
Perbandingan Kinerja Arsitektur CNN: Studi Kasus pada LeNet-5, AlexNet, dan ResNet-50 Menggunakan Dataset CIFAR-10

Onixa Shafa Putri Wibowo¹, Silvia Febriani², Siti Nurbayanah³

¹Teknik Informatika, Universitas Islam Negeri Sunan Gunung Djati Bandung

Article Info

Article history:

Received month dd, yyyy

Revised month dd, yyyy

Accepted month dd, yyyy

Keywords:

Convolutional Neural Network (CNN)

LeNet-5

AlexNet

ResNet-50

CIFAR-10

ABSTRACT

Penelitian ini membandingkan kinerja tiga arsitektur Convolutional Neural Network (CNN) – LeNet-5, AlexNet, dan ResNet-50 – dalam tugas klasifikasi citra menggunakan dataset CIFAR-10. Masalah utama yang diatasi adalah efektivitas dan efisiensi model CNN yang bervariasi dalam mengenali objek pada citra yang kompleks. Metode yang digunakan meliputi pra-pemrosesan data seperti normalisasi piksel dan augmentasi data (khusus ResNet-50), serta pemodelan dan pelatihan masing-masing arsitektur dengan konfigurasi *hyperparameter* yang dioptimalkan. Ketiga model dilatih menggunakan fungsi *loss Cross-Entropy* dan *optimizer* (Adam untuk LeNet-5 dan AlexNet; SGD dengan Nesterov momentum untuk ResNet-50). Hasil eksperimen menunjukkan bahwa ResNet-50 mencapai akurasi klasifikasi terbaik sebesar 91.56% pada *test set* CIFAR-10. AlexNet menempati posisi kedua dengan akurasi 73.71%, sementara LeNet-5 memiliki akurasi 63.73%. Analisis *loss* pelatihan menunjukkan bahwa semua model berhasil belajar dari data, dengan ResNet-50 menunjukkan konvergensi *loss* validasi yang paling stabil dan rendah (0.7196). Dari segi akurasi per kelas, ResNet-50 juga mendominasi, dengan akurasi tertinggi untuk kelas 'automobile' (97.30%) dan peningkatan signifikan pada kelas 'cat' (80.20%) yang sering menjadi tantangan. Akurasi per kelas terendah pada LeNet-5 adalah 45.90% (kelas 'cat'), pada AlexNet adalah 56.90% (kelas 'cat'), dan pada ResNet-50 adalah 80.20% (kelas 'cat').

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Onixa Shafa Putri Wibowo

Jurusan Teknik Informatika, UIN Sunan Gunung Djati Bandung

Email: onixashafa@gmail.com

1. PENDAHULUAN

Kemajuan teknologi kecerdasan buatan (Artificial Intelligence) dalam beberapa tahun terakhir telah memberikan kontribusi yang signifikan terhadap berbagai disiplin ilmu, termasuk dalam pengolahan data visual atau citra digital. Salah satu pendekatan dalam deep learning yang banyak digunakan dalam konteks ini adalah *Convolutional Neural Network* (CNN). CNN merupakan jenis jaringan saraf tiruan yang secara khusus dirancang untuk menangani data visual dua dimensi, dengan kemampuan dalam mengekstraksi fitur spasial secara otomatis. Melalui proses ini, CNN dapat mengenali pola dan tekstur pada gambar tanpa memerlukan teknik ekstraksi fitur secara manual. Hal ini menjadikan CNN sangat efektif untuk menyelesaikan berbagai tugas klasifikasi citra (Vincent et al., 2025).

Seiring dengan meningkatnya kebutuhan terhadap metode klasifikasi citra yang lebih efisien, berbagai arsitektur CNN telah dikembangkan. Di antara model-model tersebut, terdapat tiga arsitektur yang sering dijadikan acuan dalam penelitian karena tingkat efektivitasnya, yaitu LeNet-5, AlexNet, dan ResNet.

Arsitektur LeNet-5 yang dirancang oleh Yann LeCun pada tahun 1998 merupakan salah satu model awal CNN yang terbukti berhasil dalam klasifikasi angka tulisan tangan pada dataset MNIST. Struktur LeNet-5 terdiri dari tujuh lapisan utama, yaitu dua lapisan konvolusi, dua lapisan subsampling, dan tiga lapisan *fully connected*. Keunggulan LeNet-5 terletak pada kesederhanaan desain dan efisiensi dalam pelatihan, menjadikannya ideal untuk digunakan pada sistem dengan keterbatasan komputasi atau perangkat *edge computing*. Studi oleh Li et al. (2025) menunjukkan bahwa LeNet-5 dapat diterapkan secara efektif dalam sistem pembacaan meter digital dengan tingkat akurasi yang tinggi. Selain itu, penelitian yang dilakukan oleh Etaba et al. (2025) juga membuktikan bahwa model ini dapat dimanfaatkan dalam proses deteksi dini penyakit tanaman ubi jalar melalui pengolahan citra daun secara otomatis, yang berkontribusi terhadap peningkatan produktivitas pertanian. Lubis dan Susilawati (2025) juga mendemonstrasikan efektivitas LeNet-5 dalam mengenali tulisan tangan angka pada dataset MNIST dengan akurasi tinggi, menegaskan relevansi model ini pada domain pengenalan pola visual sederhana.

Seiring dengan meningkatnya kompleksitas data yang dihadapi dalam pengolahan citra, maka diperlukan arsitektur yang lebih dalam dan kuat. AlexNet, yang diperkenalkan oleh Krizhevsky dan rekan-rekannya pada tahun 2012, menjadi solusi atas kebutuhan tersebut dan mencatatkan pencapaian penting dengan memenangkan kompetisi *ImageNet Large Scale Visual Recognition Challenge* (ILSVRC). Model ini terdiri dari delapan lapisan (lima konvolusi dan tiga *fully connected*), serta memanfaatkan fungsi aktivasi ReLU, teknik *dropout*, dan augmentasi data untuk mengurangi risiko *overfitting*. Menurut Muda et al. (2025), AlexNet berhasil diadaptasi dalam bidang medis, khususnya untuk mengevaluasi postur tubuh pasien kanker payudara menggunakan data *skeleton*. Hasilnya menunjukkan bahwa model ini mampu menghasilkan klasifikasi yang akurat terhadap gerakan tubuh pasien yang sedang menjalani rehabilitasi.

