**BULLETIN OF COMPUTER SCIENCE RESEARCH**
**ISSN 2774-3659 (Media Online)**
Vol 6, No 6, October 2026 | Hal 2469-2481
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i6.1334

# Implementasi Pendekatan Ensemble Deep Learning untuk Klasifikasi Citra Ilustrasi Asli dan Hasil Generative AI

# Implementation of Ensemble Deep Learning for Classifying Original and Generative AI Illustrations

**Ibni Haziq Pryandara*, Yayan Hendrian**

Fakultas Teknik dan Informatika, Program Studi Informatika, Universitas Bina Sarana Informatika, Depok, Indonesia
Email: [1,*]ibnihaziq2@gmail.com, [2]yayan.yhn@bsi.ac.id
Email Penulis Korespondensi: ibnihaziq2@gmail.com
Received: **11/08/2026**, Revised: **12/10/2026**, Accepted: **13/10/2026**, Published**: 13/10/2026**

**Abstrak**−Perkembangan *Generative Artificial Intelligence* (AI) semakin memungkinkan pembuatan ilustrasi digital dengan kemiripan visual tinggi terhadap karya manusia, sehingga menyulitkan proses identifikasi serta menimbulkan tantangan terkait orisinalitas dan perlindungan hak cipta. Namun, penelitian deteksi citra *Generative AI* masih banyak berfokus pada citra fotografi atau objek dunia nyata, sedangkan klasifikasi ilustrasi digital dengan karakteristik artistik dan kemiripan visual tinggi masih terbatas. Penelitian ini bertujuan mengimplementasikan dan mengevaluasi metode klasifikasi otomatis untuk membedakan ilustrasi asli dan hasil *Generative AI*. Metode yang digunakan adalah *Ensemble Deep Learning* dengan teknik *Soft Voting* yang menggabungkan *MobileNetV2*, *ResNet50*, dan *DenseNet121*. Pemodelan mengikuti metodologi *CRISP-DM* dan dievaluasi menggunakan *5-Fold Stratified Cross Validation*. Hasil validasi internal memperoleh akurasi sebesar 96,00% dan *AUC* sebesar 99,06%. Pada pengujian eksternal, akurasi mencapai 87,60% dengan *AUC* sebesar 95,09%, sedangkan pengujian sampel individu menghasilkan akurasi sebesar 80,00%. Penurunan akurasi pada data eksternal menunjukkan adanya tantangan dalam generalisasi model terhadap data dengan karakteristik berbeda dari data pelatihan dan validasi. Kontribusi penelitian ini terletak pada penerapan dan evaluasi *Soft Voting Ensemble* untuk klasifikasi ilustrasi digital bergaya anime dan manga melalui pengujian internal, eksternal, dan sampel individu. Hasil penelitian menunjukkan bahwa pendekatan ini berpotensi digunakan untuk membedakan ilustrasi asli dan hasil *Generative AI*, meskipun peningkatan kemampuan generalisasi model masih diperlukan.

**Kata Kunci:** Kecerdasan Buatan; Ilustrasi Digital; Klasifikasi Citra; Generative AI; Soft Voting Ensemble

**Abstract**−The rapid development of Generative Artificial Intelligence (AI) has increasingly enabled the creation of digital illustrations with high visual similarity to human-created artworks, making identification more difficult and raising challenges related to originality and copyright protection. However, research on Generative AI image detection has largely focused on photographic images or real-world objects, while the classification of digital illustrations with artistic characteristics and high visual similarity remains limited. This study aims to implement and evaluate an automated classification method to distinguish original illustrations from Generative AI-generated illustrations. The proposed approach employs Ensemble Deep Learning with a Soft Voting technique that combines MobileNetV2, ResNet50, and DenseNet121. The modeling process follows the CRISP-DM methodology and is evaluated using 5-Fold Stratified Cross Validation. Internal validation achieved an accuracy of 96.00% and an AUC of 99.06%. In external testing, accuracy reached 87.60% with an AUC of 95.09%, while individual-sample testing achieved an accuracy of 80.00%. The decrease in accuracy on external data indicates challenges in generalizing the model to data with characteristics different from those in the training and validation sets. This study contributes by applying and evaluating a Soft Voting Ensemble for classifying anime- and manga-style digital illustrations through internal validation, external testing, and individual-sample testing. The results indicate that this approach has the potential to distinguish original illustrations from Generative AI-generated illustrations, although further improvements in the model's generalization capability are needed.

