

Pengaruh Augmentasi Data Terhadap Akurasi Pelatihan Model CNN untuk Klasifikasi Jenis Ikan

Muhammad Abel Al-Fahrezi[#]

[#] Jurusan Teknologi Informasi, Politenik Negeri Padang, Limau Manis, Padang, 25164, Indonesia
E-mail: [abelalfahrezi11\[at\]gmail.com](mailto:abelalfahrezi11[at]gmail.com)

ABSTRACTS

Sustainability of marine resources and management of aquatic ecosystems depend on accurate fish classification. CNNs have proven successful in image classification tasks; however, they often face the problem of limited data variation. The purpose of this study was to examine how data augmentation affects the training accuracy of CNN models for fish species classification. Two scenarios were studied: the first scenario involved training without data augmentation, and the second scenario involved training with data augmentation. In both scenarios, a custom CNN architecture for ten epochs was used. Experimental results showed that using data augmentation with the configuration used actually caused the model performance to deteriorate. Loss values on both datasets increased, with training accuracy dropping from 76.08% to 63.81%, and validation accuracy also dropping from 91.13% to 84.55%. Overly aggressive augmentation parameters or insufficient training time for the introduced data variation could have caused this decline. Interestingly, validation accuracy was consistently higher than training accuracy in both situations, indicating that certain datasets have specific features. This study emphasizes the importance of carefully optimizing augmentation parameters and training duration to maximize the benefits of data augmentation in image classification.

Manuscript received Jun 02, 2025;
revised Jun 7, 2025. accepted Jun
9, 2025 Date of publication Jun
30, 2025. International Journal,
JITSI : Jurnal Ilmiah Teknologi
Sistem Informasi licensed under a
Creative Commons Attribution-
Share Alike 4.0 International
License



ABSTRAK

Keberlanjutan sumber daya laut dan manajemen ekosistem perairan bergantung pada klasifikasi ikan yang tepat. CNN telah terbukti berhasil dalam tugas klasifikasi gambar; namun, ia sering menghadapi masalah keterbatasan variasi data. Tujuan penelitian ini adalah untuk melihat bagaimana peningkatan data mempengaruhi akurasi pelatihan model CNN untuk klasifikasi jenis ikan. Dipelajari dua skenario: skenario pertama melibatkan pelatihan tanpa augmentasi data, dan skenario kedua melibatkan pelatihan dengan augmentasi data. Dalam kedua skenario, arsitektur CNN khusus selama sepuluh epoch digunakan. Hasil eksperimen menunjukkan bahwa menggunakan augmentasi data dengan konfigurasi yang digunakan justru menyebabkan kinerja model menjadi lebih buruk. Nilai loss pada kedua set data meningkat, dengan akurasi pelatihan turun dari 76,08% menjadi 63,81%, dan akurasi validasi juga turun dari 91,13% menjadi 84,55%. Parameter augmentasi yang terlalu agresif atau waktu pelatihan yang tidak memadai untuk variasi data yang diperkenalkan dapat menyebabkan penurunan ini. Unikanya, akurasi validasi secara konsisten lebih tinggi dari akurasi pelatihan dalam kedua situasi, menunjukkan bahwa dataset tertentu memiliki fitur tertentu. Studi ini menekankan betapa pentingnya mengoptimalkan parameter augmentasi dan durasi pelatihan dengan hati-hati untuk memaksimalkan keuntungan dari augmentasi data dalam klasifikasi gambar.

Keywords / Kata Kunci — *Convolutional Neural Network (CNN); Klasifikasi Ikan; Augmentasi Data; Akurasi*

CORRESPONDING AUTHOR

Muhammad Abel Al-Fahrezi
Jurusan Teknologi Informasi, Politenik Negeri Padang, Limau Manis, Padang, 25164, Indonesia
Email: abelfahrezi11[at]gmail.com

1. PENDAHULUAN

Untuk menjaga kelestarian ekosistem perairan dan mendukung keberlanjutan sumber daya laut, klasifikasi jenis ikan sangat penting. Identifikasi spesies ikan yang akurat memungkinkan pemantauan populasi, menjaga kesehatan ekosistem, dan mendukung pengelolaan perikanan yang berkelanjutan. Selain itu, klasifikasi ikan juga membantu ilmu taksonomi dan pelestarian keanekaragaman hayati, yang merupakan dasar konservasi laut dan sumber daya laut [1].

Teknologi komputer vision dan machine learning, terutama Convolutional Neural Network (CNN), telah membantu mengatasi keterbatasan metode klasifikasi ikan secara manual yang lama dan rentan kesalahan. CNN banyak digunakan dalam pengolahan gambar untuk klasifikasi objek, termasuk jenis ikan, karena dapat secara otomatis dan efisien mengekstraksi fitur penting dari gambar. Penelitian sebelumnya, seperti yang dilakukan oleh Nizar Abdurrahman et al. (2023), telah menunjukkan potensi CNN dalam klasifikasi ikan, bahkan mampu mencapai akurasi hingga 88% dibandingkan metode tradisional seperti k-NN [2].