Selanjutnya, ResNet (*Residual Network*) dikembangkan oleh Kaiming He dan timnya pada tahun 2015 sebagai penyempurnaan terhadap arsitektur CNN yang sangat dalam. Inovasi utama dari ResNet terletak pada penggunaan *residual connections* atau *skip connections*, yang memungkinkan sinyal informasi melewati satu atau beberapa lapisan secara langsung. Teknik ini secara efektif mengatasi permasalahan *vanishing gradient* yang sering terjadi pada jaringan saraf yang dalam, dan memungkinkan pelatihan jaringan dengan ratusan lapisan. Muda et al. (2025) juga menemukan bahwa ResNet memberikan performa pelatihan yang lebih stabil dibandingkan dengan model CNN konvensional maupun RNN dalam klasifikasi data *skeleton*, serta menunjukkan ketahanan yang lebih baik terhadap *overfitting* meskipun menggunakan arsitektur dalam. Selain itu, Setiawan (2025) menunjukkan bahwa arsitektur ResNet mampu mendeteksi malaria berbasis citra sel darah dengan akurasi hingga 97,2%, menjadikannya sangat andal dalam domain medis. Hasil serupa juga diperoleh oleh Agustin et al. (2025) dalam konteks klasifikasi irama murattal Al-Qur'an, di mana arsitektur ResNet50 unggul dalam mendeteksi karakteristik audio dibandingkan dengan VGG16.

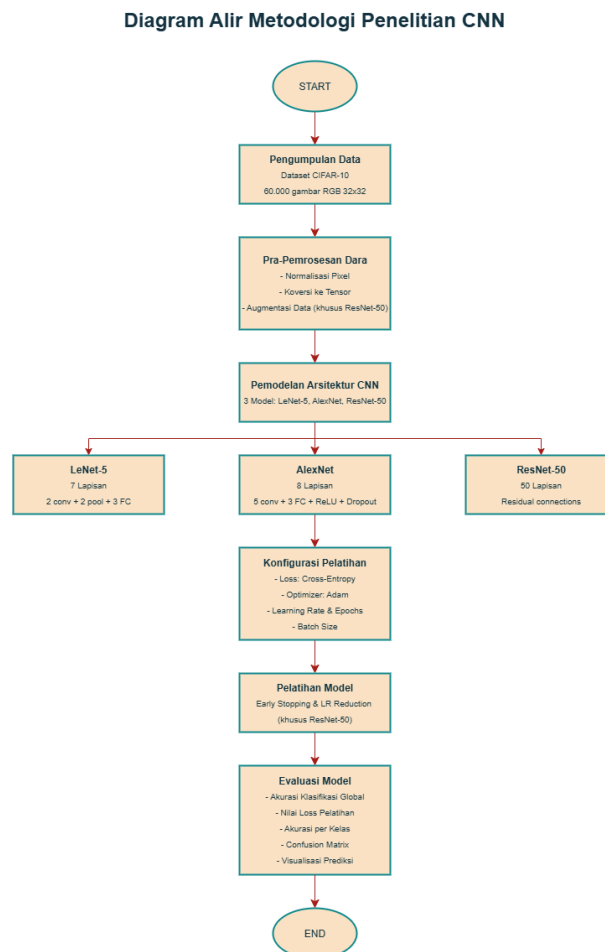
Untuk mengukur dan membandingkan performa dari ketiga arsitektur CNN tersebut, digunakan dataset CIFAR-10, yaitu himpunan data *benchmark* yang terdiri dari 60.000 gambar RGB berukuran 32×32 piksel dengan sepuluh kelas objek berbeda, seperti pesawat, mobil, katak, dan kucing. Dataset ini dipilih karena memiliki tingkat kompleksitas menengah yang cukup menantang bagi arsitektur sederhana seperti LeNet-5, namun tetap memungkinkan untuk diuji oleh model yang lebih kompleks seperti ResNet. Sebagaimana dijelaskan oleh Risman Hadi et al. (2025), CIFAR-10 merupakan dataset yang sesuai untuk menguji kemampuan generalisasi model CNN karena keragaman objek dan ukuran gambar yang terkontrol. Studi lain oleh Fahri et al. (2025) menunjukkan bahwa kombinasi CNN seperti EfficientNet-B0 dan InceptionV3 juga mampu menghasilkan klasifikasi akurat dalam deteksi tumor ginjal, menegaskan pentingnya pemilihan arsitektur dalam konteks aplikasi citra medis. Bahkan, perbandingan model-model CNN juga telah digunakan dalam domain non-visual seperti rekayasa sipil, misalnya oleh Pradana et al. (2025), yang menilai efektivitas arsitektur dalam konteks *Building Information Modeling* (BIM).

Dengan latar belakang tersebut, penelitian ini bertujuan untuk membandingkan kinerja arsitektur LeNet-5, AlexNet, dan ResNet dalam klasifikasi citra menggunakan dataset CIFAR-10. Evaluasi dilakukan terhadap beberapa aspek, yaitu akurasi klasifikasi, nilai loss, waktu pelatihan, serta efisiensi arsitektur dalam hal jumlah parameter dan kecepatan inferensi. Tujuan dari penelitian ini adalah untuk memberikan gambaran

empiris yang jelas mengenai efektivitas dan efisiensi masing-masing arsitektur dalam skenario yang setara. Selain itu, hasil dari penelitian ini diharapkan dapat memberikan kontribusi nyata dalam pengembangan sistem klasifikasi berbasis deep learning yang adaptif, baik dalam bidang medis, pertanian, pendidikan Al-Qur'an, maupun sistem *edge computing*, serta memperkaya literatur mengenai pemilihan arsitektur CNN yang tepat berdasarkan konteks aplikatif dan sumber daya yang tersedia.

