

Pengaruh Penambahan Arsitektur Model dalam Klasifikasi Citra Bencana Alam Menggunakan Ensemble Learning

Ibnu Richo Kurniawan¹, Irsyad Umar Fakhri², Felicia Amanda Cahyadewi³, Fahrizal Denta Saputra⁴,
Achmad⁵, Muhammad Naufal⁶, Ricardus Anggi Pramunendar⁷

^{1,2,3,4,5,6,7}Research Center for Intelligent Distributed Surveillance and Security, Universitas Dian Nuswantoro, Indonesia

¹111202214063@mhs.dinus.ac.id, ²111202214094@mhs.dinus.ac.id, ³111202214090@mhs.dinus.ac.id *,

⁴111202214059@mhs.dinus.ac.id, ⁵111202214062@mhs.dinus.ac.id, ⁶m.naufal@dsn.dinus.ac.id,

⁷ricardus.anggi@dsn.dinus.ac.id

Abstract

This study focuses on improving the accuracy of natural disaster image classification by integrating three deep learning architectures, namely Convolutional Neural Network (CNN), InceptionV3, and InceptionResNetV2, within an ensemble learning approach. The main issues addressed are the limited accuracy of single models and the imbalance of performance across disaster categories, particularly flood and earthquake, which often lead to misclassification. The model was trained on a multi-class dataset consisting of four disaster categories: earthquake, flood, wildfire, and cyclone. Experimental results demonstrate that the ensemble approach based on soft voting significantly enhanced classification performance, achieving an accuracy of 96.02%. This value is higher than CNN at 88.3%, InceptionV3 at 94.1%, and InceptionResNetV2 at 92.4%. The improvement corresponds to a 7.5% reduction in error compared to the best single model, InceptionV3, thereby confirming the effectiveness of ensemble learning in combining the strengths of multiple architectures. Furthermore, the ensemble model provided consistent improvements in other evaluation metrics, including precision, recall, and F1-score, while exhibiting stronger generalization, particularly for more complex categories. With stable performance across all classes, the proposed approach is highly relevant for image-based disaster detection systems that require both speed and reliability. This research highlights that integrating diverse deep learning models not only improves classification accuracy but also offers a more robust and dependable solution for real-time disaster detection, supporting mitigation strategies and emergency decision-making in critical scenarios.

Keywords: disaster classification, CNN, InceptionV3, InceptionResNetV2, ensemble learning

Abstrak

Penelitian ini berfokus pada peningkatan akurasi klasifikasi citra bencana alam melalui integrasi tiga arsitektur *deep learning*, yaitu *Convolutional Neural Network (CNN)*, *InceptionV3*, dan *InceptionResNetV2*, dalam suatu pendekatan *ensemble learning*. Permasalahan utama yang diangkat adalah keterbatasan akurasi model tunggal dan ketidakseimbangan performa antar kategori bencana, khususnya pada *flood* dan *earthquake* yang sering menimbulkan kesalahan klasifikasi. Model dilatih menggunakan dataset multikelas yang terdiri dari empat kategori bencana: *earthquake*, *flood*, *wildfire*, dan *cyclone*. Hasil eksperimen menunjukkan bahwa pendekatan *ensemble* berbasis *soft voting* mampu meningkatkan akurasi secara signifikan, mencapai 96,02%. Nilai ini lebih tinggi dibandingkan CNN dengan 88,3%, *InceptionV3* dengan 94,1%, dan *InceptionResNetV2* dengan 92,4%. Peningkatan akurasi tersebut setara dengan pengurangan *error* sebesar 7,5% dibandingkan model terbaik tunggal, yaitu *InceptionV3*, sehingga membuktikan efektivitas *ensemble learning* dalam mengombinasikan keunggulan tiap arsitektur. Selain itu, model *ensemble* memberikan peningkatan yang konsisten pada metrik evaluasi lain seperti *precision*, *recall*, dan *F1-score*, serta menunjukkan kemampuan generalisasi yang lebih baik, khususnya pada kategori yang lebih sulit diklasifikasikan. Dengan kinerja yang stabil pada seluruh kelas, pendekatan ini relevan untuk sistem deteksi bencana berbasis citra yang membutuhkan kecepatan dan reliabilitas tinggi. Penelitian ini menegaskan bahwa integrasi berbagai model *deep learning* tidak hanya mampu meningkatkan akurasi, tetapi juga menghadirkan solusi yang lebih robust dan dapat diandalkan untuk implementasi sistem deteksi bencana waktu nyata dalam mendukung mitigasi serta pengambilan keputusan darurat.

Kata kunci: klasifikasi bencana, CNN, *InceptionV3*, *InceptionResNetV2*, *ensemble learning*

©This work is licensed under a Creative Commons Attribution - ShareAlike 4.0 International License

1. Pendahuluan

Bencana alam seperti *cyclone*, *earthquake*, *flood*, dan *wildfire* adalah peristiwa tak terduga yang sering menyebabkan kerusakan masif terhadap manusia, infrastruktur, dan lingkungan [1]. Dampak destruktif dari bencana ini semakin parah dengan adanya perubahan iklim yang memperburuk frekuensi dan intensitas bencana, yang menuntut sistem klasifikasi bencana yang cepat, akurat, dan adaptif. Salah satu

solusi yang dapat digunakan adalah sistem klasifikasi berbasis citra digital, yang dapat membantu dalam pengambilan keputusan secara real-time dalam manajemen bencana. Seiring dengan meningkatnya kebutuhan akan sistem klasifikasi yang lebih efisien, kemajuan dalam teknologi *deep learning*, khususnya dalam bidang *Computer Vision*, memungkinkan penggunaan model *Convolutional Neural Network (CNN)* sebagai metode utama dalam klasifikasi citra bencana. CNN unggul dalam mengekstraksi fitur

spasial dari data visual, tanpa memerlukan rekayasa fitur manual [2]. CNN telah banyak diterapkan pada berbagai masalah klasifikasi citra, termasuk di sektor bencana alam, dan memberikan hasil yang sangat baik dalam deteksi objek visual yang kompleks [3]. Model CNN ini telah terbukti efektif dalam berbagai aplikasi visual, termasuk dalam klasifikasi citra bencana alam, karena kemampuannya untuk menangani kompleksitas data yang sangat besar dan beragam [5,6]. *InceptionV3*, sebagai varian dari CNN, dirancang untuk menangani variasi skala dan kompleksitas visual yang tinggi melalui mekanisme konvolusi multiskala. Dengan arsitektur ini, model mampu mengklasifikasikan citra dengan berbagai tingkat kedalaman dan resolusi yang beragam [5]. Di sisi lain, arsitektur *InceptionResNetV2* menggabungkan *Inception* dengan *residual learning* dari ResNet, yang mempercepat konvergensi model serta meningkatkan akurasi [6].

Perkembangan teknologi *deep learning* dan *ensemble learning* dalam konteks bencana alam juga menunjukkan tren yang semakin signifikan. Studi terbaru membuktikan bahwa kombinasi arsitektur CNN seperti *InceptionV3* dengan teknik ensemble dapat meningkatkan akurasi dan ketahanan model dalam klasifikasi citra bencana, sehingga sistem deteksi dapat lebih andal dalam mendukung mitigasi dan respons darurat [3,8].