Namun, keterbatasan jumlah dan variasi data merupakan masalah utama dalam pelatihan model CNN untuk klasifikasi gambar ikan. Kualitas dan kuantitas data ikan yang tersedia seringkali terbatas, seperti variasi posisi, pencahayaan, dan oklusi, yang dapat mengurangi kinerja model. Untuk mengatasi hal ini, metode augmentasi data digunakan untuk meningkatkan variasi data pelatihan tanpa perlu mengumpulkan data baru secara menyeluruh. Melalui augmentasi data, model CNN diharapkan dapat belajar dari berbagai variasi citra ikan, termasuk perbedaan posisi, pencahayaan, dan oklusi, sehingga secara signifikan meningkatkan akurasi klasifikasi jenis ikan [3],[4].

Tujuan penelitian ini adalah untuk mengevaluasi bagaimana akurasi pelatihan model CNN dalam klasifikasi jenis ikan dipengaruhi oleh augmentasi. Dengan memfokuskan pada berbagai metode augmentasi data, penelitian ini diharapkan dapat membantu meningkatkan kinerja model klasifikasi gambar ikan dan mengatasi keterbatasan data yang sering terjadi dalam aplikasi nyata

2. METODOLOGI PENELITIAN

Penelitian ini mengadopsi pendekatan eksperimental kuantitatif untuk menganalisis pengaruh augmentasi data terhadap akurasi pelatihan model Convolutional Neural Network (CNN) dalam klasifikasi jenis ikan. Dalam desain penelitian, performa model dibandingkan dengan dua situasi utama: pelatihan tanpa augmentasi data dan pelatihan dengan menggunakan metode augmentasi data. Pengumpulan dan pra-pemrosesan dataset gambar ikan, pengembangan dan konfigurasi arsitektur model CNN, penerapan berbagai strategi augmentasi data, dan evaluasi komparatif akurasi pelatihan untuk masing-masing skenario adalah tujuan utama penelitian. Proses penelitian secara keseluruhan dirancang untuk memastikan validitas dan objektivitas hasil yang diperoleh, sehingga dapat memberikan pemahaman yang jelas tentang kontribusi yang diberikan augmentasi data.

2.1. Sumber Data

Kumpulan data gambar ikan, yang disebut "Dataset Ikan", dapat diakses secara publik melalui platform Kaggle: <https://www.kaggle.com/datasets/markdaniellampa/fish-dataset/>. Dataset ini dipilih karena relevansinya untuk tujuan klasifikasi jenis ikan dan karena mudah diakses, yang memungkinkan replikasi penelitian. Dataset ini, secara khusus, terdiri dari 13.304 foto dari 31 jenis ikan yang berbeda. Dataset biasanya memiliki setiap gambar dalam format JPG dengan resolusi yang berbeda, tetapi masing-masing menampilkan detail visual yang cukup untuk memudahkan identifikasi. Nama spesies ikan yang digunakan pada eksperimen ini seperti pada Tabel 1.

2.2. Pra-pemrosesan Data

Sebelum dimasukkan ke dalam model Convolutional Neural Network (CNN), tahap pra-pemrosesan data sangat penting untuk memastikan kualitas dan konsistensi citra. Proses ini bertujuan untuk mengoptimalkan karakteristik gambar dan format input untuk meningkatkan efisiensi dan akurasi pelatihan model.

Langkah-langkah pra-pemrosesan yang diterapkan adalah sebagai berikut:

2.2.1. Pengubahan Ukuran Citra(Resizing)

Setiap gambar dalam kumpulan data ikan memiliki resolusi awal yang berbeda. Setiap gambar diubah ukurannya menjadi 128 x 128 piksel. Ini dilakukan untuk menyeragamkan dimensi input dan memenuhi persyaratan arsitektur model CNN yang akan digunakan. Praktik pra-pemrosesan ini sangat esensial untuk mempersiapkan data agar model CNN dapat belajar secara efektif dan efisien [5],[6].

2.2.2. Normalisasi Nilai Piksel

Nilai piksel asli gambar berada dalam rentang 0–255. Mereka dinormalisasi untuk membantu proses konvergensi model selama pelatihan dan mencegah masalah vanishing atau exploding gradients. Ini dilakukan dengan membagi setiap nilai piksel dengan 255.0, sehingga rentangnya berubah menjadi 0.0–1.0. Teknik normalisasi ini biasa digunakan dalam pengolahan gambar untuk *deep learning* [7].

2.3. Augmentasi Data

Salah satu teknik penting dalam pelatihan Convolutional Neural Network (CNN) adalah augmentasi data. Teknik ini membantu CNN mengatasi masalah jumlah dan variasi data pelatihan yang terbatas. Augmentasi data meningkatkan kekuatan model, mencegah overfitting, dan meningkatkan kemampuan generalisasi model terhadap data yang belum pernah dilihat sebelumnya dengan menghasilkan variasi baru dari gambar yang sudah ada [8], [9]. Set pelatihan (training set) dari *Fish Dataset* digunakan untuk augmentasi data dalam penelitian ini.

Teknik-teknik augmentasi data yang diimplementasikan meliputi transformasi geometris dan perubahan intensitas piksel, seperti pada Tabel 2.