2. METODE

Penelitian ini mengimplementasikan dan membandingkan kinerja tiga arsitektur Convolutional Neural Network (CNN) yang berbeda -LeNet-5, AlexNet, dan ResNet-50 -untuk tugas klasifikasi gambar pada dataset CIFAR-10. Metodologi yang digunakan dalam studi ini dirancang untuk memastikan perbandingan yang adil dan akurat antar model, mencakup tahapan mulai dari pengumpulan data hingga evaluasi kinerja model.



Gambar 1. Diagram Alir Metodologi Penelitian

2.1. Pengumpulan Data

Dataset utama yang digunakan dalam eksperimen ini adalah CIFAR-10. CIFAR-10 merupakan salah satu *benchmark* dataset yang paling sering digunakan dalam penelitian klasifikasi objek. Dataset ini terdiri dari 60.000 citra berwarna berukuran 32×32 piksel, yang terbagi rata ke dalam 10 kelas objek yang berbeda. Kelas-kelas objek tersebut meliputi pesawat (airplane), mobil (automobile), burung (bird), kucing (cat), rusa (deer), anjing (dog), katak (frog), kuda (horse), kapal (ship), dan truk (truck). Pembagian dataset secara standar adalah 50.000 gambar untuk pelatihan (*training set*) dan 10.000 gambar untuk pengujian (*test set*).

Pemilihan dataset CIFAR-10 didasarkan pada beberapa pertimbangan. Pertama, CIFAR-10 menawarkan tingkat kompleksitas visual yang lebih tinggi dibandingkan dengan dataset sederhana seperti MNIST (yang hanya berisi angka tulisan tangan) namun masih dapat dikelola dengan sumber daya komputasi yang wajar untuk perbandingan model. Gambar-gambar dalam CIFAR-10 memiliki variasi latar belakang, pose objek, dan kondisi pencahayaan yang lebih beragam, sehingga menjadi tantangan yang baik untuk menguji kemampuan generalisasi model CNN. Kedua, ukuran citra 32×32 piksel memungkinkan pelatihan yang relatif cepat, sehingga memfasilitasi eksperimen komparatif pada berbagai arsitektur model.

2.2. Pra-Pemrosesan Data

Tahap pra-pemrosesan data sangat krusial untuk memastikan data dalam format yang optimal sebelum disalurkan ke model jaringan saraf tiruan. Dalam penelitian ini, beberapa langkah pra-pemrosesan diterapkan pada citra dataset CIFAR-10:

1. **Normalisasi Piksel:** Semua nilai piksel dalam citra dinormalisasi. Normalisasi ini mengubah rentang nilai piksel dari 0-255 menjadi rentang yang lebih kecil dan konsisten.
 - Untuk LeNet-5 dan ResNet-50, nilai piksel dinormalisasi ke rentang $[0, 1]$ dengan membagi setiap nilai piksel dengan 255.0. Setelah itu, untuk ResNet-50 dilakukan normalisasi tambahan menggunakan mean dan std deviasi CIFAR-10.
 - Untuk AlexNet, normalisasi dilakukan ke rentang $[-1, 1]$ menggunakan rata-rata (0.5, 0.5, 0.5) dan standar deviasi (0.5, 0.5, 0.5) untuk setiap saluran RGB. Normalisasi ini membantu mempercepat konvergensi model dan meningkatkan stabilitas pelatihan dengan memastikan semua fitur memiliki skala yang serupa.
2. **Konversi ke Tensor:** Setelah normalisasi, citra dikonversi menjadi format tensor. Tensor adalah struktur data fundamental dalam *framework deep learning* (PyTorch dan TensorFlow/Keras) yang memungkinkan operasi matematika efisien pada GPU. Citra CIFAR-10 yang awalnya berbentuk array NumPy dengan dimensi (tinggi, lebar, saluran) diubah menjadi tensor dengan dimensi (saluran, tinggi, lebar) untuk PyTorch dan (tinggi, lebar, saluran) untuk TensorFlow/Keras, sesuai dengan persyaratan *input* masing-masing *framework*.
3. **Augmentasi Data (khusus ResNet-50):** Untuk meningkatkan kemampuan generalisasi model dan mengurangi risiko *overfitting*, khususnya pada arsitektur yang lebih dalam seperti ResNet-50, teknik augmentasi data diterapkan pada *training set*. Augmentasi data melibatkan pembuatan variasi baru dari citra pelatihan yang ada tanpa menambah citra fisik. Teknik-teknik yang digunakan meliputi:
 - a. **Rotasi** (*rotation_range* = 15): Memutar gambar secara acak hingga 15 derajat.
 - b. **Pergeseran Lebar dan Tinggi** (*width_shift_range* = 0.125, *height_shift_range* = 0.125): Menggeser gambar secara horizontal atau vertikal hingga 12.5% dari total lebar atau tinggi gambar.
 - c. **Horizontal Flip** (*horizontal_flip* = True): Membalik gambar secara horizontal.
 - d. **Zoom** (*zoom_range* = 0.1): Memperbesar atau memperkecil bagian gambar secara acak hingga 10%.
 - e. **Shear** (*shear_range* = 0.1): Memberikan efek geser pada gambar.
 - f. **Channel Shift** (*channel_shift_range* = 0.1): Menggeser nilai intensitas piksel pada setiap saluran warna.
 - g. **Brightness** (*brightness_range* = $[0.9, 1.1]$): Mengubah kecerahan gambar secara acak.
 - h. **Cutout:** Secara acak menghilangkan bagian persegi dari gambar. Penerapan augmentasi data ini secara signifikan memperkaya variasi *training set*, membantu model untuk belajar fitur yang lebih *robust* dan tidak terlalu sensitif terhadap variasi kecil dalam data *input*.

