar 2.11 Proses Inverted Residual (Xu et al., 2025)

2.11.3 Linear Bottleneck

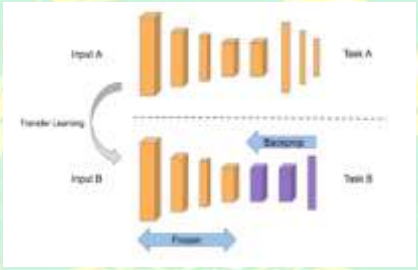
Linear Bottleneck merupakan salah satu komponen pada arsitektur MobileNetV2 yang berfungsi mempertahankan informasi penting selama proses ekstraksi fitur. Berbeda dengan CNN konvensional yang menggunakan fungsi aktivasi nonlinier pada setiap lapisan, MobileNetV2 menerapkan transformasi linear pada lapisan bottleneck untuk mengurangi kehilangan informasi saat jumlah kanal diperkecil. Pendekatan ini membuat model tetap efisien sekaligus mampu mempertahankan performa klasifikasi yang baik (Aji *et al.*, 2025).



Gambar 2.12 Proses Linear Bottleneck (Tragoudaras *et al.*, 2022)

2.12 Transfer Learning

Transfer Learning merupakan salah satu pendekatan dalam *Deep Learning* yang memanfaatkan model yang telah melalui proses pelatihan (*pre-trained model*) pada dataset berskala besar untuk diterapkan pada permasalahan baru yang memiliki karakteristik serupa. Dengan metode ini, proses pelatihan tidak dilakukan dari awal, tetapi menggunakan bobot (*weights*) yang telah dipelajari sebelumnya dan kemudian disesuaikan dengan dataset yang digunakan. Pendekatan tersebut dapat mempercepat proses pelatihan, mengurangi kebutuhan data, serta meningkatkan kinerja model dalam melakukan klasifikasi citra (Rismiyati, 2021).



Gambar 2.13 Proses Transfer Learning (Putra *et al.*, 2023)

Pada penelitian ini, metode transfer learning diterapkan menggunakan MobileNetV2 sebagai *pre-trained model* yang telah dilatih pada dataset ImageNet. Penerapan metode ini melibatkan tahapan feature extraction dan fine-tuning untuk menyesuaikan model dengan dataset penelitian. Dengan memanfaatkan pengetahuan yang telah diperoleh dari proses pelatihan sebelumnya, model diharapkan mampu mengenali karakteristik visual gambar manusia asli maupun gambar hasil Gemini AI secara lebih akurat dan efisien.

Dalam implementasinya, transfer learning melibatkan tiga konsep utama, yaitu:

1. *Pre-trained model* merupakan model *deep learning* yang telah dilatih sebelumnya menggunakan dataset berskala besar, seperti ImageNet, sehingga memiliki kemampuan untuk mengenali berbagai pola dan karakteristik visual pada citra. Model tersebut dapat dimanfaatkan kembali sebagai dasar dalam menyelesaikan tugas klasifikasi yang berbeda tanpa perlu melakukan pelatihan dari awal (*from scratch*). Beberapa contoh *pre-trained model* yang banyak digunakan antara lain MobileNetV2, ResNet50, VGG16, EfficientNet, dan InceptionV3.
2. *Feature extraction* merupakan teknik dalam *transfer learning* yang memanfaatkan pre-trained model untuk mengekstraksi karakteristik atau fitur penting dari suatu citra. Pada teknik ini, sebagian besar bobot (*weights*) model dipertahankan (*freeze*) sehingga tidak diperbarui selama proses pelatihan. Hanya lapisan klasifikasi pada bagian akhir yang dilatih menggunakan dataset baru. Dengan pendekatan ini, proses pelatihan menjadi lebih cepat sekaligus tetap mempertahankan kemampuan model dalam mengenali fitur-fitur visual pada citra.
3. *Fine-tuning* merupakan teknik dalam *transfer learning* yang dilakukan dengan melatih kembali sebagian lapisan pada pre-trained model menggunakan dataset baru. Berbeda dengan feature extraction yang mempertahankan sebagian besar bobot model, pada *fine-tuning* beberapa lapisan dibuka kembali sehingga bobotnya dapat diperbarui. Proses ini memungkinkan model beradaptasi dengan karakteristik dataset baru dan meningkatkan performa klasifikasi.