TABEL 1. Nama Spesies Ikan

Class	Jenis Ikan
0	<i>Bangus</i>
1	<i>Big Head Carp</i>
2	<i>Black Spotted Barb</i>
3	<i>Catfish</i>
4	<i>Climbing Perch</i>
5	<i>Fourfinger Threadfin</i>
6	<i>Freshwater Eel</i>
7	<i>Glass Perchlet</i>
8	<i>Goby</i>
9	<i>Gold Fish</i>
10	<i>Gourami</i>
11	<i>Grass Carp</i>
12	<i>Green Spotted Puffer</i>
13	<i>Indian Carp</i>
14	<i>Indo-Pacific Tarpon</i>
15	<i>Jaguar Gapote</i>
16	<i>Janitor Fish</i>
17	<i>Knife fish</i>
18	<i>Long-Snouted Pipefish</i>
19	<i>Mosquito Fish</i>
20	<i>Mudfish</i>
21	<i>Mullet</i>
22	<i>Pangasius</i>
23	<i>Perch</i>
24	<i>Scat Fish</i>
25	<i>Silver Barb</i>
26	<i>Silver Carp</i>
27	<i>Silver Perch</i>
28	<i>Snakehead</i>
29	<i>Tenpounder</i>
30	<i>Tilapia</i>

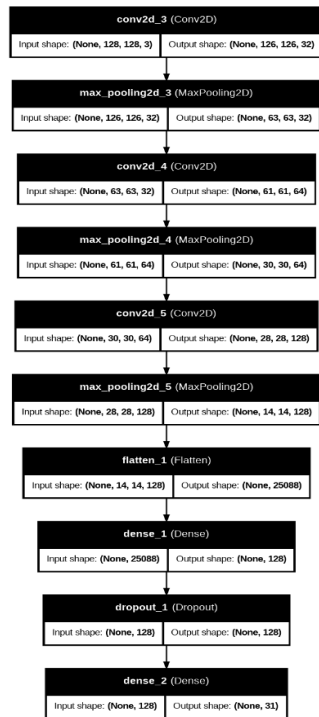
TABEL 2. Teknik Augmentasi Data yang Digunakan

Teknik Augmentasi	Deskripsi	Parameter
Rotasi (Rotation)	Mengubah orientasi citra untuk membantu model mengenali objek tanpa terpengaruh sudut pandang.	15°
Pergeseran Lebar & Tinggi (Width/Height Shift)	Menggeser citra secara horizontal atau vertikal untuk mensimulasikan variasi posisi objek dalam <i>frame</i> .	0.1
Pergeseran Kecerahan (Brightness Shift)	Mengubah intensitas cahaya pada citra untuk meningkatkan toleransi model terhadap variasi kondisi pencahayaan.	0.8-1.2
Zoom Acak (Random Zoom)	Memperbesar atau memperkecil citra secara acak untuk melatih model mengenai objek pada skala yang berbeda.	0.8-1.2
Horizontal Flip	Membalik citra secara horizontal, efektif untuk objek yang simetris atau dapat memiliki orientasi kiri/kanan.	True

Selama proses pelatihan, implementasi augmentasi data ini dilakukan secara *on-the-fly*. Artinya, citra yang di-augmentasi tidak disimpan secara fisik; sebaliknya, mereka dibuat secara dinamis setiap periode pelatihan. Metode ini memastikan bahwa model selalu melihat variasi data yang sedikit berbeda di setiap epoch. Ini memungkinkan untuk memaksimalkan manfaat dari augmentasi data tanpa menghabiskan banyak penyimpanan.

2.4. Arsitektur Model Convolutional Neural Network (CNN)

Model Convolutional Neural Network (CNN) yang digunakan dalam penelitian ini dirancang khusus untuk mengklasifikasikan berbagai jenis ikan. Arsitektur ini dirancang khusus untuk mengekstraksi fitur hierarkis dari citra input berukuran 128 x 128 piksel dengan tiga kanal warna. Arsitektur ini dirancang untuk melakukan klasifikasi ke dalam 31 kategori jenis ikan yang berbeda. Struktur model terdiri dari beberapa lapisan konvolusi dan pooling yang diikuti oleh lapisan *fully connected*, seperti yang ditunjukkan pada Gambar 1, yang merupakan diagram contoh arsitektur model.



GAMBAR 1. Arsitektur CNN

2.5. Konfigurasi Pelatihan Model

Setelah tahap pra-pemrosesan data dan definisi arsitektur model CNN, langkah selanjutnya adalah konfigurasi dan prosedur pelatihan model. Tahap ini sangat penting untuk memastikan bahwa model dapat belajar dari data pelatihan dan mengoptimalkan kinerjanya dalam klasifikasi jenis ikan. Pelatihan dilakukan menggunakan framework TensorFlow dan Keras di lingkungan Google Colab.

Detail konfigurasi parameter pelatihan model disajikan pada Tabel 3.

2.6. Prosedur Eksperimen

Prosedur eksperimen dalam penelitian ini dirancang untuk menganalisis secara komparatif bagaimana augmentasi data mempengaruhi akurasi pelatihan model Convolutional Neural Network (CNN) untuk klasifikasi jenis ikan. Eksperimen dilakukan dalam dua skenario utama yang berbeda, tetapi dengan parameter model dan pelatihan yang sama, kecuali untuk elemen augmentasi data. Perbandingan skenario ditunjukkan dalam Tabel 4.