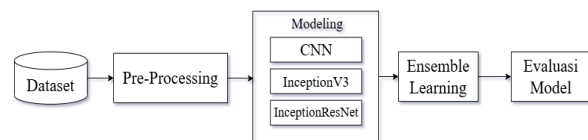
Sejumlah studi sebelumnya telah menunjukkan efektivitas penggunaan model CNN dalam klasifikasi citra bencana alam. Seperti menggabungkan *InceptionV3* dengan data blok wilayah untuk menilai kerusakan bangunan pasca-gempa dan mencapai akurasi sebesar 90,07% [3]. Selain itu, mengembangkan model ensemble yang menggabungkan CNN dan *InceptionV3* untuk klasifikasi multi-kelas bencana alam, dengan akurasi mencapai 92,79% [2]. Peneliti lain juga menemukan bahwa meskipun CNN tunggal dengan augmentasi data dapat mencapai akurasi sekitar 90% [8]. Pendekatan ini sering kali tidak cukup robust terhadap ketidakseimbangan kelas dan variasi visual yang ekstrem pada dataset bencana. Masalah ketidakseimbangan kelas dalam dataset bencana alam merupakan tantangan serius karena kelas minoritas sering terwakili jauh lebih sedikit. Penelitian terbaru menunjukkan bahwa integrasi teknik augmentasi data dengan pendekatan ensemble learning mampu meningkatkan stabilitas prediksi pada kelas minoritas, menjadikan metode ini sangat relevan untuk klasifikasi citra bencana [10,11].

Untuk mengatasi tantangan tersebut, pendekatan ensemble learning yang menggabungkan beberapa arsitektur deep learning seperti CNN, *InceptionV3*, dan *InceptionResNetV2* dapat menjadi solusi yang lebih adaptif dan akurat dalam klasifikasi citra bencana [11]. Pendekatan ensemble menggabungkan keunggulan masing-masing model untuk mengurangi kesalahan dan meningkatkan generalisasi pada berbagai jenis bencana [12]. Penelitian ini bertujuan untuk

mengembangkan model klasifikasi citra bencana alam berbasis ensemble learning yang mengintegrasikan ketiga arsitektur tersebut, dengan harapan dapat menghasilkan akurasi klasifikasi yang lebih tinggi dan robust dibandingkan dengan model-model tunggal [13].

2. Metode Penelitian

Penelitian ini dirancang dengan mengikuti serangkaian tahapan yang sistematis dan terstruktur, dimulai dari proses pengumpulan dataset yang berkaitan dengan kejadian bencana alam hingga tahap klasifikasi untuk menentukan jenis-jenis bencana yang terjadi. Tahapan-tahapan tersebut mencakup proses *pre-processing* untuk mempersiapkan data agar dapat diolah dengan baik, pengembangan model untuk analisis data, serta penerapan teknik *ensemble learning* guna meningkatkan ketepatan dan akurasi hasil klasifikasi. Untuk memberikan gambaran yang lebih jelas mengenai alur kerja penelitian ini, metodologi secara keseluruhan dapat dilihat secara rinci pada Gambar 1.



Gambar 1. Diagram Blok

Pengumpulan dataset citra bencana alam dilakukan dengan mengumpulkan gambar dari berbagai sumber, yang mencakup beberapa kelas seperti banjir (*flood*), siklon (*cyclone*), gempa bumi (*earthquake*), dan kebakaran hutan (*wildfire*). Pada tahap *pre-processing* dan filtrasi, citra yang dikumpulkan mengalami normalisasi, augmentasi, serta pembagian menjadi data pelatihan dan validasi untuk mempersiapkan model pelatihan. Seperti yang dapat dilihat pada Gambar 1 di atas, citra-citra tersebut melalui proses augmentasi untuk meningkatkan keragaman data. Selanjutnya, citra yang telah diproses digunakan untuk melatih tiga model klasifikasi, yaitu CNN, *InceptionV3*, dan *InceptionResNetV2*. Teknik ensemble learning diterapkan dengan melatih beberapa model secara paralel, dan hasil prediksi digabungkan menggunakan metode seperti *voting* atau *stacking* untuk meningkatkan akurasi klasifikasi akhir. Evaluasi dilakukan untuk mengukur performa model melalui matriks-matriks seperti akurasi, presisi, *recall*, dan *F1-score* pada data validasi, serta perbandingan antara model tunggal dan model ensemble untuk menilai peningkatan kinerja yang dicapai.

2.1. Dataset

Dataset yang digunakan dalam penelitian ini merupakan kumpulan citra (gambar) yang diklasifikasikan ke dalam empat kategori bencana alam utama, yaitu *flood* (banjir), *earthquake* (gempa bumi), *cyclone* (angin topan), dan *wildfire* (kebakaran hutan). Dataset ini diperoleh dari platform Kaggle melalui tautan berikut ini:

https://www.kaggle.com/datasets/rupakroy/cyclone-wildfire-flood-earthquake-database?select=Cyclone_Wildfire_Flood_Earthquake_Database. Pada sumber aslinya, jumlah citra per kategori mendekati 1000 gambar, namun dalam penelitian ini hanya digunakan subset terpilih sebanyak 600 gambar per kategori, sehingga total jumlah data yang digunakan dalam penelitian ini mencapai 2.400 gambar.

Pemilihan subset ini dilakukan secara acak untuk menjaga keseimbangan jumlah data antar kategori, menghilangkan citra duplikat atau berkualitas rendah, serta memastikan keragaman visual tetap terjaga [14]. Langkah ini juga bertujuan untuk mengurangi potensi bias model yang diakibatkan karena ketidakseimbangan jumlah data pada dataset asli. Setiap gambar dalam subset mewakili situasi nyata dari jenis bencana tertentu, yang diambil dari berbagai lokasi dan sudut pandang. Rincian distribusi data untuk setiap kategori dapat dilihat pada Tabel 1.

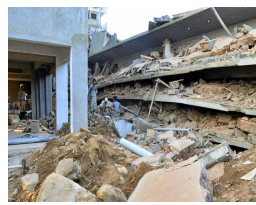
Tabel 1. Tabel Persebaran Data Citra Bencana Alam

Name Disaster	Total
Flood	600
Cyclone	600
Earthquake	600
Wildfire	600

Dataset ini berisi gambar-gambar representatif dari berbagai jenis bencana alam. Gambar-gambar ini dapat digunakan sebagai referensi visual untuk mendeteksi atau mengklasifikasikan jenis bencana berdasarkan citra.



Gambar 2. Cyclone



Gambar 3. Earthquake



Gambar 4. Flood



Gambar 5. Wildfire

Dataset ini digunakan dalam proses pelatihan dan pengujian model klasifikasi gambar berbasis *deep learning*, yang memanfaatkan fitur visual dari masing-masing gambar untuk mengenali jenis bencana secara otomatis. Untuk menunjang performa model secara optimal, dataset ini dibagi menjadi tiga bagian utama, yaitu data *training*, *validation*, dan *testing*, sebagaimana ditampilkan pada Tabel 2.

Tabel 2. Tabel Pembagian Data

Disaster	Train	Val	Test
Flood	420	90	90

Cyclone	420	90	90
Earthquake	420	90	90
WildFire	420	90	90

Dataset dibagi menjadi tiga bagian, yaitu 70% untuk melatih model, 15% untuk memantau kinerja model selama pelatihan dan membantu mendeteksi potensi *overfitting*, serta 15% untuk mengevaluasi performa akhir model setelah proses pelatihan selesai [2]. Data training digunakan untuk mengajarkan model agar dapat mempelajari pola dari data, sementara data validation digunakan selama pelatihan untuk mengukur sejauh mana model dapat generalisasi dan menghindari *overfitting*. Setelah pelatihan selesai, data *testing* digunakan untuk memberikan gambaran tentang seberapa baik model dapat memprediksi data yang belum pernah dilihat sebelumnya.