2.13 Dataset Gambar

Dataset merupakan sekumpulan data yang digunakan sebagai dasar dalam proses pembelajaran model *machine learning*. Kualitas dan kuantitas data yang terdapat dalam dataset memiliki pengaruh yang signifikan terhadap performa model klasifikasi, sehingga pemilihan dataset yang sesuai menjadi salah satu faktor penting dalam menghasilkan model dengan tingkat akurasi yang baik. Dataset yang digunakan dalam penelitian ini terdiri atas dua kategori citra manusia, yaitu gambar manusia asli (*real image*) yang diperoleh dari dataset publik yang tersedia secara daring (online) serta hasil pengambilan gambar menggunakan kamera, dan gambar manusia hasil generatif Gemini AI (*AI-generated image*) yang dihasilkan melalui *prompt* deskriptif pada aplikasi Gemini. Kedua kategori dataset disusun dengan mempertimbangkan variasi pose, ekspresi wajah, pencahayaan, serta latar belakang agar model dapat mempelajari karakteristik visual yang representatif dari masing-masing kelas (Baharuddin dan Tjahyanto, 2022).

2.14 Image Pre-processing

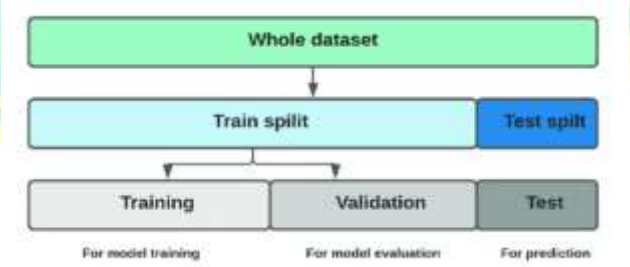
Image Preprocessing merupakan proses awal dalam pengolahan citra yang bertujuan menyiapkan gambar sebelum diproses oleh model *Deep Learning*. Pada tahap ini, citra disesuaikan agar memiliki ukuran, format, dan kualitas yang seragam sehingga data menjadi lebih konsisten. Selain itu, proses ini membantu mengurangi variasi dan gangguan pada citra sehingga model dapat melakukan ekstraksi fitur dan klasifikasi dengan lebih baik. Tahapan *image preprocessing* umumnya meliputi *split dataset*, *resizing*, normalisasi nilai piksel, serta *data*

augmentation untuk meningkatkan kualitas dan keberagaman data pelatihan (Alkausar *et al.*, 2022).

Berdasarkan proses yang dilakukan pada penelitian ini, tahapan *image preprocessing* yang diterapkan terdiri atas dataset splitting, resize image, normalization, dan data augmentation. Masing-masing tahapan dijelaskan sebagai berikut.

1. Dataset Splitting

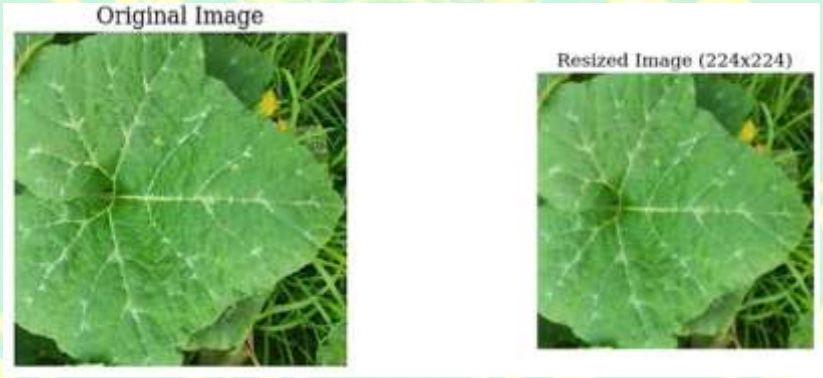
   Dataset Splitting merupakan tahapan yang dilakukan untuk membagi kumpulan data ke dalam beberapa bagian sebelum proses pelatihan model dimulai. Pembagian ini bertujuan agar setiap kelompok data memiliki peran yang berbeda, yaitu sebagai data training, data validation, dan data testing. Dengan pemisahan tersebut, model dapat dilatih, divalidasi, dan diuji menggunakan data yang berbeda sehingga hasil evaluasi menjadi lebih objektif. Selain itu, proses ini juga membantu meningkatkan kemampuan model dalam melakukan generalisasi terhadap data baru serta mengurangi risiko terjadinya *overfitting* (Nurhopipah dan Hasanah, 2020).