TABEL 3. Konfigurasi Parameter Pelatihan Model CNN

Parameter Pelatihan	Detail Konfigurasi	Keterangan
Fungsi Kerugian (Loss Function)	Categorical Cross-Entropy	Mengukur perbedaan antara distribusi probabilitas yang diprediksi model dan distribusi probabilitas kelas sebenarnya.
Optimizer	Adam (Adaptive Moment Estimation)	Algoritma optimasi yang mengadaptasi <i>learning rate</i> untuk setiap parameter model, mendorong konvergensi cepat dan stabil [10].
Learning Rate	0.001 (default)	Mengontrol seberapa besar langkah yang diambil optimizer saat memperbarui bobot model [11].
Jumlah Epoch	10	Jumlah siklus lengkap di mana seluruh dataset pelatihan dilewatkan melalui model.
Batch Size	32	Jumlah sampel pelatihan yang diproses dalam satu iterasi sebelum bobot model diperbarui.
Metrik Evaluasi	Accuracy	Proporsi prediksi yang benar dari total prediksi, digunakan untuk memonitor performa model.

TABEL 4. Perbandingan Prosedur Eksperimen Skenario Pelatihan

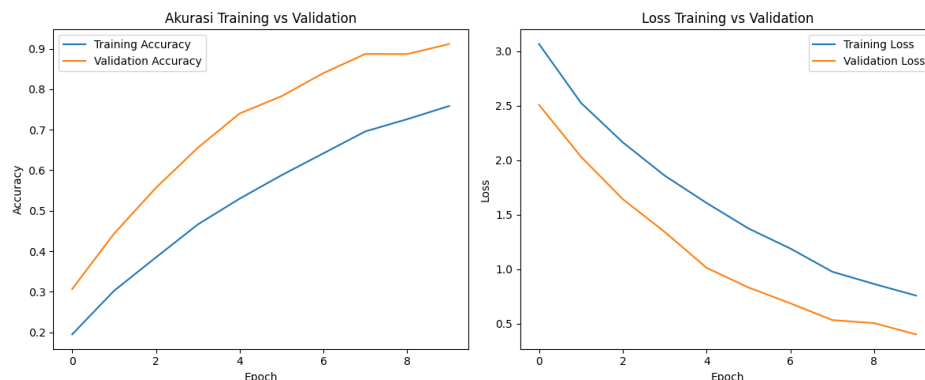
Fitur Prosedur	Skenario 1 : Tanpa Augmentasi Data	Skenario 2 : Dengan Augmentasi Data
Dataset Pelatihan	Citra asli dari <i>Fish Dataset</i>	Citra <i>Fish Dataset</i> dengan augmentasi <i>on-the-fly</i>
Arsitektur Model	Identik (Mengacu ke sub-bab 2.4)	Identik (Mengacu ke sub-bab 2.4)
Konfigurasi Pelatihan	Identik (Mengacu ke sub-bab 2.5)	Identik (Mengacu ke sub-bab 2.5)
Penerapan Augmentasi	Tidak diterapkan	Diterapkan (teknik dan parameter di sub-bab 2.3)
Metrik Utama Monitor	Akurasi Pelatihan, Loss	Akurasi Pelatihan, Loss
Tujuan Skenario	Membangun <i>baseline</i> performa model	Mengukur peningkatan performa model dengan augmentasi

3. HASIL DAN PEMBAHASAN

Hasil dari berbagai eksperimen yang telah disajikan dalam penelitian ini menunjukkan pengaruh augmentasi data terhadap akurasi pelatihan model Convolutional Neural Network (CNN) dalam mengklasifikasikan jenis ikan. Dua skenario utama—Skenario 1 yang melibatkan pelatihan model tanpa augmentasi data dan Skenario 2 yang melibatkan pelatihan model dengan augmentasi data—dilaksanakan sesuai dengan prosedur yang diuraikan pada sub-bab 2.6. Hasil yang disajikan termasuk metrik performa utama seperti akurasi dan *loss*, serta analisis komparatif untuk mengevaluasi bagaimana peningkatan kapabilitas model dipengaruhi oleh augmentasi data. Hasil-hasil ini akan dibahas lebih lanjut mengenai tujuan penelitian.

3.1. Hasil Pelatihan Model Tanpa Augmentasi Data (Skenario 1)

Dalam skenario pertama penelitian ini, model Convolutional Neural Network (CNN) dilatih tanpa menggunakan metode augmentasi data pada set pelatihan. Tujuan skenario ini adalah untuk menetapkan baseline performa model untuk situasi di mana hanya menggunakan citra *Fish Dataset* yang asli. Pelatihan dilakukan dalam sepuluh epoch.