2.3. Pemodelan Arsitektur CNN

Tiga arsitektur CNN yang berbeda dipilih untuk studi komparatif ini, merepresentasikan evolusi arsitektur CNN dari yang klasik hingga yang modern:

2.3.1. LeNet-5

LeNet-5, yang diperkenalkan oleh Yann LeCun pada tahun 1998, adalah salah satu pionir arsitektur CNN yang berhasil diterapkan dalam klasifikasi angka tulisan tangan pada dataset MNIST. Struktur LeNet-5 terdiri dari tujuh lapisan utama: dua lapisan konvolusi,

dua lapisan *subsampling* (pooling), dan tiga lapisan *fully connected*. Keunggulan LeNet-5 terletak pada kesederhanaan desainnya dan efisiensi komputasi, menjadikannya model dasar yang ideal untuk perbandingan. Dalam penelitian ini, implementasi LeNet-5 disesuaikan untuk menerima input gambar berwarna (3 *in_channels*) dan menghasilkan 10 kelas output sesuai CIFAR-10.

2.3.2. AlexNet

AlexNet, yang diperkenalkan oleh Krizhevsky, Sutskever, dan Hinton pada tahun 2012, menjadi titik balik dalam sejarah *deep learning* dengan memenangkan ImageNet Large Scale Visual Recognition Challenge (ILSVRC) secara signifikan. Arsitektur ini lebih dalam dan kompleks dibandingkan LeNet-5, terdiri dari delapan lapisan: lima lapisan konvolusi dan tiga lapisan *fully connected*. Fitur-fitur inovatif AlexNet meliputi penggunaan fungsi aktivasi ReLU (Rectified Linear Unit), yang membantu mengatasi masalah *vanishing gradient* dan mempercepat pelatihan; teknik *dropout* untuk mengurangi *overfitting* dengan secara acak menonaktifkan neuron selama pelatihan; dan augmentasi data yang ekstensif. Dalam penelitian ini, AlexNet diadaptasi untuk dataset CIFAR-10 dengan input 32×32 piksel dan penyesuaian pada lapisan konvolusi awal untuk ukuran input yang lebih kecil.

2.3.3. ResNet-50

ResNet (Residual Network), yang dikembangkan oleh Kaiming He dan timnya pada tahun 2015, merevolusi arsitektur CNN yang sangat dalam. Inovasi utama ResNet adalah pengenalan *residual connections* atau *skip connections*, yang memungkinkan sinyal informasi melewati satu atau beberapa lapisan secara langsung. Teknik ini secara efektif mengatasi permasalahan *vanishing/exploding gradient* dan *degradation* yang sering terjadi pada jaringan saraf yang sangat dalam, memungkinkan pelatihan jaringan dengan ratusan bahkan ribuan lapisan. ResNet-50 adalah versi ResNet dengan 50 lapisan *deep*. Untuk dataset CIFAR-10, implementasi ResNet-50 dalam penelitian ini disesuaikan dengan mengubah ukuran *kernel_size* dan *strides* pada lapisan konvolusi awal, serta menghilangkan lapisan *max pooling* awal yang biasanya ada pada arsitektur ResNet untuk ImageNet, agar *feature maps* tidak terlalu kecil untuk ukuran gambar 32×32 . Lapisan *GlobalAveragePooling2D* dan lapisan *Dropout* ditambahkan di bagian akhir sebelum lapisan *Dense* output.

2.4. Konfigurasi Pelatihan

Setiap model dilatih dengan konfigurasi yang telah dioptimalkan untuk performa terbaik pada dataset CIFAR-10, dengan mempertimbangkan karakteristik unik masing-masing arsitektur:

1. **Fungsi Loss:** *Cross-Entropy Loss* (atau *Categorical Crossentropy* untuk TensorFlow/Keras) digunakan sebagai fungsi kerugian standar untuk semua model. Fungsi ini sangat sesuai untuk tugas klasifikasi multikelas, mengukur perbedaan antara distribusi probabilitas yang diprediksi model dan distribusi probabilitas sebenarnya dari label.
 2. **Optimizer:** *Optimizer Adam (Adaptive Moment Estimation)* dipilih untuk mengoptimalkan bobot model pada LeNet-5 dan AlexNet. Adam dikenal karena efisiensinya dalam menangani *sparse gradients* dan kemampuannya untuk mengadaptasi *learning rate* selama pelatihan. Untuk ResNet-50, *optimizer* SGD (Stochastic Gradient Descent) dengan *Nesterov momentum* dan *weight decay* ($5e-4$) digunakan.
 - a. Untuk **LeNet-5** dan **AlexNet**, *learning rate* awal diatur sebesar **0.001**.
 - b. Untuk **ResNet-50**, *learning rate* awal diatur sebesar **0.1**, dengan *cosine annealing scheduler* untuk penyesuaian *learning rate* seiring *epoch*. Selain itu, *ReduceLROnPlateau callback* diimplementasikan pada pelatihan ResNet-50, yang secara otomatis mengurangi *learning rate* sebesar faktor 0.5 jika *validation loss* tidak membaik selama 10 *epoch* berturut-turut, dengan *minimum learning rate* $1e-6$. Mekanisme ini membantu model untuk menemukan konvergensi yang lebih baik.
 3. **Epochs:** Jumlah *epoch* menentukan berapa kali model akan melihat seluruh *training set*.
-