2.2. Pre-processing and Filtration

Pada penelitian ini seluruh *citra* diubah ukurannya menjadi 299×299 piksel untuk disesuaikan dengan arsitektur model yang digunakan [15]. Augmentasi yang digunakan meliputi rotasi acak hingga 20°, pergeseran horizontal dan vertikal sebesar 10%, pembesaran (*zoom*) hingga 15%, serta pembalikan horizontal. Sementara itu, data validasi dan pengujian hanya dikenakan proses *rescaling* tanpa augmentasi agar hasil evaluasi tidak bias oleh modifikasi sintesis.

2.3. Proposed Models

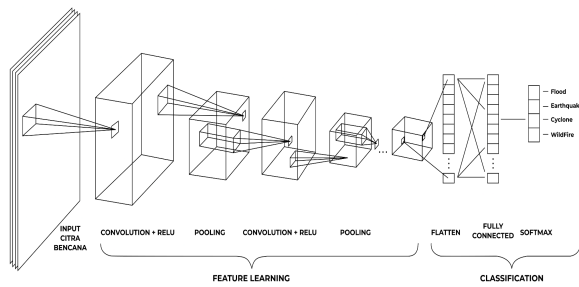
Untuk mendeteksi bencana alam dari citra (gambar), kami mengusulkan penggunaan beberapa arsitektur *deep learning* yang terbukti efektif dalam tugas klasifikasi gambar, yaitu *Convolutional Neural Network (CNN)*, *InceptionV3*, dan *InceptionResNetV2*. Setiap arsitektur memiliki karakteristik dan keunggulan tersendiri dalam mengekstraksi fitur penting dari citra.

2.3.1 CNN

Convolutional Neural Network (CNN) adalah salah satu algoritma dari *deep learning* yang merupakan pengembangan dari *Multilayer Perceptron (MLP)* yang dirancang untuk mengolah data dalam bentuk dua dimensi, misalnya gambar atau suara [16]. CNN digunakan untuk membuat klasifikasi data yang telah diberi label. Lapisan CNN memiliki susunan neuron 3 dimensi (lebar, tinggi, kedalaman). Lebar dan tinggi merupakan ukuran lapisan sedangkan kedalaman mengacu pada jumlah lapisan. Sebuah CNN dapat memiliki puluhan hingga ratusan lapisan yang masing-masing mempelajari mendeteksi berbagai fitur pada gambar.

Struktur CNN terdiri dari input, proses ekstraksi fitur, proses klasifikasi dan output (Gambar 6). Proses ekstraksi fitur pada CNN tersusun dari beberapa lapisan, yaitu lapisan konvolusi, fungsi aktivasi (ReLU) dan pooling. CNN bekerja secara hierarki, jadi output pada lapisan konvolusi pertama digunakan sebagai input pada lapisan konvolusi selanjutnya. Pada proses klasifikasi terdiri dari *fully connected* dan

fungsi aktivasi (*softmax*) yang outputnya berupa hasil klasifikasi [17].



Gambar 6. Arsitektur CNN

CNN menganalisis data masukan, terutama gambar, melalui beberapa array. Hal ini memungkinkan CNN untuk memproses fitur spasiotemporal dengan resolusi yang lebih tinggi, kemudian mengubah fitur tersebut menjadi fitur yang lebih rumit pada resolusi yang lebih rendah [18]. Rumus dasar CNN yang digunakan untuk proses konvolusi adalah ditampilkan pada Persamaan (1), dengan y adalah keluaran dari neuron, f adalah fungsi aktivasi, w_i adalah bobot masukkan ke neuron, x_i adalah masukkan ke neuron, b adalah bias, dan n adalah jumlah masukan ke neuron.

$$a_{ij} = \sum_{m=0}^s \sum_{n=0}^s w_{m,n} \cdot x_{i+m,j+n} \quad (1)$$

Selain itu, model juga dilengkapi dengan penggunaan *Dropout* sebesar 0.5 pada lapisan *dense*, yang berfungsi untuk mengurangi risiko *overfitting* dengan cara menonaktifkan sejumlah unit secara acak selama proses pelatihan. Meskipun tidak dijelaskan secara eksplisit dalam landasan teori [17]. Teknik *dropout* telah menjadi praktik umum dalam arsitektur CNN modern sebagai upaya untuk meningkatkan kemampuan generalisasi model. *Dropout* adalah teknik regularisasi yang mengabaikan neuron yang dipilih secara acak selama pelatihan diusulkan oleh Hinton [19]. *Dropout* memungkinkan pembelajaran fitur yang lebih kompleks dan kira-kira menggandakan jumlah iterasi yang diperlukan untuk konvergen.

2.3.2 InceptionV3

Jaringan konvolusional merupakan blok penyusun arsitektur model *InceptionV3*. Model ini mencapai 3,46% dalam hal "error rate 5 teratas" setelah dilatih pada set data *ImageNet* [20]. *InceptionV3* merupakan salah satu model CNN yang memiliki 42 layer yang dinilai lebih efisien dalam segi kedalaman arsitektur, dan *error rate* yang lebih kecil. *InceptionV3* mengusung sistem yang berbeda, yaitu menggunakan sistem pemfaktoran *convolutional layer* menjadi multi-layers dengan ukuran kernel yang terbilang lebih kecil. Arsitektur ini memiliki parameter yang lebih tepat karena mampu mengurangi nilai parameternya dengan membagikan bobotnya menjadi beberapa multi-layer [19].

2.3.3 InceptionResNetV2

Tahap ini menggabungkan dua arsitektur *deep learning*, *Inception* dan *Residual Network (ResNet)*, dalam arsitektur *InceptionResNetV2* untuk meningkatkan efisiensi, stabilitas, dan akurasi dalam segmentasi citra, khususnya untuk deteksi bangunan dari citra udara [20]. *Inception* menggunakan berbagai ukuran kernel konvolusi secara paralel dalam satu blok untuk menangkap fitur dari berbagai skala, bentuk, dan tekstur. Sementara itu, *ResNet* menggunakan *residual connections* untuk menjaga informasi tetap mengalir antar layer, sehingga menghindari masalah *vanishing gradient* saat jaringan menjadi sangat dalam. *InceptionResNetV2* mengoptimalkan kedua pendekatan ini dengan memodifikasi blok *Inception* agar lebih ringan melalui faktorisasi kernel besar menjadi beberapa kernel kecil (misalnya 5×5 menjadi dua 3×3), serta membagi kernel $n \times n$ menjadi $n \times 1$ dan $1 \times n$. Setelah itu, ditambahkan *filter-expansion layer* berupa konvolusi 1×1 tanpa aktivasi untuk menyamakan dimensi sebelum dilakukan penjumlahan residual. Persamaan residual connection dirumuskan pada persamaan (2).

$$y = x + F(x) \quad (2)$$

di mana x adalah input, $F(x)$ adalah hasil transformasi fitur melalui layer konvolusi, dan y adalah output setelah penjumlahan residual. Untuk meningkatkan stabilitas, terutama ketika jumlah filter melebihi 1000, nilai $F(x)$ dikalikan dengan skalar s sebelum dijumlahkan, seperti pada persamaan (3).

$$y = x + s \times F(x) \quad (3)$$

di mana s biasanya bernilai antara 0.1 hingga 0.3. Strategi ini terbukti membantu menghindari kegagalan pelatihan pada jaringan dalam yang kompleks. Arsitektur *InceptionResNetV2* yang terdiri dari 164 layer ini berhasil mempercepat proses pelatihan dan menghasilkan akurasi yang lebih tinggi dibandingkan model-model sebelumnya seperti *InceptionV3* maupun *InceptionV4* pada dataset besar seperti *Inria Aerial Image Dataset*.