Gambar 2.14 Pembagian dataset menjadi training, validation, dan testing (Viola Seda, 2023)

2. Resize Image

Resize Image merupakan salah satu tahapan pada *image preprocessing* yang dilakukan untuk menyesuaikan ukuran dimensi citra dengan ukuran masukan (*input size*) yang dibutuhkan oleh model *Deep Learning*. Proses ini bertujuan menghasilkan seluruh gambar dengan dimensi yang seragam sehingga data dapat diproses secara konsisten oleh model. Selain itu, penyesuaian ukuran citra juga membantu meningkatkan efisiensi komputasi dan mempercepat proses pelatihan tanpa menghilangkan informasi visual yang penting. Pada penelitian ini, seluruh citra diubah menjadi ukuran 224 × 224 piksel agar sesuai dengan spesifikasi masukan arsitektur MobileNetV2, sehingga proses ekstraksi fitur dan klasifikasi dapat dilakukan secara lebih optimal.



Gambar 2.15 Ilustrasi proses Resize Image (Bhuria *et al.*, 2026)

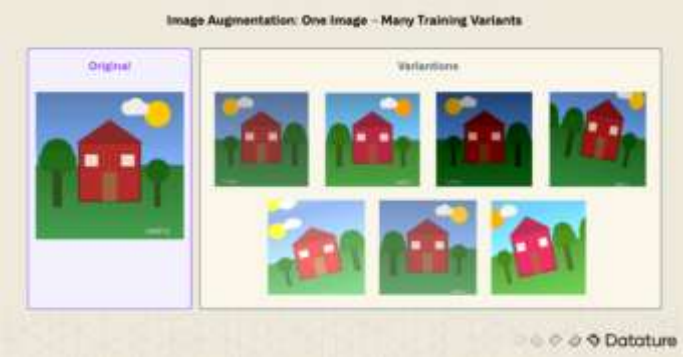
3. Normalization

Normalisasi merupakan tahapan *image preprocessing* yang bertujuan menyamakan rentang nilai piksel citra sebelum diproses oleh model CNN dengan arsitektur MobileNetV2. Pada penelitian ini, citra yang digunakan berupa gambar manusia, baik gambar asli (*non-Gemini*) maupun gambar hasil generatif Gemini AI. Nilai piksel citra yang semula berada pada rentang 0–255

dinormalisasi menggunakan metode Min-Max Normalization dengan membagi setiap nilai piksel terhadap 255, sehingga rentangnya berubah menjadi 0–1. Proses ini diterapkan pada seluruh data training, validation, testing, serta pada citra yang diprediksi melalui aplikasi. Dengan demikian, data masukan memiliki skala yang seragam sehingga proses pelatihan menjadi lebih stabil dan model dapat mengenali karakteristik citra dengan lebih baik (Sarah Rambe *et al.,* 2025)

4. Data Augmentation

Data Augmentation merupakan teknik *image preprocessing* yang bertujuan meningkatkan keragaman data pelatihan melalui pembuatan variasi baru dari citra tanpa mengubah kelas atau labelnya. Variasi tersebut diperoleh dengan menerapkan beberapa transformasi, seperti rotasi (*rotation*), pergeseran (*shift*), pembesaran (*zoom*), dan pembalikan gambar (*flip*). Penerapan teknik ini membantu model mengenali objek dalam berbagai kondisi, meningkatkan kemampuan generalisasi, serta mengurangi risiko *overfitting* selama proses pelatihan. Pada penelitian ini, *data augmentation* diterapkan pada data training yang terdiri atas gambar manusia asli dan gambar hasil Gemini Artificial Intelligence, sehingga model MobileNetV2 dapat mempelajari karakteristik visual dari setiap kelas secara lebih optimal (Solihin *et al.*, 2022)



Gambar 2.16 Contoh proses Data Augmentation pada citra (Keechin, 2026)