- a. **LeNet-5** dan **AlexNet** dilatih selama 10 *epoch*.
- b. **ResNet-50** diatur untuk melatih hingga 100 *epoch*. Namun, untuk mencegah *overfitting* dan menghemat waktu komputasi, *callback EarlyStopping* diimplementasikan dengan *patience* 30. Pelatihan akan berhenti jika *validation accuracy* tidak menunjukkan peningkatan selama jumlah *patience* yang ditentukan berturut-turut, dan bobot model terbaik akan dikembalikan (*restore best weights=True*).
4. **Batch Size:** *Batch size* mengacu pada jumlah sampel yang diproses sebelum *update* bobot model.
 - a. **LeNet-5** menggunakan *batch size* 10.
 - b. **AlexNet** dan **ResNet-50** menggunakan *batch size* 64. Untuk ResNet-50, jika GPU tidak tersedia, *batch size* disesuaikan menjadi 32. Penggunaan *batch size* yang lebih besar umumnya mempercepat pelatihan pada GPU.
5. **Perangkat Komputasi:** Pelatihan dilakukan pada perangkat yang mendukung komputasi paralel. Untuk implementasi PyTorch (LeNet-5 dan AlexNet), model akan secara otomatis beralih ke GPU (CUDA) jika tersedia; jika tidak, CPU akan digunakan. Untuk implementasi TensorFlow/Keras (ResNet-50), sistem secara eksplisit memeriksa ketersediaan GPU dan menggunakannya jika ada.

2.5. Evaluasi Model

Evaluasi komprehensif dilakukan untuk membandingkan kinerja ketiga arsitektur CNN. Metrik-metrik berikut digunakan:

1. **Akurasi Klasifikasi Global:** Akurasi keseluruhan model pada *test set* dihitung sebagai persentase prediksi yang benar dari total sampel. Metrik ini memberikan gambaran umum tentang kinerja model.
2. **Nilai Loss Pelatihan:** Nilai *loss* pelatihan dicatat pada setiap *epoch* untuk memantau proses konvergensi model. Tren penurunan *loss* yang stabil mengindikasikan bahwa model belajar secara efektif.
3. **Akurasi per Kelas:** Untuk mendapatkan pemahaman yang lebih rinci tentang kinerja model pada setiap kelas objek, akurasi per kelas dihitung secara individual. Hal ini membantu mengidentifikasi kelas-kelas yang menjadi tantangan bagi model atau kelas-kelas di mana model berkinerja sangat baik.
4. **Confusion Matrix (khusus AlexNet dan ResNet-50):** *Confusion matrix* dihitung untuk AlexNet untuk memberikan visualisasi rinci tentang kinerja model pada setiap kelas, menunjukkan jumlah *True Positives (TP)*, *False Positives (FP)*, *False Negatives (FN)*, dan *True Negatives (TN)*. Ini memungkinkan analisis yang lebih mendalam mengenai jenis kesalahan yang dilakukan model (misklasifikasi antar kelas).
5. **Contoh Prediksi (khusus AlexNet dan ResNet-50):** Beberapa contoh gambar dari *test set* ditampilkan bersama dengan prediksi model dan label sebenarnya. Ini menyediakan inspeksi kualitatif langsung tentang bagaimana model melakukan prediksi pada citra individu.

3. HASIL PENELITIAN

Pada bagian ini, performa ketiga arsitektur Convolutional Neural Network (CNN) – LeNet-5, AlexNet, dan ResNet-50 – akan disajikan dan dianalisis secara komprehensif. Eksperimen dilakukan menggunakan dataset CIFAR-10, dan evaluasi kinerja model didasarkan pada metrik-metrik kunci seperti akurasi klasifikasi global, akurasi per kelas, nilai *loss* pelatihan, jumlah parameter model, dan waktu yang diperlukan untuk proses pelatihan. Hasil yang diperoleh akan memberikan gambaran empiris mengenai efektivitas dan efisiensi masing-masing arsitektur dalam skenario klasifikasi citra pada dataset CIFAR-10.

3.1. Akurasi Klasifikasi Global

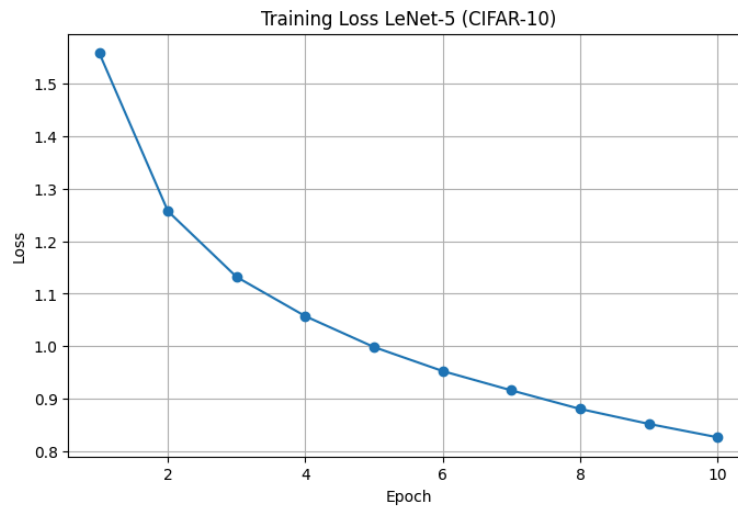
Tabel 1. Akurasi Klasifikasi Global Model CNN

Model	Akurasi Global (%)
LeNet-5	63.73
AlexNet	73.71
ResNet-50	91.56

Dari Tabel 1, terlihat jelas bahwa ResNet-50 mencapai akurasi tertinggi sebesar 91.56%, diikuti oleh AlexNet dengan 73.71%, dan LeNet-5 dengan 63.73%. Hasil ini menunjukkan bahwa arsitektur yang lebih dalam dan kompleks, seperti ResNet-50, mampu mengekstraksi fitur yang lebih *robust* dan performa klasifikasi yang lebih unggul pada dataset CIFAR-10.