Perbedaan *InceptionResNetV2* dengan *InceptionV3* terletak pada penggunaan *residual connection*. *InceptionV3* menggunakan modul *Inception* untuk mengekstraksi fitur multi skala dari gambar, sehingga mampu menangkap informasi dari berbagai ukuran fitur secara efisien. Sementara itu, *InceptionResNetV2* menggabungkan arsitektur *Inception* dengan *residual connection* yang memungkinkan aliran informasi dari layer sebelumnya langsung diteruskan ke layer berikutnya. Hal ini mempermudah proses pelatihan pada jaringan yang sangat dalam dan membantu mengatasi masalah *vanishing gradient*, sehingga lebih stabil saat digunakan untuk klasifikasi citra [21].

2.4. Ensemble Learning

Tahap *ensemble learning* merupakan salah satu pendekatan yang bertujuan untuk meningkatkan akurasi hasil klasifikasi secara keseluruhan serta memperkuat stabilitas dalam proses prediksi [22].

Pendekatan ini dilakukan dengan cara menggabungkan *output* dari beberapa model pembelajaran mesin yang telah dilatih secara independen sebelumnya, sehingga kelemahan dari masing-masing model dapat saling dikompensasi dan menghasilkan prediksi akhir yang lebih andal dan generalisasi yang lebih baik terhadap data baru.

Metode *ensemble learning* yang digunakan adalah metode *Soft Voting*, yang menggunakan keluaran probabilitas dari masing-masing model *deep learning*. Proses ini dilakukan dengan cara menjumlahkan semua probabilitas di setiap model *deep learning*, yang kemudian membaginya dengan jumlah model *deep learning* yang digunakan. Prediksi akhir ditentukan berdasarkan kelas yang memiliki nilai probabilitas tertinggi dari hasil perhitungan. Prosedur *ensemble learning* ditunjukkan pada persamaan (4),

$$P_{ensemble} = \frac{P_{inception} + P_{inceptionresnet} + P_{cnn}}{3} \quad (4)$$

di mana *Pensemble* merupakan hasil akhir berupa probabilitas dari metode *ensemble*, yang diperoleh dari rata-rata nilai probabilitas output masing-masing model. Adapun *Pinception*, *Pinceptionresnet*, dan *Pcnn* masing-masing mewakili hasil *output* probabilitas dari model *InceptionV3*, *InceptionResNetV2*, dan *CNN Custom*.

Soft Voting memiliki keunggulan karena mempertimbangkan tingkat keyakinan (*confidence score*) setiap model, bukan hanya keputusan akhir berbasis suara mayoritas. Hal ini sangat relevan dalam klasifikasi citra bencana karena beberapa kategori, seperti *Flood* dan *Cyclone*, memiliki kemiripan fitur visual yang tinggi sehingga memerlukan pembobotan prediksi yang lebih halus untuk mengurangi kesalahan klasifikasi. Studi terkini menunjukkan bahwa *Soft Voting* mampu memberikan peningkatan *F1-score* dan *AUROC* yang signifikan dibanding *Hard Voting* pada berbagai domain, termasuk deteksi penipuan kartu kredit, prediksi penyakit jantung, dan prediksi likuefaksi tanah [23].

Teknik *Stacking* juga sering digunakan dalam *ensemble learning* karena memanfaatkan *meta-learner* untuk menggabungkan prediksi model dasar. Namun, metode ini memerlukan waktu pelatihan tambahan dan memiliki kompleksitas yang lebih tinggi [24]. Dalam konteks sistem deteksi bencana yang membutuhkan kecepatan, efisiensi komputasi, dan stabilitas, *Soft Voting* menjadi pilihan yang optimal karena dapat memadukan kekuatan tiap model tanpa menambah beban komputasi yang signifikan.

2.5. Evaluasi Model

Confusion matrix merupakan salah satu metode yang digunakan untuk mengevaluasi performa model klasifikasi. Pada tahap ini, *confusion matrix* digunakan untuk mengetahui performansi dari setiap kernel algoritma, yaitu *CNN*, *InceptionV3*, *InceptionResNetV2*, dan *Ensemble Learning*. Evaluasi

dilakukan dengan menghasilkan nilai-nilai *accuracy*, *precision*, *recall*, dan *F1-score* sebagai ukuran kinerja model.

Accuracy menunjukkan proporsi prediksi yang benar terhadap seluruh data yang diuji, yang dihitung menggunakan persamaan (5). Nilai ini menggambarkan seberapa sering model memprediksi dengan tepat dari seluruh sampel yang tersedia. *Precision* menunjukkan proporsi prediksi *True Positive* dari seluruh prediksi positif seperti pada persamaan (6). *Recall* menunjukkan proporsi data positif yang benar-benar terprediksi sebagai positif, seperti pada persamaan (7). Sementara *F1-Score* merupakan rata-rata harmonis dari *precision* dan *recall* seperti pada persamaan (8).

$$Accuracy = \frac{TP}{TP+FP} \times 100\% \quad (5)$$

$$Precision = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (6)$$

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (7)$$

$$F1score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \times 100\% \quad (8)$$

dengan TP_c adalah *True Positive*, TN_c adalah *True Negative*, FP_c adalah *False Positive*, dan FN_c adalah *False Negative* [25].

Struktur umum *confusion matrix* untuk kasus multi-kelas dalam penelitian ini diperlihatkan pada Tabel 3. Tabel tersebut menunjukkan distribusi antara kelas aktual dan hasil prediksi model. Setiap baris mewakili kelas sebenarnya, sedangkan setiap kolom menunjukkan kelas prediksi. Klasifikasi yang benar berada pada posisi diagonal utama, sedangkan prediksi salah terletak di luar diagonal. Penyajian dalam bentuk ini memudahkan evaluasi kinerja model secara lebih terperinci.

Tabel 3. Tabel Struktur *Confusion Matrix*

True\Pred	Cyclone	Wildfire	Earthquake	Flood
Cyclone	$C_{00}(TP)$	C_{01}	C_{02}	C_{03}
Wildfire	C_{10}	$C_{11}(TP)$	C_{12}	C_{13}
Earthquake	C_{20}	C_{21}	$C_{22}(TP)$	C_{23}
Flood	C_{30}	C_{31}	C_{32}	$C_{33}(TP)$

Untuk kasus multi-kelas, struktur *confusion matrix* diperluas menjadi matriks berukuran $n \times n$, di mana n merepresentasikan jumlah kelas yang diklasifikasikan. Setiap elemen c_{ij} pada matriks menunjukkan jumlah data aktual dari kelas ke- i yang diprediksi sebagai kelas ke- j . Visualisasi dalam bentuk heatmap sangat membantu dalam mengevaluasi kinerja model, karena memudahkan identifikasi prediksi yang tepat yang ditandai pada diagonal utama. Selain itu, *classification report* digunakan untuk merangkum hasil evaluasi performa model, baik pada klasifikasi biner maupun multi-kelas [26].