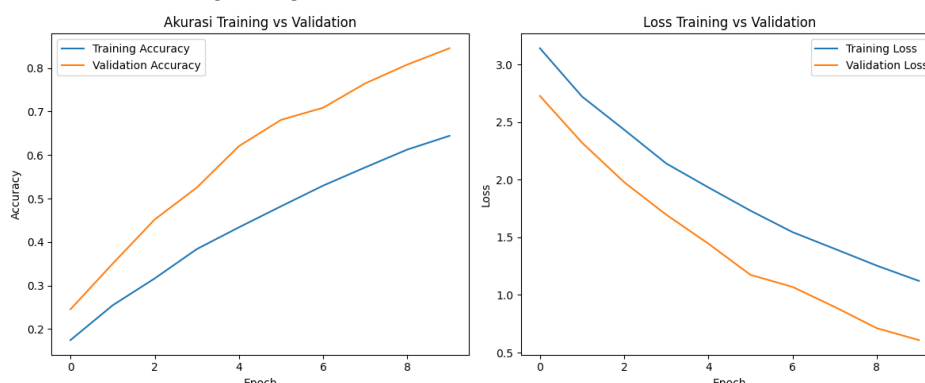


GAMBAR 2. Plot Akurasi dan Loss per Epoch Skenario 1

Gambar 2 menunjukkan bagaimana *loss* pelatihan dan *accuracy* pelatihan model berkembang selama pelatihan, serta metrik validasi. *Loss* pelatihan (garis biru pada plot *loss*) menurun secara konsisten dari nilai awal di atas 3.0 hingga mencapai sekitar 0.75 pada epoch ke-9. Sebaliknya, akurasi pelatihan (garis biru pada plot akurasi) meningkat secara konsisten, meningkat dari sekitar 15% pada epoch 0 hingga 73,09 % pada epoch 9.

Sangat menarik bahwa kinerja set validasi menunjukkan tren yang lebih baik. *Loss* validasi (garis oranye pada plot *loss*) selalu berada di bawah *loss* pelatihan dan terus menurun hingga mencapai sekitar 0,40 pada akhir pelatihan. Selain itu, akurasi validasi (garis oranye pada plot akurasi) secara konsisten lebih tinggi dari akurasi pelatihan, mencapai puncak 88,66% pada epoch ke-9. Hasil ini menunjukkan bahwa model memiliki kemampuan dasar untuk mengklasifikasikan jenis ikan dengan data terbatas; namun, terdapat fenomena di mana model bekerja lebih baik dengan set validasi daripada set pelatihan.

3.2. Hasil Pelatihan Model Dengan Augmentasi Data (Skenario 2)



GAMBAR 3. Plot Akurasi dan Loss per Epoch Skenario 2

Dalam skenario kedua penelitian ini, kami melihat bagaimana penggunaan metode augmentasi data berdampak pada set pelatihan model Convolutional Neural Network (CNN). Tujuan skenario ini adalah untuk melihat bagaimana akurasi pelatihan model dan kemampuan generalisasinya dapat dipengaruhi oleh penambahan variasi data melalui augmentasi. Pengaruh ini termasuk dalam klasifikasi jenis ikan. Dengan arsitektur dan konfigurasi

yang sama seperti dalam Skenario 1, model dilatih selama sepuluh epoch. Namun, metode augmentasi diterapkan secara on-the-fly.

Gambar 3 menunjukkan bagaimana *loss* pelatihan dan akurasi pelatihan model berkembang selama pelatihan dengan augmentasi data dan metrik validasi. *Loss* pelatihan (garis biru pada plot *loss*) menunjukkan penurunan yang konsisten dari nilai awal sekitar 3.2 hingga nilai terakhir sekitar 1.1 pada epoch pertama. Sebaliknya, akurasi pelatihan (garis biru pada plot akurasi) menunjukkan peningkatan yang konsisten, dari sekitar 14% pada epoch 1 hingga mencapai 63.81% pada epoch ke-10.

Seperti dalam skenario pertama, kinerja pada set validasi menunjukkan tren yang lebih baik daripada kinerja pelatihan. *Loss* validasi menunjukkan penurunan yang konsisten hingga mencapai sekitar 0.60 pada akhir pelatihan. Sementara itu, akurasi validasi terus meningkat, mencapai 84,55% pada epoch ke-10. Hasil skenario ini memberikan gambaran tentang kinerja model dengan augmentasi, yang akan dibahas secara komparatif dalam subbab berikutnya.

3.3. Analisis Komparatif dan Pembahasan

3.3.1. Perbandingan Akurasi dan Kurva Pelatihan

Untuk menganalisis secara komparatif performa model CNN pada kedua skenario, hasil akhir metrik akurasi dan *loss* pelatihan serta validasi dirangkum dalam Tabel 5.

TABEL 5. Perbandingan Performa Akhir Model CNN pada Skenario 1 dan Skenario 2

Metrik Performa	Skenario 1 : Tanpa Augmentasi Data	Skenario 2 : Dengan Augmentasi Data	Perbedaan (Skenario 2-Skenario 1)
Akurasi Pelatihan	76,08%	63,81%	-12,27%
<i>Loss</i> Pelatihan	0.7496	1.1371	+0.3875
Akurasi Validasi	91,13%	84,55%	-6,58%
<i>Loss</i> Validasi	0.40	0.6084	+0.2084

Skenario 1 (tanpa augmentasi data) memiliki akurasi pelatihan akhir sebesar 76,08%, sedangkan Skenario 2 (dengan augmentasi data) memiliki akurasi pelatihan akhir sebesar 63,81%. Perbandingan ini menunjukkan bahwa ketika augmentasi data diterapkan, akurasi pelatihan menurun 12,27%.