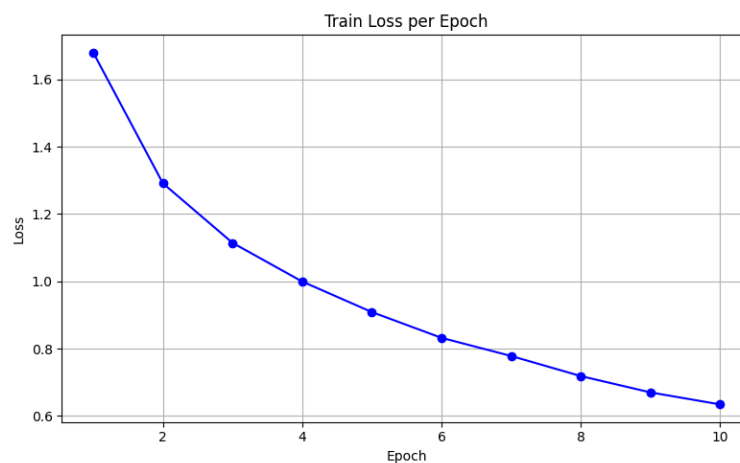
3.2. Nilai *Loss* Pelatihan

Visualisasi *loss* pelatihan untuk masing-masing model disajikan pada Gambar 2, 3, dan 4. Penurunan *loss* menunjukkan kemampuan model untuk belajar dari data pelatihan dan mengoptimalkan bobotnya.



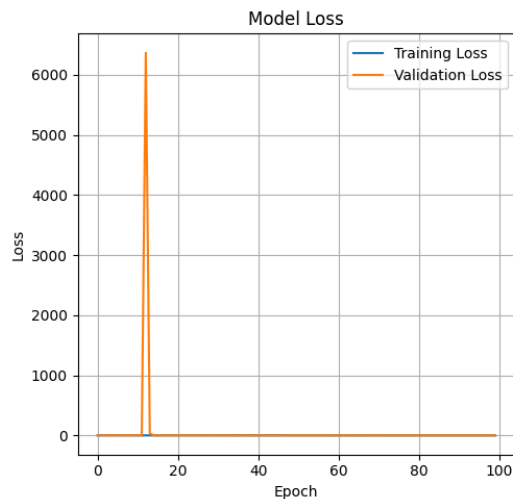
Gambar 2. Training Loss LeNet-5

- **LeNet-5:** *Loss* pelatihan LeNet-5 menurun dari 1.5579 pada *epoch* pertama menjadi 0.8267 pada *epoch* ke-10 (Gambar 2). Penurunan ini menunjukkan proses pembelajaran yang stabil, meskipun masih ada potensi untuk peningkatan akurasi jika dilatih lebih lama atau dengan *hyperparameter* yang berbeda.



Gambar 3. Training Loss AlexNet

- **AlexNet:** AlexNet menunjukkan penurunan *loss* yang lebih cepat dan mencapai nilai *loss* akhir yang lebih rendah (0.6340) dibandingkan LeNet-5 dalam 10 *epoch* (Gambar 3). Hal ini mengindikasikan efisiensi AlexNet dalam mempelajari pola data berkat arsitekturnya yang lebih dalam dan penggunaan fungsi aktivasi ReLU serta *dropout*.



Gambar 4. Model Loss ResNet-5

- **ResNet-50:** ResNet-50, yang dilatih selama 100 *epoch* dengan *learning rate scheduler* dan *early stopping*, menunjukkan konvergensi *loss* yang sangat baik. *Loss* validasi berhasil mencapai 0.7196. Kurva *loss* pada Gambar 4 menunjukkan penurunan yang stabil dan konvergen, dengan *gap* yang relatif kecil antara *training loss* dan *validation loss*, mengindikasikan kemampuan *generalisasi* yang baik.

3.3. Akurasi per Kelas

Analisis akurasi per kelas memberikan wawasan lebih dalam mengenai performa model pada setiap kategori objek.

```
Akurasi untuk kelas plane: 67.50%
Akurasi untuk kelas car : 80.50%
Akurasi untuk kelas bird : 46.40%
Akurasi untuk kelas cat : 45.90%
Akurasi untuk kelas deer : 53.80%
Akurasi untuk kelas dog : 55.40%
Akurasi untuk kelas frog : 75.10%
Akurasi untuk kelas horse: 69.60%
Akurasi untuk kelas ship : 77.10%
Akurasi untuk kelas truck: 66.00%
```

Gambar 5. Akurasi per Kelas untuk LeNet-5

- **LeNet-5:** Pada gambar 5, LeNet-5 memiliki akurasi bervariasi per kelas, dengan 'car' sebagai kelas dengan akurasi tertinggi (80.50%) dan 'cat' sebagai kelas terendah (45.90%). Model ini menunjukkan kesulitan dalam membedakan kelas yang memiliki kemiripan visual tinggi, seperti 'cat' dan 'dog'.

Akurasi per kelas:	
airplane	: 78.70%
automobile	: 82.50%
bird	: 68.40%
cat	: 56.90%
deer	: 66.00%
dog	: 66.30%
frog	: 72.20%
horse	: 78.80%
ship	: 84.30%
truck	: 83.00%

Gambar 6. Akurasi per Kelas untuk AlexNet

- **AlexNet:** Pada gambar 6, AlexNet menunjukkan peningkatan akurasi yang signifikan di sebagian besar kelas dibandingkan LeNet-5. Akurasi tertinggi dicapai pada kelas 'ship' (84.30%) dan 'automobile' (82.50%), sementara kelas 'cat' (56.90%) tetap menjadi yang terendah. Ini mengindikasikan bahwa AlexNet lebih baik dalam menangani keragaman visual, namun tantangan pada kelas 'cat' masih ada.