2.15 Visual Studio Code (VS Code)

Visual Studio Code (VS Code) merupakan perangkat lunak penyunting kode (*source code editor*) yang dikembangkan oleh Microsoft dan banyak dimanfaatkan dalam pengembangan perangkat lunak. Aplikasi ini mendukung berbagai bahasa pemrograman, seperti Python, Java, C++, dan JavaScript, serta menyediakan antarmuka yang ringan dan mudah digunakan. Selain itu, VS Code dilengkapi dengan berbagai fitur, seperti *IntelliSense*, *debugging*, terminal terintegrasi, dan dukungan ekstensi (*extensions*) yang dapat meningkatkan produktivitas dalam proses pengembangan aplikasi. Berdasarkan penelitian Murani dkk. (2025), Visual Studio Code dinilai efektif sebagai editor pemrograman karena memiliki antarmuka yang intuitif dan fitur-fitur pendukung yang memudahkan pengguna, terutama *IntelliSense* dan *debugging* terintegrasi, sehingga mampu meningkatkan efisiensi serta meminimalkan kesalahan dalam penulisan kode program.(Murani *et al.*, 2025).

Dalam penelitian ini, Visual Studio Code digunakan sebagai lingkungan pengembangan (*Integrated Development Environment* atau IDE) untuk mengimplementasikan model Convolutional Neural Network (CNN) dengan

arsitektur MobileNetV2 menggunakan bahasa pemrograman Python. Melalui VS Code, proses penulisan kode program, *image pre-processing*, pelatihan (*training*), pengujian (*testing*), hingga evaluasi model dilakukan secara terintegrasi dengan pustaka TensorFlow, Keras, OpenCV, dan NumPy. Penggunaan Visual Studio Code diharapkan dapat mendukung proses pengembangan sistem deteksi gambar manusia hasil Gemini Artificial Intelligence secara lebih efektif dan efisien.

2.16 Training Model

Training model merupakan tahapan pembelajaran pada model deep learning untuk mengenali pola dan karakteristik yang terdapat pada dataset. Selama proses ini, model mempelajari hubungan antara data masukan (*input*) dan label (*output*) melalui pembaruan parameter secara berulang hingga mampu menghasilkan prediksi yang akurat terhadap data baru. Pada penelitian ini, proses pelatihan dilakukan menggunakan arsitektur MobileNetV2 dengan pendekatan transfer learning untuk mengklasifikasikan gambar manusia asli (*real image*) dan gambar hasil generatif Gemini AI. Keberhasilan proses pelatihan dipengaruhi oleh beberapa parameter penting, yaitu:

1. *Epoch* merupakan satu kali siklus pelatihan ketika seluruh data pada training dataset diproses oleh model dari awal hingga akhir. Penambahan jumlah *epoch* memberikan kesempatan bagi model untuk mempelajari karakteristik data secara lebih mendalam. Namun, penggunaan *epoch* yang terlalu banyak dapat menyebabkan overfitting, yaitu kondisi ketika model mampu mengenali data pelatihan dengan sangat baik, tetapi mengalami penurunan kinerja saat dihadapkan pada data baru.

2. *Batch size* adalah jumlah sampel data yang diproses dalam satu kali pembaruan parameter model selama pelatihan. Dataset dibagi menjadi beberapa kelompok (*batch*) sehingga proses komputasi menjadi lebih efisien dan tidak memerlukan seluruh data diproses secara bersamaan. Besarnya nilai *batch size* akan memengaruhi penggunaan memori, kecepatan pelatihan, serta stabilitas proses pembelajaran.
3. *Learning rate* merupakan parameter yang menentukan besarnya perubahan bobot (*weights*) pada setiap proses pembaruan model. Nilai *learning rate* yang terlalu besar dapat menyebabkan model sulit mencapai kondisi optimal, sedangkan nilai yang terlalu kecil membuat proses pelatihan berlangsung lebih lambat. Oleh karena itu, pemilihan *learning rate* yang sesuai menjadi salah satu faktor penting untuk memperoleh performa model yang optimal.
4. *Training* adalah tahap pembelajaran model menggunakan training dataset. Pada tahap ini, MobileNetV2 mempelajari berbagai karakteristik visual dari gambar manusia asli dan gambar hasil Gemini AI melalui mekanisme forward propagation dan backpropagation. Proses tersebut bertujuan untuk memperbarui bobot model agar kesalahan prediksi semakin kecil dan kemampuan klasifikasi semakin meningkat.
5. *Validation* merupakan proses evaluasi model menggunakan validation dataset, yaitu data yang tidak digunakan selama pelatihan. Tahapan ini bertujuan untuk mengukur kemampuan model dalam melakukan generalisasi terhadap data yang belum pernah dipelajari, sekaligus

mendeteksi kemungkinan terjadinya overfitting maupun underfitting selama proses pelatihan.