Penjelasan mengenai metrik-metrik yang umum disertakan dalam *classification report* dibahas secara mendalam. Metrik-metrik tersebut meliputi *accuracy*,

precision, *recall* (yang juga dikenal sebagai *sensitivity*), *specificity*, dan *F1-score*. Semua metrik ini dihitung berdasarkan nilai-nilai dalam *confusion matrix*, yaitu tabel yang mencatat jumlah prediksi benar dan salah dari model klasifikasi, termasuk *true positive (TP)*, *true negative (TN)*, *false positive (FP)*, dan *false negative (FN)*.

3. Hasil dan Pembahasan

Pada bab ini disajikan hasil dari rangkaian eksperimen yang telah dilakukan dalam penelitian, mulai dari evaluasi performa masing-masing model secara individu hingga hasil penggabungan model menggunakan pendekatan *ensemble learning*. Evaluasi dilakukan untuk mengukur kemampuan model dalam mengklasifikasikan citra bencana alam ke dalam empat kategori, yaitu *Cyclone*, *Wildfire*, *Earthquake*, dan *Flood*.

3.1. Evaluasi Model Individu

Pada tahap ini dilakukan pelatihan dan evaluasi terhadap tiga arsitektur model deep learning secara terpisah, yaitu *InceptionV3*, CNN konvensional, dan *InceptionResNetV2*. Ketiganya digunakan untuk mengklasifikasikan citra bencana alam ke dalam empat kelas: *Cyclone*, *Wildfire*, *Earthquake*, dan *Flood*. Tujuan utama dari tahap ini adalah untuk mengetahui kemampuan masing-masing model sebelum dikombinasikan dalam pendekatan *ensemble learning*. Adapun parameter pelatihan masing-masing model ditampilkan pada Tabel 4.

Tabel 4. Parameter Setiap Model

Parameter	CNN	<i>InceptionV3</i>	<i>InceptionResNetV2</i>
Optimizer	Adam	Adam	Adam
Learning Rate	0.0001	0.0001	0.0001
Batch Size	32	32	32
Epochs	50	50	50
Dropout	0.5	0.5	0.5
Input Size	299x299x3	299x299x3	299x299x3

Model pertama yang diuji adalah *InceptionV3*, sebuah arsitektur konvolusional yang dirancang oleh tim Google dengan konsep pemrosesan multi-skala dalam satu layer, yang dikenal sebagai *Inception Module* [15]. Berdasarkan hasil pengujian, model ini menunjukkan performa terbaik berdasarkan metrik klasifikasi seperti *precision*, *recall*, dan *F1-score* pada Tabel 5. Hasil ini menunjukkan bahwa arsitektur *InceptionV3* efektif dalam mengenali pola visual dari berbagai jenis bencana. Keunggulan presisi *InceptionV3* dibandingkan CNN konvensional terutama terlihat pada kategori bencana dengan ciri visual yang kompleks, seperti *wildfire* dan *flood*. Hal ini dapat dijelaskan oleh kemampuan *Inception Module* dalam mengekstraksi fitur pada berbagai skala secara simultan. Pendekatan ini memungkinkan model menangkap detail lokal dan global secara bersamaan, sehingga lebih akurat dalam membedakan objek yang mirip secara tekstur atau warna. Perbedaan hasil ini juga dapat dipengaruhi oleh kedalaman arsitektur dan

teknik regularisasi yang digunakan *InceptionV3* dibandingkan dengan CNN yang lebih sederhana.

Tabel 5. Tabel Model *InceptionV3*

Class	Precision	Recall	F1-Score	Support
<i>Cyclone</i>	100.00	94.00	97.00	90
<i>Wildfire</i>	93.00	98.00	95.00	90
<i>Earthquake</i>	97.00	87.00	92.00	90
<i>Flood</i>	87.00	97.00	92.00	90
Akurasi	94.00			360

Model kedua yang dievaluasi dalam penelitian ini adalah *Convolutional Neural Network (CNN)* konvensional, yang dirancang dengan struktur lapisan standar yang terdiri dari lapisan konvolusi, *pooling*, dan *fully connected* [27]. Model ini memberikan hasil yang cukup baik namun sedikit lebih rendah dibandingkan *InceptionV3* pada Tabel 6. Beberapa variasi performa antar kelas masih terlihat, menunjukkan keterbatasan model dalam menangani kompleksitas fitur tertentu.

Tabel 6. Tabel Model CNN

Class	Precision	Recall	F1-Score	Support
<i>Cyclone</i>	87.00	94.00	90.00	90
<i>Wildfire</i>	88.00	91.00	90.00	90
<i>Earthquake</i>	96.00	89.00	87.00	90
<i>Flood</i>	92.00	78.00	84.00	90
Akurasi	88.00			360

Model ketiga yang dievaluasi dalam penelitian ini adalah *InceptionResNetV2*, sebuah arsitektur yang menggabungkan konsep *Inception* dengan *residual connections* dari ResNet [28]. Arsitektur ini berhasil menunjukkan performa mendekati *InceptionV3*, dengan stabilitas klasifikasi yang cukup baik di sebagian besar kelas pada Tabel 7. Meskipun tidak menjadi yang terbaik, model ini menunjukkan kinerja yang kompetitif dalam skenario klasifikasi multi-kelas.

Tabel 7. Tabel Model *InceptionResNetV2*

Class	Precision	Recall	F1-Score	Support
<i>Cyclone</i>	99.00	97.00	98.00	90
<i>Wildfire</i>	93.00	98.00	95.00	90
<i>Earthquake</i>	93.00	78.00	85.00	90
<i>Flood</i>	83.00	94.00	89.00	90
Akurasi	92.00			360

Secara keseluruhan, ketiga model memiliki kekuatan yang berbeda-beda dalam mengenali masing-masing kelas. *InceptionV3* unggul dalam stabilitas performa dan presisi tinggi, CNN menunjukkan ketahanan terhadap *noise* meskipun dengan performa moderat, sementara *InceptionResNetV2* menawarkan keseimbangan antara kedalaman arsitektur dan akurasi. Perbedaan karakteristik ini menjadi dasar yang kuat untuk menerapkan strategi *ensemble*, dengan harapan bahwa kombinasi model mampu menutupi kelemahan masing-masing arsitektur dan menghasilkan sistem klasifikasi yang lebih andal dan akurat.

3.2. Hasil *Ensemble Learning*

Setelah mengevaluasi model secara individu, tahap berikutnya adalah penerapan teknik *ensemble learning* untuk meningkatkan performa klasifikasi. Teknik yang

digunakan adalah *soft voting*, yaitu menggabungkan probabilitas prediksi dari ketiga model individu.

Hasil dari model *ensemble* menunjukkan adanya peningkatan performa secara menyeluruh dibandingkan model-model individu. Peningkatan ini terlihat dari kenaikan metrik akurasi, presisi, *recall*, dan *F1-score* yang lebih stabil di semua kelas bencana. Informasi lengkap mengenai hasil ensemble dapat dilihat pada Tabel 8.