Pada metrik *loss* pelatihan, tren yang serupa juga terlihat. Skenario 1 menunjukkan *loss* yang lebih rendah (0.7496) dibandingkan Skenario 2 (1.1371), dan Skenario 1 juga menunjukkan akurasi validasi yang lebih tinggi (91,13%) dan *loss* validasi yang lebih rendah (0.40) dibandingkan Skenario 2 (84,55% dan 0.6084, masing-masing). Hal ini menunjukkan bahwa menggunakan augmentasi data pada dataset dan konfigurasi tertentu menghasilkan kinerja model yang lebih buruk pada set pelatihan dan validasi.

Gambar 2 dan Gambar 3 menunjukkan lebih jelas perkembangan kurva akurasi dan *loss* pelatihan serta validasi dari kedua skenario. Ini menunjukkan bagaimana kinerja model berubah selama epoch pelatihan untuk kedua kondisi. Dalam skenario 1 dan 2, kurva akurasi model tanpa peningkatan selalu berada di atas kurva akurasi model dengan peningkatan; sebaliknya, kurva penurunan model dengan peningkatan (Skenario 2) selalu berada di atas kurva akurasi model tanpa peningkatan (Skenario 1). Ini menunjukkan secara visual bahwa ketika augmentasi data diterapkan, performa model menurun baik dalam *loss* maupun akurasi.

3.3.2. Pembahasan Pengaruh Augmentasi Data

Analisis perbandingan skenario 1 (tanpa augmentasi data) dan skenario 2 (dengan augmentasi data) memberikan informasi penting tentang pengaruh augmentasi data terhadap akurasi pelatihan model CNN untuk klasifikasi jenis ikan. Ringkasan performa yang ditunjukkan pada Tabel 5 (sub-bab 3.3.1) menunjukkan bahwa penerapan peningkatan data justru menyebabkan penurunan akurasi pelatihan model, dari 76,08% dalam skenario 1 menjadi 63,81% dalam skenario 2. Akurasi validasi juga menurun, dengan model skenario 2 mencapai 84,55% dibandingkan 91,13% dalam skenario 1. Selain itu, nilai *loss* pelatihan dan validasi juga meningkat (meskipun lebih rendah) dalam skenario 2.

Fakta ini berbeda dengan harapan umum bahwa peningkatan data meningkatkan performa dan generalisasi model, terutama dalam kasus keterbatasan data [12-14]. Sebaliknya, peningkatan data tidak menghasilkan peningkatan akurasi pelatihan atau bahkan penurunan.

Beberapa faktor potensial dapat menjelaskan hasil yang tidak terduga ini:

1. Intensitas Augmentasi Data: Parameter augmentasi seperti rotasi $\pm 15^\circ$, shift 0.1, dan zoom 0.8–1.2 mungkin terlalu agresif atau tidak ideal untuk karakteristik khusus dataset citra ikan ini. Transformasi yang berlebihan dapat menyebabkan sampel buatan yang tidak sepenuhnya menunjukkan variasi dunia nyata, atau noise dapat menghambat model dalam mempelajari karakteristik yang penting untuk klasifikasi [15].

2. Jumlah Epoch Pelatihan yang Terbatas: Model mungkin tidak memiliki waktu yang cukup untuk sepenuhnya mengkonvergen dengan hanya sepuluh epoch pelatihan; ini terutama berlaku untuk variasi data yang lebih besar yang dihasilkan oleh augmentasi. Agar model dapat menginternalisasi keragaman baru dan mencapai kinerja terbaiknya, augmentasi data seringkali membutuhkan waktu pelatihan yang lebih lama [12].
3. Karakteristik Set Data dan Anomali Performa Validasi: Hal yang konsisten di kedua situasi (tanpa dan dengan peningkatan) adalah bahwa akurasi validasi secara signifikan lebih tinggi dan *loss* validasi lebih rendah dibandingkan dengan metrik pelatihan. Ini merupakan anomali yang perlu diperhatikan. Ini dapat menunjukkan bahwa ada bias dalam pembagian dataset; misalnya, set validasi mungkin memiliki sampel yang "lebih mudah" atau kurang representatif dari kerumitan data pelatihan. Selain itu, itu juga dapat menunjukkan bahwa ada perbedaan karakteristik intrinsik yang belum sepenuhnya teridentifikasi antara set validasi dan pelatihan. Di masa mendatang, penyelidikan lebih lanjut diperlukan untuk fenomena ini.

Secara keseluruhan, temuan penelitian ini menunjukkan bahwa, meskipun augmentasi data secara teoretis bertujuan untuk meningkatkan kekuatan dan generalisasi model, penerapannya harus dioptimalkan dengan cermat. Ini terutama berlaku untuk parameter augmentasi yang tepat dan durasi pelatihan yang memadai. Untuk klasifikasi jenis ikan dalam dataset ini, peningkatan data dengan konfigurasi yang digunakan belum menunjukkan hasil pelatihan yang lebih baik; ini bahkan menunjukkan bahwa model menghadapi masalah dalam menginternalisasi variasi yang diperkenalkan dalam rentang waktu yang tersedia.