6. *Loss* adalah nilai yang menggambarkan tingkat kesalahan antara hasil prediksi model dan label sebenarnya. Semakin kecil nilai *loss*, semakin baik kemampuan model dalam melakukan klasifikasi. Nilai ini digunakan sebagai acuan untuk memperbarui bobot model agar kesalahan prediksi terus berkurang pada setiap iterasi pelatihan.
7. *Optimizer* merupakan algoritma yang bertugas memperbarui bobot (*weights*) dan bias pada jaringan saraf selama proses pelatihan. Pembaruan tersebut dilakukan berdasarkan hasil backpropagation dengan tujuan meminimalkan nilai *loss* dan meningkatkan performa model. Salah satu optimizer yang paling banyak digunakan adalah Adam (Adaptive Moment Estimation) karena mampu mempercepat proses konvergensi serta menghasilkan pelatihan yang lebih stabil dibandingkan beberapa metode optimasi lainnya.

2.17 Fungsi Loss

Fungsi loss merupakan ukuran yang digunakan untuk mengetahui tingkat kesalahan antara hasil prediksi model dan nilai sebenarnya (*ground truth*) selama proses pelatihan. Nilai *loss* menjadi dasar bagi model dalam melakukan pembaruan bobot (*weights*) agar kemampuan klasifikasi semakin baik. Semakin rendah nilai *loss* yang diperoleh, semakin kecil kesalahan prediksi yang dihasilkan oleh model sehingga performa klasifikasi menjadi lebih optimal (Wijaya *et al.*, 2025).

Pada penelitian ini, model digunakan untuk mengklasifikasikan dua kategori citra, yaitu gambar manusia asli (*real image*) dan gambar hasil generatif Gemini AI (*AI-generated image*). Oleh karena itu, fungsi *loss* yang diterapkan adalah Binary Cross Entropy, yaitu fungsi yang dirancang untuk menangani permasalahan klasifikasi biner (*binary classification*) dengan membandingkan probabilitas hasil prediksi terhadap label sebenarnya

2.17.1 Binary Cross Entropy

Binary Cross Entropy merupakan fungsi *loss* yang umum digunakan pada model klasifikasi dengan dua kelas. Fungsi ini mengukur perbedaan antara probabilitas yang diprediksi oleh model dan nilai target sebenarnya, sehingga dapat diketahui besarnya kesalahan yang terjadi selama proses pelatihan. Nilai *loss* yang semakin kecil menunjukkan bahwa hasil prediksi model semakin mendekati label yang sebenarnya (Salsabil *et al.*, 2025).

Persamaan Binary Cross Entropy ditunjukkan sebagai berikut.

$$L = -[y \log(\hat{y}) + (1 - y) \log(1 - \hat{y})]$$

Keterangan:

- L = Nilai *loss* atau tingkat kesalahan model.
- y = Label sebenarnya (*ground truth*), bernilai **0** atau **1**.
- $\hat{y}$ = Probabilitas hasil prediksi model terhadap kelas positif.
- log = Fungsi logaritma natural.

2.18 Optimizer

Optimizer merupakan algoritma yang digunakan untuk memperbarui bobot (*weights*) pada model selama proses pelatihan. Tujuan utama optimizer adalah

mengurangi nilai loss sehingga model dapat menghasilkan prediksi yang lebih akurat. Pembaruan bobot dilakukan secara bertahap berdasarkan tingkat kesalahan yang diperoleh pada setiap iterasi pelatihan. Pada penelitian ini, optimizer yang digunakan adalah Adam (Adaptive Moment Estimation). Adam merupakan salah satu optimizer yang banyak digunakan pada model Convolutional Neural Network (CNN) karena memiliki proses pembelajaran yang cepat, stabil, dan mampu menyesuaikan nilai learning rate secara otomatis pada setiap parameter model. Penggunaan Adam dipilih karena mampu meningkatkan efisiensi proses pelatihan serta membantu model mencapai hasil klasifikasi yang lebih baik (Astuti *et al.*, 2022).