Tabel 8. Tabel Hasil *Ensemble*

Class	Precision	Recall	F1-Score	Support
Cyclone	99.00	98.00	98.00	90
Wildfire	97.00	98.00	97.00	90
Earthquake	96.00	91.00	94.00	90
Flood	91.00	96.00	93.00	90
Akurasi			96.00	360

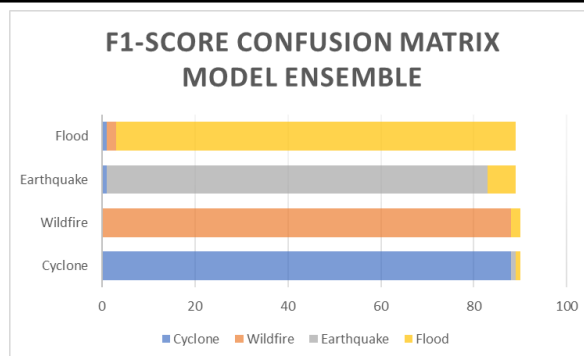
Analisis terhadap hasil klasifikasi menunjukkan bahwa kesalahan paling sering terjadi pada kelas *earthquake* dan *flood*. Kedua jenis bencana ini memiliki kemiripan visual yang tinggi, seperti dominasi warna abu-abu dan tekstur permukaan yang tidak terlalu kontras, sehingga menyulitkan model dalam membedakan citra secara akurat. Selain itu, variasi pencahayaan dan sudut pengambilan gambar dalam dataset turut memperbesar potensi kesalahan klasifikasi.

Untuk mengatasi hal tersebut, dilakukan beberapa penyesuaian pada proses pelatihan, termasuk augmentasi data yang lebih agresif seperti rotasi, zooming, dan peningkatan kontras, serta penerapan *class weighting* untuk memperkuat sensitivitas model terhadap kelas yang sering mengalami kesalahan. Penambahan model ketiga dalam sistem ensemble juga berperan penting dalam menyeimbangkan prediksi, karena masing-masing model memiliki kekuatan dalam mengekstraksi fitur yang berbeda.

Penerapan *soft voting* terbukti mampu meminimalisasi kelemahan dari masing-masing model. Ketika salah satu model memiliki akurasi rendah terhadap kelas tertentu, kontribusi model lain dengan performa lebih baik mampu menutupi kekurangan tersebut. Pendekatan ini menghasilkan klasifikasi yang lebih seimbang dan akurat, serta mengurangi fluktuasi performa antar kelas. Keunggulan metode ensemble menjadi sangat penting dalam konteks klasifikasi citra bencana, karena kesalahan klasifikasi dapat berdampak besar terhadap pengambilan keputusan dalam situasi darurat. Oleh karena itu, sistem klasifikasi yang andal dan konsisten menjadi sangat krusial dalam aplikasi dunia nyata.

3.3. Evaluasi Kinerja Klasifikasi

Evaluasi kinerja model tidak hanya dilakukan dengan melihat metrik numerik, tetapi juga melalui analisis *classification report* dan *confusion matrix* [29].



Gambar 7. *F1-score* per Kelas dari *Confusion Matrix* Model Ensemble

Laporan klasifikasi yang disajikan pada Tabel 9 memberikan gambaran performa model berdasarkan tiga metrik utama—*precision*, *recall*, dan *F1-score*—untuk setiap kelas. Keempat kelas utama, yaitu *Cyclone*, *Wildfire*, *Earthquake*, dan *Flood*, menunjukkan konsistensi performa dengan nilai metrik yang relatif tinggi. Akurasi makro dan rata-rata berbobot keduanya mencapai 96%, menandakan bahwa model mampu menjaga konsistensi prediksi tanpa kecenderungan bias terhadap kelas tertentu.

Visualisasi performa model ditampilkan pada Gambar 7, yang menunjukkan distribusi nilai *F1-score* untuk masing-masing kelas berdasarkan hasil *confusion matrix* dari model ensemble. Kelas *Cyclone* dan *Wildfire* memperoleh *F1-score* tertinggi, menunjukkan bahwa model mampu mengenali pola visual dari kedua jenis bencana dengan akurasi tinggi. Sebaliknya, kelas *Flood* dan *Earthquake* menunjukkan nilai *F1-score* yang sedikit lebih rendah, yang mengindikasikan adanya tantangan dalam membedakan citra dengan kemiripan tekstur dan warna. Grafik ini memperkuat temuan bahwa pendekatan ensemble tidak hanya meningkatkan akurasi keseluruhan, tetapi juga membantu menyeimbangkan performa antar kelas.

Tabel 9. Tabel *Classification Report*

Class	Precision	Recall	F1-Score	Support
Cyclone	0.99	0.98	0.98	90
Wildfire	0.97	0.98	0.97	90
Earthquake	0.96	0.91	0.94	90
Flood	0.91	0.96	0.93	90
Akurasi			0.96	360
Macro Avg	0.96	0.96	0.96	360
Weighted Avg	0.96	0.96	0.96	360

Secara umum, model ensemble berhasil mempertahankan performa yang tinggi dan merata di seluruh kelas. Tidak terdapat indikasi adanya dominasi model terhadap kelas tertentu, yang menunjukkan bahwa arsitektur ensemble mampu mengenali variasi visual dari gambar setiap bencana secara seimbang.

Untuk memperkuat temuan dari laporan klasifikasi, dilakukan juga analisis terhadap *confusion matrix* yang ditampilkan pada Tabel 10. Matriks ini menunjukkan distribusi prediksi model terhadap kelas sebenarnya. Mayoritas prediksi yang dihasilkan tergolong tepat sasaran.

Tabel 10. Tabel *Confusion Matrix*

TruePred	Cyclone	Wildfire	Earthquake	Flood
Cyclone	88	0	1	1
Wildfire	0	88	0	2
Earthquake	1	1	82	6
Flood	0	2	0	86

Kesalahan klasifikasi yang terjadi cenderung tidak sistematis dan tidak terpusat pada satu kelas tertentu. Hal ini memperlihatkan bahwa meskipun tantangan dalam klasifikasi gambar bencana cukup kompleks—misalnya, kemiripan warna dan tekstur antara gambar *Cyclone* dan *Flood*—model masih mampu mempertahankan presisi yang tinggi. Penerapan teknik ensemble terbukti dapat mereduksi kesalahan tersebut dengan menggabungkan prediksi dari beberapa model, sehingga menghasilkan keputusan akhir yang lebih stabil.

Stabilitas performa model ini juga diperkuat oleh penerapan teknik augmentasi data selama tahap pelatihan. Proses augmentasi seperti rotasi, pembalikan (*flipping*), *zooming*, dan pergeseran (*shifting*) secara efektif meningkatkan keragaman sampel pelatihan. Dampaknya, model menjadi lebih adaptif terhadap variasi data dan tidak mudah mengalami *overfitting* [30]. Augmentasi geometris seperti rotasi dan *flipping* meningkatkan ketahanan model terhadap perubahan orientasi dan simetri objek, yang relevan dalam citra satelit dan UAV [31]. *Zooming* dan *shifting* membantu model mengenali objek dalam berbagai skala dan posisi, memperkuat sensitivitas spasial terhadap fitur penting [32]. Hal ini menunjukkan bahwa teknik augmentasi sederhana dapat meningkatkan akurasi klasifikasi hingga 4.9% pada dataset *benchmark* dan secara signifikan mengurangi *overfitting* pada data terbatas [33]. Selain itu, augmentasi berfungsi sebagai regularisasi eksplisit, mendorong model untuk belajar dari distribusi fitur yang lebih luas dan meningkatkan generalisasi terhadap *noise* dan *distorsi* visual [34].