Per-class Accuracy:	
airplane	: 0.9190 (91.90%)
automobile	: 0.9730 (97.30%)
bird	: 0.8960 (89.60%)
cat	: 0.8020 (80.20%)
deer	: 0.9190 (91.90%)
dog	: 0.8340 (83.40%)
frog	: 0.9690 (96.90%)
horse	: 0.9460 (94.60%)
ship	: 0.9400 (94.00%)
truck	: 0.9580 (95.80%)

Gambar 7. Akurasi per Kelas untuk ResNet-50

- **ResNet-50:** Pada gambar 7, ResNet-50 secara konsisten menunjukkan akurasi yang sangat tinggi di semua kelas. Akurasi tertinggi mencapai 97.30% untuk kelas 'automobile', dan akurasi terendah untuk kelas 'cat' meningkat drastis menjadi 80.20%. Performa yang merata dan tinggi ini menyoroti kapabilitas ResNet-50 dalam mempelajari representasi fitur yang sangat diskriminatif, bahkan untuk kelas yang sulit dibedakan.

3.4. Confusion Matrix

Confusion matrix untuk AlexNet dan ResNet-50 memberikan detail mengenai *True Positives* (TP), *False Positives* (FP), *False Negatives* (FN), dan *True Negatives* (TN) per kelas.

- **AlexNet:** Pada AlexNet, terlihat bahwa kelas 'cat' memiliki TP terendah (569) dan FN tertinggi (431), serta FP yang signifikan (517). Ini berarti model sering salah mengklasifikasikan gambar 'cat' sebagai kelas lain (FP) dan gagal mengenali gambar 'cat' yang sebenarnya (FN).
- **ResNet-50:** *Confusion matrix* ResNet-50 menunjukkan nilai TP yang sangat tinggi dan nilai FP/FN yang jauh lebih rendah di semua kelas. Khususnya untuk kelas 'cat', meskipun masih menjadi yang terendah, nilai TP-nya (802) jauh lebih tinggi dibandingkan AlexNet, dan nilai FP (198) serta FN (198) jauh berkurang, menunjukkan peningkatan performa yang drastis.

Tabel 2. TP, FP, FN, TN per kelas AlexNet

Kelas	TP	FP	FN	TN
airplane	787	300	213	8700
automobile	825	87	175	8913
bird	684	478	316	8522
cat	569	517	431	8483
deer	660	256	340	8744
dog	663	374	337	8626
frog	722	116	278	8884
horse	788	216	212	8784
ship	843	157	157	8843
truck	830	128	170	8872

Tabel 3. TP, FP, FN, TN per kelas ResNet-50

Kelas	TP	FP	FN	TN
airplane	919	81	81	8919
automobile	973	27	27	8973
bird	896	104	104	8896
cat	802	198	198	8802
deer	919	81	81	8919
dog	834	166	166	8834
frog	969	31	31	8969
horse	946	54	54	8946
ship	940	60	60	8940
truck	958	42	42	8958

3.5 Visualisasi Contoh Prediksi Pada AlexNet dan ResNet-50

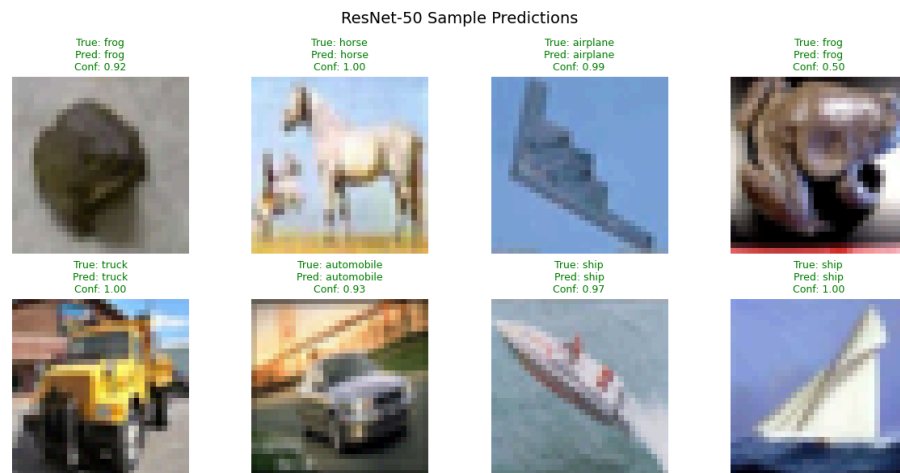
Untuk memberikan pemahaman kualitatif mengenai kinerja model, beberapa contoh gambar dari *test set* ditampilkan bersama dengan prediksi model dan label sebenarnya. Gambar 8 menyajikan contoh prediksi dari AlexNet, dan Gambar 9 menampilkan contoh prediksi dari ResNet-50.

1. **AlexNet:** Pada contoh prediksi AlexNet, terlihat bahwa model mampu mengklasifikasikan beberapa objek dengan benar, seperti "ship" dan "frog". Namun, terdapat satu kasus misklasifikasi yang signifikan, yaitu gambar kucing ("T: cat") yang diprediksi sebagai anjing ("P: dog"). Kesalahan ini konsisten dengan temuan akurasi per kelas AlexNet, di mana model masih kesulitan dalam membedakan objek dengan kemiripan visual yang tinggi, seperti kucing dan anjing.



Gambar 8. Visualisasi Prediksi AlexNet

2. **ResNet-50:** Visualisasi prediksi ResNet-50 menunjukkan performa yang sangat impresif. Dari contoh yang ditampilkan, ResNet-50 berhasil mengklasifikasikan semua gambar dengan tepat. Selain itu, model juga memberikan nilai *confidence* (Conf) yang tinggi untuk sebagian besar prediksinya, menunjukkan tingkat kepastian yang tinggi. Misalnya, gambar kuda diprediksi sebagai "horse" dengan *confidence* 1.00, dan gambar truk diprediksi sebagai "truck" dengan *confidence* 1.00. Bahkan untuk gambar katak yang mungkin terlihat kurang jelas, model berhasil memprediksi sebagai "frog" dengan *confidence* 0.92 dan 0.50. Performa akurat ini mendukung temuan akurasi global ResNet-50 yang superior, menyoroti kemampuannya dalam mengekstraksi fitur kompleks dan membuat prediksi yang andal.