3.4. Keterbatasan Penelitian dan Saran Penelitian Lanjutan

Studi ini menunjukkan bagaimana akurasi pelatihan model CNN untuk klasifikasi jenis ikan dipengaruhi oleh peningkatan data. Namun, penelitian ini memiliki beberapa keterbatasan yang harus diakui, seperti semua studi ilmiah. Keterbatasan ini juga membuka peluang untuk penelitian lebih lanjut di masa depan.

Berikut adalah beberapa keterbatasan yang teridentifikasi dalam penelitian ini:

1. Jumlah *Epoch* yang Terbatas: Model CNN yang dipelajari dalam penelitian ini dilatih hanya selama 10 *epoch*. Proses pelatihan yang singkat ini, terutama ketika menggunakan data yang diperkaya oleh augmentasi, mungkin tidak memungkinkan model untuk mengkonvergen dan menginternalisasi seluruh variasi fitur yang diperkenalkan. Akibatnya, model mungkin tidak memiliki kinerja yang ideal.
2. Optimasi Parameter Augmentasi yang Terbatas: Konfigurasi umum menentukan parameter augmentasi data seperti rotasi, shift, dan zoom. Seperti yang mungkin terjadi pada penelitian ini, augmentasi data yang terlalu agresif atau salah dapat menghasilkan sampel yang tidak realistis dan merugikan proses pembelajaran model.
3. Anomali Validasi dan Karakteristik Dataset: Fenomena yang konsisten di mana *loss* validasi selalu lebih rendah dan akurasi validasi selalu lebih tinggi dibandingkan metrik pelatihan di kedua situasi adalah anomali yang belum sepenuhnya diketahui penyebabnya. Hal ini dapat menunjukkan bias dalam pembagian dataset atau fitur unik pada set validasi yang membuat model "lebih mudah" untuk digunakan.

Berdasarkan keterbatasan di atas, beberapa saran untuk penelitian lanjutan yang dapat dilakukan adalah sebagai berikut:

1. Eksplorasi Durasi Pelatihan yang Lebih Panjang: Untuk memastikan konvergensi terbaik, lakukan pelatihan model dengan lebih banyak epoch, terutama ketika menggunakan augmentasi data. Untuk menghentikan pelatihan saat performa validasi tidak lagi meningkat, juga dapat menggunakan teknik *early stopping*.
2. Optimasi Hyperparameter Augmentasi Data: Optimumkan parameter augmentasi secara menyeluruh. Ini dapat dilakukan dengan menggunakan metode pencarian grid, pencarian acak, atau teknik otomatisasi augmentasi seperti AutoAugment. Tujuannya adalah untuk menemukan kombinasi augmentasi yang paling efisien dan sesuai dengan karakteristik dataset ikan.
3. Investigasi Karakteristik Set Data Mendalam: Analisis distribusi data dan karakteristik citra di set pelatihan, validasi, dan pengujian untuk menemukan bias atau perbedaan yang signifikan yang mungkin memengaruhi kinerja model.
4. Uji dengan Arsitektur Model Berbeda dan Pembelajaran Transfer: Menentukan apakah hasil yang sama tetap muncul atau apakah arsitektur CNN lain, seperti model yang dilatih sebelumnya, seperti ResNet dan EfficientNet, menanggapi peningkatan data dengan cara yang berbeda.
5. Variasi Jenis Augmentasi: Melihat dampak dari teknik augmentasi yang lebih canggih atau spesifik domain (seperti augmentasi berbasis GAN atau augmentasi yang memperhitungkan anatomi ikan).

4. KESIMPULAN

Tujuan penelitian ini adalah untuk melihat bagaimana pengaruh augmentasi data terhadap akurasi pelatihan model Convolutional Neural Network (CNN) untuk klasifikasi jenis ikan. Seperti yang ditunjukkan oleh eksperimen yang membandingkan pelatihan model tanpa augmentasi data (Skenario 1) dan dengan augmentasi data (Skenario 2), dapat disimpulkan bahwa penerapan augmentasi data dengan konfigurasi yang digunakan dalam

penelitian ini justru menyebabkan performa model menurun. Akurasi pelatihan model menurun dari 76,08% menjadi 63,81%, dan akurasi validasi juga menunjukkan penurunan, dari 91,13% menjadi 84,55%. Selain itu, nilai loss pada kedua set data juga meningkat, menunjukkan bahwa pembelajaran model tidak optimal.

Beberapa alasan untuk penurunan kinerja ini termasuk jumlah epoch pelatihan yang terbatas sehingga model belum dapat menginternalisasi variasi data yang lebih besar dari augmentasi dan kemungkinan parameter augmentasi yang terlalu agresif untuk dataset ikan tertentu. Fakta bahwa akurasi validasi lebih besar daripada akurasi pelatihan di kedua situasi juga merupakan fitur dataset yang perlu diperhatikan lebih lanjut. Namun, penelitian ini menunjukkan bahwa untuk memaksimalkan manfaat augmentasi data dalam klasifikasi citra, optimalisasi parameter augmentasi dan durasi pelatihan yang memadai sangat penting

UCAPAN TERIMAKASIH

Penulis mengucapkan puji syukur kehadiran Allah SWT. atas rahmat dan karunia-Nya, sehingga penelitian ini dapat terselesaikan dengan baik. Keberhasilan penyusunan jurnal ini tidak lepas dari dukungan berbagai pihak.