Untuk mengevaluasi apakah terdapat perbedaan performa yang signifikan antar model yang digunakan, baik model individual seperti *InceptionV3*, *InceptionResNetV2*, *CNN Custom*, maupun model gabungan, dilakukan analisis statistik terhadap hasil akurasi yang diperoleh dari beberapa percobaan. Analisis ini bertujuan untuk memberikan gambaran yang lebih jelas mengenai efektivitas masing-masing model dalam mengklasifikasikan citra bencana alam. Melalui pengujian ini, dapat diketahui apakah penambahan model tertentu, seperti *InceptionResNetV2*, memberikan kontribusi yang signifikan terhadap peningkatan performa. Hasil perbandingan performa antar model tersebut disajikan dalam Tabel 11, yang menggambarkan nilai akurasi, variansi, serta perbedaan signifikan di antara model-model yang diuji. Dengan melihat hasil tersebut, kita dapat menarik kesimpulan apakah strategi penggabungan model memberikan dampak positif pada klasifikasi.

Tabel 11. Tabel Hasil Uji Statistik Antar Model

Sumber Variasi	SS	df	MS	F	P-value	F kritis
Between Group	7.952	4	1.988	5.964	0.0044	3.056
Within Group	5.000	15	0.333	2		
Total	12.952	19				

Dari hasil tersebut, dapat disimpulkan bahwa terdapat perbedaan yang signifikan dalam kinerja antar model ($P\text{-value} < 0.05$). Ini mengindikasikan bahwa penambahan model seperti *InceptionResNetV2* ke dalam skema ensemble tidak hanya bersifat kosmetik, tetapi memberikan kontribusi nyata terhadap peningkatan performa. Dengan demikian, pendekatan ensemble dapat dianggap lebih unggul dibandingkan pendekatan berbasis model tunggal.

4. Kesimpulan

Penelitian ini membuktikan bahwa pendekatan *ensemble learning* yang menggabungkan *CNN*, *InceptionV3*, dan *InceptionResNetV2* mampu meningkatkan akurasi klasifikasi citra bencana alam secara signifikan. Dengan metode *soft voting*, model ensemble mencapai akurasi 96%, lebih tinggi dibandingkan model individu terbaik (*InceptionV3*) yang hanya mencapai 94%. Selain itu, evaluasi menggunakan *precision*, *recall*, dan *F1-score* memperlihatkan peningkatan kinerja yang konsisten di seluruh kategori, meskipun masih terdapat ambiguitas pada kelas *Earthquake* dan *Flood*.

Implikasi praktis dari penelitian ini sangat relevan untuk penerapan sistem deteksi bencana waktu nyata. Model ensemble yang lebih akurat dan stabil dapat diintegrasikan ke dalam platform monitoring visual seperti UAV (*drone*), satelit, maupun kamera pengawas, sehingga memungkinkan identifikasi cepat terhadap jenis bencana yang terjadi. Dengan dukungan sistem ini, lembaga pemerintah seperti BPBD (Badan Penanggulangan Bencana Daerah), BNPB (Badan Nasional Penanggulangan Bencana), maupun organisasi kemanusiaan, dapat mempercepat proses peringatan dini, pengambilan keputusan, dan alokasi sumber daya, sehingga penanganan bencana dapat dilakukan secara lebih efektif, tepat sasaran, dan efisien.

Saran penelitian lanjutan meliputi beberapa aspek penting. Pertama, ukuran dan keragaman dataset perlu diperluas agar model lebih robust terhadap variasi citra dari berbagai kondisi geografis dan sumber data. Kedua, penerapan metode interpretabilitas seperti Grad-CAM penting untuk meningkatkan transparansi dan kepercayaan pengguna terhadap sistem. Ketiga, eksplorasi metode ensemble yang lebih kompleks seperti *stacking* atau *boosting* berpotensi mendorong performa klasifikasi ke tingkat yang lebih tinggi. Keempat, uji coba implementasi pada sistem real-time dengan perangkat keras terbatas (*edge devices*) perlu

dilakukan agar model benar-benar siap digunakan di lapangan.