2.19 Overfitting dan Underfitting

Dalam proses pelatihan deep learning, kemampuan model dalam melakukan generalisasi terhadap data baru merupakan aspek yang sangat penting. Model yang baik tidak hanya memiliki performa tinggi pada data pelatihan, tetapi juga mampu memberikan hasil yang akurat pada data yang belum pernah dipelajari. Salah satu permasalahan yang sering terjadi adalah overfitting, yaitu kondisi ketika model terlalu menyesuaikan diri dengan data pelatihan sehingga menghasilkan akurasi tinggi pada training dataset, namun performanya menurun pada validation dataset atau data baru. Sebaliknya, underfitting terjadi ketika model belum mampu mempelajari pola pada data pelatihan dengan baik, sehingga menghasilkan performa yang rendah baik pada data pelatihan maupun data validasi. Kedua kondisi tersebut dapat diamati melalui perubahan nilai training loss, validation loss,

training accuracy, dan validation accuracy selama proses pelatihan model (Firmansyach *et al.*, 2023).

2.20 Evaluasi Model

Pada sebuah model klasifikasi memiliki tingkat performa dari proses klasifikasi yang dihasilkan. Tingkat performa model klasifikasi tersebut dapat diukur untuk mengetahui seberapa efisien dan akurat untuk mengklasifikasi sebuah objek dalam menyelesaikan suatu permasalahan. Confusion Matrix merupakan salah satu metode evaluasi yang digunakan untuk mengukur kinerja model klasifikasi dengan membandingkan hasil prediksi model terhadap label sebenarnya (actual label). Melalui Confusion Matrix, dapat diketahui jumlah prediksi yang benar maupun salah pada setiap kelas sehingga performa model dapat dianalisis secara lebih rinci. Selain menghitung tingkat akurasi (accuracy), Confusion Matrix juga menjadi dasar dalam memperoleh nilai precision, recall, dan F1-score, yang digunakan untuk mengevaluasi kemampuan model dalam melakukan klasifikasi (Karnadi dan Handhayani, 2024). Terdapat beberapa istilah umum yang dapat dipakai dalam proses pengukuran kinerja dari model klasifikasi yatu:

1. *True Positive* (TP) merupakan data positif yang diprediksi benar
2. *True Negative* (TN) merupakan data negatif yang diprediksi benar
3. *False Positive* (FP) merupakan data negatif namun diprediksi sebagai data positif
4. *False Negative* (FN) merupakan data positif namun diprediksi sebagai data negatif

Tabel 2.2 *Confusion Matrix*

		Predicted Class	
Actual Class		TP	FN
		FP	TN

Berdasarkan nilai TP, TN, FP, dan FN, performa model dievaluasi menggunakan beberapa metrik, yaitu Accuracy, Precision, Recall, dan F1-Score.

1. Accuracy

   Accuracy merupakan ukuran yang menunjukkan tingkat ketepatan model dalam melakukan klasifikasi terhadap seluruh data uji. Nilai ini dihitung berdasarkan jumlah prediksi yang benar dibandingkan dengan total data yang digunakan, sehingga mencerminkan performa model secara keseluruhan, berikut rumusnya:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

2. Precision

   Precision merupakan metrik evaluasi yang digunakan untuk mengukur ketepatan model dalam memprediksi data pada kelas positif. Nilai *precision* menunjukkan perbandingan antara jumlah prediksi positif yang benar dengan seluruh data yang diprediksi sebagai kelas positif. Adapun rumus dari precision:

$$\text{Precision} = \frac{TP}{TP + FP} \times 100\%$$

3. Recall

Recall merupakan metrik evaluasi yang digunakan untuk mengukur kemampuan model dalam mengenali seluruh data yang termasuk ke dalam kelas positif. Nilai *recall* menunjukkan perbandingan antara jumlah data positif yang berhasil diprediksi dengan benar terhadap seluruh data positif yang sebenarnya. Adapun rumus recall:

$$\text{Recall} = \frac{TP}{TP + FN} \times 100\%$$

4. F1-Score

F1-Score merupakan metrik evaluasi yang menggabungkan nilai precision dan recall dalam bentuk rata-rata harmonis. Metrik ini digunakan untuk memberikan penilaian yang lebih seimbang terhadap performa model, terutama ketika distribusi jumlah data pada setiap kelas tidak seimbang. Adapun rumus F1-Score:

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$