Gambar 9. Visualisasi Prediksi ResNet-50

4. CONCLUSION

Penelitian ini telah berhasil membandingkan kinerja tiga arsitektur Convolutional Neural Network (CNN) klasik dan modern—LeNet-5, AlexNet, dan ResNet-50—dalam tugas klasifikasi citra menggunakan dataset CIFAR-10. Hasil penelitian secara empiris menunjukkan bahwa kompleksitas arsitektur CNN berkorelasi positif dengan akurasi klasifikasi pada dataset dengan keragaman visual menengah seperti CIFAR-10. ResNet-50 muncul sebagai model dengan performa terbaik, mencapai akurasi global optimal sebesar 91.56%. Keunggulan ini dapat diatribusikan pada arsitektur yang sangat dalam dengan *residual connections* dan penerapan teknik optimasi lanjutan seperti *data augmentation* yang ekstensif, *cosine annealing scheduler*, *label smoothing*, *weight decay*, dan *cutout augmentation*. Akurasi tertinggi ResNet-50 tercatat pada kelas 'automobile' (97.30%), dan secara signifikan meningkatkan performa pada kelas yang menantang seperti 'cat' (80.20%). Di sisi lain, AlexNet menunjukkan peningkatan akurasi yang substansial (73.71%) dibandingkan dengan LeNet-5 (63.73%). AlexNet berhasil memvalidasi efektivitas penggunaan fungsi aktivasi ReLU dan *dropout* untuk menangani *overfitting*. Sementara itu, LeNet-5, meskipun sederhana dan efisien, menunjukkan keterbatasan dalam menangani kompleksitas dataset CIFAR-10, terutama pada kelas dengan variasi tinggi seperti 'cat' (45.90%). Berdasarkan temuan ini, dapat disimpulkan bahwa untuk kasus klasifikasi citra pada dataset yang memiliki tingkat kompleksitas visual menengah, model dengan arsitektur yang lebih dalam dan strategi pelatihan yang lebih canggih, seperti ResNet-50, memberikan hasil yang jauh lebih akurat. Namun, perlu dipertimbangkan *trade-off* antara akurasi dan sumber daya komputasi yang dibutuhkan. ResNet-50 memiliki jumlah parameter yang jauh lebih besar (24.634.250 parameter) dan waktu pelatihan yang lebih lama (2.51 jam) dibandingkan LeNet-5 dan AlexNet.

Untuk penelitian selanjutnya, disarankan untuk:

- Menguji model dengan dataset yang lebih besar dan lebih bervariasi untuk mengeksplorasi kemampuan generalisasi model secara lebih luas.
- Menerapkan kombinasi teknik *ensemble learning* pada model-model CNN untuk melihat apakah akurasi dapat ditingkatkan lebih lanjut.
- Menjelajahi arsitektur *deep learning* lain yang lebih mutakhir atau teknik *transfer learning* untuk kasus-kasus spesifik yang memerlukan akurasi sangat tinggi dengan data terbatas.
- Melakukan analisis efisiensi model (misalnya, jumlah FLOPs dan ukuran model) secara lebih rinci untuk perbandingan yang lebih holistik.

DAFTAR PUSTAKA

- [1] Vincent, G., Darian, G., & Surantha, N. (2025). *Performance Evaluation of CNN for Skin Cancer Detection on Edge Computing Devices*. BINUS Graduate Program.
- [2] Muda, M., Abd Aziz, A. Z., & Wattanapanich, C. (2025). *Computer Vision-Based Approach Using Deep Learning for Breast Cancer Rehabilitation Evaluation: A Comparative Performance of CNN and RNN*. Universiti Sultan Zainal Abidin.
- [3] Li, C., Meng, Y., Li, J., Liu, Y., & Lv, C. (2025). *Digital Meter Reading Recognition Method Based on Improved LeNet-5*. Tianjin University of Technology.

-
- [4] Etaba, O. P., Abubakar, J. J. A., Ishaya, G. A., & Enyetere, A. (2025). *Adapting Deep-Learning in Early Yam Disease Detection Using LeNet-5*. International Journal of Current Science Research and Review.
 - [5] Risman Hadi, R., Dwiyanaputra, R., & Irfan, P. (2025). *Implementasi LeNet-5 dan MobileNet-V2 untuk Klasifikasi Kematangan Buah Cabai*. JATI – Jurnal Mahasiswa Teknik Informatika.
 - [6] Agustin, I. R., Wahana, A., & Atmadja, A. R. (2025). *Klasifikasi Irama Murottal Al-Quran Menggunakan Metode CNN dengan Perbandingan Arsitektur ResNet50 dan VGG16*. Journal of Information System Research (JOSH), 6(2), 914–922.
 - [7] Setiawan, A. W. (2025). *Perbandingan Kinerja Arsitektur CNN pada Deteksi Malaria Menggunakan Citra Sel Darah*. Jurnal Teknologi Informasi dan Ilmu Komputer, 11(3), 643–652.
 - [8] Fahri, M., Yanto, F., Syafria, F., & Abdillah, R. (2025). *Perbandingan Akurasi Arsitektur EfficientNet-B0, VGG16, dan Inception V3 dalam Deteksi Tumor Ginjal pada Citra CT-Scan*. Bulletin of Computer Science Research, 5(4), 696–705.
 - [9] Pradana, C. G., Vitria, A. H. P., & Fahrudin, D. M. (2025). *Perbandingan Pekerjaan Arsitektur antara RAB Proyek dengan BIM Cubicost pada Gedung Laboratorium ISI Surakarta*. Universitas Sebelas Maret.
 - [10] Lubis, M. F. F., & Susilawati. (2025). *Pengenalan Tulisan Tangan Angka Menggunakan CNN dengan Arsitektur DenseNet-201 pada Dataset MNIST*. Universitas Medan Area.
-