Daftar Rujukan

- [1] M. T. Chaudhary and A. Piracha, "Natural Disasters—Origins, Impacts, Management," *Encyclopedia*, vol. 1, no. 4, pp. 1101–1131, Oct. 2021, doi: 10.3390/encyclopedia1040084.
- [2] K. A. Sheth, R. P. Kulkarni, and G. K. Revathi, "Enhancing natural disaster image classification: an ensemble learning approach with inception and CNN models," *Geomat. Nat. Hazards Risk*, vol. 15, no. 1, p. 2407029, Dec. 2024, doi: 10.1080/19475705.2024.2407029.
- [3] H. Ma, Y. Liu, Y. Ren, D. Wang, L. Yu, and J. Yu, "Improved CNN Classification Method for Groups of Buildings Damaged by Earthquake, Based on High Resolution Remote Sensing Images," *Remote Sens.*, vol. 12, no. 2, p. 260, Jan. 2020, doi: 10.3390/rs12020260.
- [4] M. Haqqi, L. Rochmah, A. D. Safitri, R. A. Pratama, and Tarwoto, "Implementation Of Machine Learning To Identify Types Of Waste Using CNN Algorithm," *J. FASILKOM*, vol. 14, no. 3, pp. 761–765, Dec. 2024, doi: 10.37859/jf.v14i3.8116.
- [5] S. Godbole, A. Kattukaran, S. Savla, V. Pradhan, P. Kanani, and D. Patil, "Enhancing Paediatric Pneumonia Detection and Classification Using Customized CNNs and Transfer Learning Based Ensemble Models," *Int. Res. J. Multidiscip. Technovation*, pp. 38–53, Oct. 2024, doi: 10.54392/irjm2463.
- [6] A. M. Mustafa, R. Agha, L. Ghazalat, and T. Sha'ban, "Natural disasters detection using explainable deep learning," *Intell. Syst. Appl.*, vol. 23, p. 200430, Sept. 2024, doi: 10.1016/j.iswa.2024.200430.
- [7] A. Akhyar *et al.*, "Deep artificial intelligence applications for natural disaster management systems: A methodological review," *Ecol. Indic.*, vol. 163, p. 112067, June 2024, doi: 10.1016/j.ecolind.2024.112067.
- [8] M. Sobhana, S. C. Chaparala, D. N. V. S. L. S. Indira, and K. K. Kumar, "A disaster classification application using convolutional neural network by performing data augmentation," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 27, no. 3, p. 1712, Sept. 2022, doi: 10.11591/ijeecs.v27.i3.pp1712-1720.
- [9] R. Eltehwewy, A. Abouelfarag, and S. N. Saleh, "Efficient Classification of Imbalanced Natural Disasters Data Using Generative Adversarial Networks for Data Augmentation," *ISPRS Int. J. Geo-Inf.*, vol. 12, no. 6, p. 245, June 2023, doi: 10.3390/ijgi12060245.
- [10] A. A. Khan, O. Chaudhari, and R. Chandra, "A review of ensemble learning and data augmentation models for class imbalanced problems: Combination, implementation and evaluation," *Expert Syst. Appl.*, vol. 244, p. 122778, June 2024, doi: 10.1016/j.eswa.2023.122778.
- [11] "ELREI: Ensemble Learning of ResNet, EfficientNet, and Inception-v3 for Lung Disease Classification based on Chest X-Ray Image," *Int. J. Intell. Eng. Syst.*, vol. 16, no. 5, pp. 149–161, Oct. 2023, doi: 10.22266/ijies2023.1031.14.
- [12] A. Rathod, V. Pariawala, M. Surana, and K. Saxena, "Leveraging CNNs and Ensemble Learning for Automated Disaster Image Classification," in *Artificial Intelligence and Sustainable Computing*, M. Pandit, M. K. Gaur, and S. Kumar, Eds., in Algorithms for Intelligent Systems., Singapore: Springer Nature Singapore, 2024, pp. 311–326. doi: 10.1007/978-981-97-0327-2_23.
- [13] T. Yasuno, M. Amakata, and M. Okano, "Natural Disaster Classification Using Aerial Photography Explainable for Typhoon Damaged Feature," in *Pattern Recognition. ICPR International Workshops and Challenges*, vol. 12666, A. Del Bimbo, R. Cucchiara, S. Sclaroff, G. M. Farinella, T. Mei, M. Bertini, H. J. Escalante, and R. Vezzani, Eds., in Lecture Notes in Computer Science, vol. 12666., Cham: Springer International Publishing, 2021, pp. 15–25. doi: 10.1007/978-3-030-68780-9_2.
- [14] C. Yang, E. A. Fridgeirsson, J. A. Kors, J. M. Reys, and P. R. Rijnbeek, "Impact of random oversampling and random undersampling on the performance of prediction models developed using observational health data," *J. Big Data*, vol. 11, no. 1, p. 7, Jan. 2024, doi: 10.1186/s40537-023-00857-7.
- [15] Y. Fan *et al.*, "A Multi-Watermarking Algorithm for Medical Images Using Inception V3 and CT," *Comput. Mater. Contin.*, vol. 74, no. 1, pp. 1279–1302, 2023, doi: 10.32604/cmc.2023.031445.
- [16] H. Mukhtar, F. Alfano, H. Fu'adah Amran, F. Handayani, and R. Medikawati Taufiq, "Deep Learning Untuk Klasifikasi Kematangan Buah Mangrove Berdasarkan Warna," *J. FASILKOM*, vol. 13, no. 3, pp. 563–569, Dec. 2023, doi: 10.37859/jf.v13i3.6292.
- [17] W. D. Setyawan, A. Nilogiri, and Q. A'yun, "IMPLEMENTASI CONVOLUTION NEURAL NETWORK (CNN) UNTUK KLASIFIKASI PADA CITRA IKAN CUPANG HIAS," *JTIK J. Tek. Inform. Kaputama*, vol. 7, no. 1, pp. 101–110, Jan. 2023, doi: 10.59697/jtik.v7i1.45.
- [18] V. Wulandari, W. J. Sari, Z. H. Al-Sawaff, and S. Manickam, "Comparative Analysis of Weather Image Classification Using CNN Algorithm with InceptionV3, DenseNet169 and NASNetMobile Architecture Models," vol. 2, 2025, doi: https://doi.org/10.57152/predatecs.v2i2.1608.
- [19] M. I. Fathur Rozi, N. O. Adiwijaya, and D. I. Swasono, "Identifikasi Kinerja Arsitektur Transfer Learning Vgg16, Resnet-50, Dan Inception-V3 Dalam Pengklasifikasian Citra Penyakit Daun Tomat," *J. Ris. Rekayasa Elektro*, vol. 5, no. 2, p. 145, Dec. 2023, doi: 10.30595/jrre.v5i2.18050.
- [20] S. Aghayari, A. Hadavand, S. Mohamadnezhad Niazi, and M. Omidalizarandi, "BUILDING DETECTION FROM AERIAL IMAGERY USING INCEPTION RESNET UNET AND UNET ARCHITECTURES," *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. X-4/W1-2022, pp. 9–17, Jan. 2023, doi: 10.5194/isprs-annals-X-4-W1-2022-9-2023.
- [21] F. M. Firmando, D. R. I. M. Setiadi, A. R. Muslikh, and S. W. Iriananda, "Analyzing InceptionV3 and InceptionResNetV2 with Data Augmentation for Rice Leaf Disease Classification," *J. Future Artif. Intell. Technol.*, vol. 1, no. 1, pp. 1–11, May 2024, doi: 10.62411/faith.2024-4.
- [22] A. Munandar, W. Maulana Baihaqi, and A. Nurhopipah, "A Soft Voting Ensemble Classifier to Improve Survival Rate Predictions of Cardiovascular Heart Failure Patients," *Ilk. J. Ilm.*, vol. 15, no. 2, pp. 344–352, Aug. 2023, doi: 10.33096/ilkom.v15i2.1632.344-352.
- [23] M. Azim Mim, N. Majadi, and P. Mazumder, "A soft voting ensemble learning approach for credit card fraud detection," *Heliyon*, vol. 10, no. 3, p. e25466, Feb. 2024, doi: 10.1016/j.heliyon.2024.e25466.
- [24] N. Peppes, E. Daskalakis, T. Alexakis, E. Adamopoulou, and K. Demestichas, "Performance of Machine Learning-Based Multi-Model Voting Ensemble Methods for Network Threat Detection in Agriculture 4.0," *Sensors*, vol. 21, no. 22, p. 7475, Nov. 2021, doi: 10.3390/s21227475.
- [25] K. Kusumaningtyas, I. Dwijayanti, A. Rizqi Lahitani, and M. Habibi, "Analisis Tren Topik dalam Ulasan Negatif Aplikasi M-Banking Menggunakan Latent Dirichlet Allocation," *J. FASILKOM*, vol. 14, no. 3, pp. 549–555, Nov. 2024, doi: 10.37859/jf.v14i3.8035.
- [26] O. Rainio, J. Teuho, and R. Klén, "Evaluation metrics and statistical tests for machine learning," *Sci. Rep.*, vol. 14, no. 1, p. 6086, Mar. 2024, doi: 10.1038/s41598-024-56706-x.
- [27] A. Zafar *et al.*, "Convolutional Neural Networks: A Comprehensive Evaluation and Benchmarking of Pooling Layer Variants," *Symmetry*, vol. 16, no. 11, p. 1516, Nov. 2024, doi: 10.3390/sym16111516.
- [28] M. Neshat, M. Ahmed, H. Askari, M. Thilakarathne, and S. Mirjalili, "Hybrid Inception Architecture with Residual Connection: Fine-tuned Inception-ResNet Deep Learning Model for Lung Inflammation Diagnosis from Chest Radiographs," 2023, *arXiv*. doi: 10.48550/ARXIV.2310.02591.

-
- [29] S. Sathyanarayanan, "Confusion Matrix-Based Performance Evaluation Metrics," *Afr. J. Biomed. Res.*, pp. 4023–4031, Nov. 2024, doi: 10.53555/AJBR.v27i4S.4345.
- [30] A. Mumuni and F. Mumuni, "Data augmentation: A comprehensive survey of modern approaches," *Array*, vol. 16, p. 100258, Dec. 2022, doi: 10.1016/j.array.2022.100258.
- [31] H. Wang, Q. Wang, F. Yang, W. Zhang, and W. Zuo, "Data Augmentation for Object Detection via Progressive and Selective Instance-Switching," 2019, *arXiv*. doi: 10.48550/ARXIV.1906.00358.
- [32] S. Yun, D. Han, S. J. Oh, S. Chun, J. Choe, and Y. Yoo, "CutMix: Regularization Strategy to Train Strong Classifiers with Localizable Features," Aug. 07, 2019, *arXiv*: arXiv:1905.04899. doi: 10.48550/arXiv.1905.04899.
- [33] Z. Wang *et al.*, "A Comprehensive Survey on Data Augmentation," May 02, 2025, *arXiv*: arXiv:2405.09591. doi: 10.48550/arXiv.2405.09591.
- [34] J. Li, J. Pan, K.-C. Toh, and P. Zhou, "Towards Understanding Why Data Augmentation Improves Generalization," Feb. 13, 2025, *arXiv*: arXiv:2502.08940. doi: 10.48550/arXiv.2502.08940.