Penulis ingin menyampaikan rasa terima kasih yang setulus-tulusnya kepada Bapak Ir. Rahmat Hidayat, S.T., M.Sc.IT, selaku Dosen Mata Kuliah Metodologi Penelitian, atas bimbingan, arahan, dan ilmu yang telah diberikan selama proses penelitian ini. Bantuan dan motivasi yang diberikan sangat berarti dalam menyelesaikan studi ini.

REFERENSI

- [1] D. Saleky, E. Weremba, and M. A. Welikken, "Kelimpahan Dan Keanekaragaman Jenis Ikan di Perairan Ndalir Kabupaten Merauke, Papua," *Nekt. J. Perikan. dan Ilmu Kelaut.*, vol. 1, no. 2, pp. 33–42, 2021, doi: 10.47767/nektion.v1i2.290.
- [2] N. Abdurrahman, B. Rahmat, and A. N. Sihananto, "Perbandingan Performa Klasifikasi Citra Ikan Menggunakan Metode K-Nearest Neighbor (K-NN) Dan Convolutional Neural Network (CNN)," *J. Sist. Inf. dan Inform.*, vol. 2, no. 2, pp. 84–93, 2023, doi: 10.33379/jusifor.v2i2.3728.
- [3] A. Azis, "Identifikasi Jenis Ikan Menggunakan Model Hybrid Deep Learning Dan Algoritma Klasifikasi," *Sebatik*, vol. 24, no. 2, pp. 201–206, 2020, doi: 10.46984/sebatik.v24i2.1057.
- [4] S. S. Asmoro, M. F. Amrulloh, M. A. Toybah, and M. A. Saputra, "Rancang Bangun Aplikasi Mobile Untuk Klasifikasi Jenis Ikan Koi Menggunakan Algoritma Convolutional Neural Network," *Semin. Nas. Teknol. Sains*, vol. 3, no. 1, pp. 270–277, 2024, doi: 10.29407/stains.v3i1.4312.
- [5] L. Vanneschi and M. Castelli, *Multilayer perceptrons*, vol. 1–3. 2018. doi: 10.1016/B978-0-12-809633-8.20339-7.
- [6] M. Hashemi, "Enlarging smaller images before inputting into convolutional neural network: zero-padding vs. interpolation," *J. Big Data*, vol. 6, no. 1, 2019, doi: 10.1186/s40537-019-0263-7.
- [7] A. Performance, E. Alshdaifat, D. Alshdaifat, A. Alsarhan, F. Hussein, and S. Moh, "The Effect of Preprocessing Techniques , Applied to Numeric," *Data*, vol. 6, no. 11, 2021.
- [8] M. F. Gunardi, "Implementasi Augmentasi Citra pada Suatu Dataset," *J. Inform.*, vol. 9, no. 1, pp. 1–5, 2023.
- [9] S. Calderon-ramirez et al., "Correcting data imbalance for semi-supervised COVID-19 detection using X-ray chest images," no. January, 2020.
- [10] A. H. Khan, X. Cao, S. Li, V. N. Katsikis, and L. Liao, "BAS-ADAM: An ADAM based approach to improve the performance of beetle antennae search optimizer," *IEEE/CAA J. Autom. Sin.*, vol. 7, no. 2, pp. 461–471, 2020, doi: 10.1109/JAS.2020.1003048.
- [11] Y. Shao et al., "An Improvement of Adam Based on a Cyclic Exponential Decay Learning Rate and Gradient Norm Constraints," *Electron.*, vol. 13, no. 9, 2024, doi: 10.3390/electronics13091778.
- [12] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *J. Big Data*, vol. 6, no. 1, 2019, doi: 10.1186/s40537-019-0197-0.
- [13] L. Perez and J. Wang, "The Effectiveness of Data Augmentation in Image Classification using Deep Learning," 2017, [Online]. Available: <http://arxiv.org/abs/1712.04621>

- [14] M. Elgendi et al., “The Effectiveness of Image Augmentation in Deep Learning Networks for Detecting COVID-19: A Geometric Transformation Perspective,” *Front. Med.*, vol. 8, no. March, pp. 1–12, 2021, doi: 10.3389/fmed.2021.629134
- [15] Hidra Amnur, A. K. Vadreass, and M. Ridwan, “Aplikasi Pendeteksi Kematangan Tanaman Menggunakan Metode Transformasi Ruang Warna HSI (Hue, Saturation, Intensity) dan K-NN (K- Nearest Neighbor)”, *jitsi*, vol. 5, no. 4, pp. 161 -167, Dec. 2024.
- [16] C. Lei, B. Hu, D. Wang, S. Zhang, and Z. Chen, “A preliminary study on data augmentation of deep learning for image classification,” *ACM Int. Conf. Proceeding Ser.*, pp. 7–10, 2019, doi: 10.1145/3361242.3361259