**Keywords**: Artificial Intelligence; Digital Illustration; Generative AI; Image Classification; Soft Voting Ensemble

## 1. PENDAHULUAN

Perkembangan teknologi *Artificial Intelligence* (AI) telah mendorong munculnya *Generative Artificial Intelligence* (Generative AI) yang mampu menghasilkan berbagai bentuk karya kreatif secara otomatis, termasuk gambar, musik, dan teks. Perkembangan tersebut menimbulkan tantangan baru terkait orisinalitas dan perlindungan hak cipta karena karya yang dihasilkan AI semakin sulit dibedakan dari karya manusia [1]. Model generatif seperti *Generative Adversarial Network* (GAN) dan *Diffusion Model* juga memungkinkan terbentuknya citra sintetis dengan kualitas visual yang semakin menyerupai citra asli sehingga proses identifikasi secara manual menjadi lebih sulit [2]. Selain persoalan orisinalitas, perkembangan media sintetis juga menimbulkan risiko terhadap penyalahgunaan dan manipulasi konten digital sehingga diperlukan metode deteksi yang mampu membedakan media asli dan media sintetis [3]. Pada bidang seni digital, penggunaan Generative AI untuk menghasilkan ilustrasi dengan gaya tertentu juga telah memunculkan perdebatan mengenai kreativitas, orisinalitas, dan hak cipta [4]. Kondisi tersebut menunjukkan perlunya metode klasifikasi otomatis yang dapat membantu membedakan ilustrasi karya manusia dan ilustrasi hasil Generative AI, khususnya pada citra dengan kemiripan visual yang tinggi.

Salah satu pendekatan yang digunakan dalam klasifikasi citra adalah *Convolutional Neural Network* (CNN), yang mampu mempelajari karakteristik visual melalui proses ekstraksi fitur pada beberapa lapisan. Penerapan CNN dengan arsitektur MobileNetV2 telah digunakan dalam berbagai permasalahan klasifikasi citra dengan mempertimbangkan

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. 
This work is licensed under a Creative Commons Attribution 4.0 International License.

efisiensi arsitektur dan kebutuhan komputasi [5],[6]. Sementara itu, DenseNet121 menggunakan mekanisme *dense connectivity* yang memungkinkan informasi dari lapisan sebelumnya digunakan kembali oleh lapisan berikutnya [7]. ResNet50 menggunakan mekanisme *residual connection* yang membantu proses pembelajaran pada jaringan dengan kedalaman tinggi [8]. Selain penggunaan satu arsitektur, beberapa model *Deep Learning* dapat dikombinasikan melalui pendekatan *ensemble learning*. Salah satu teknik yang digunakan adalah *Soft Voting*, yaitu menggabungkan nilai probabilitas yang dihasilkan oleh beberapa model sebelum menentukan kelas akhir [9].

Penelitian mengenai klasifikasi citra asli dan hasil Generative AI telah dilakukan dengan berbagai arsitektur *Deep Learning*. Aria et al. [10] membandingkan EfficientNet-B0 dan *Vision Transformer* (ViT) untuk membedakan gambar asli dan gambar Generative AI, dengan akurasi masing-masing sebesar 96% dan 98%. Qing Yun [11] membandingkan *High Resolution Network* (HRNet) dan ViT dalam membedakan visual buatan manusia dan AI, dengan akurasi masing-masing sebesar 95% dan 97%. Bird dan Lotfi [12] mengembangkan dataset CIFAKE dan menggunakan CNN untuk membedakan citra nyata dan citra sintetis dengan akurasi sebesar 92,98%. Maulana et al. [13] menggunakan ResNet-50 yang dikombinasikan dengan *Squeeze-and-Excitation* (SE) Block dan memperoleh akurasi sebesar 96,7%, tetapi masih menghadapi tantangan dalam mengidentifikasi citra sintetis dengan tingkat realisme visual yang tinggi. Hasil penelitian tersebut menunjukkan bahwa berbagai arsitektur *Deep Learning* telah mampu memberikan performa yang baik dalam membedakan citra asli dan hasil Generative AI, tetapi karakteristik data dan tingkat kemiripan visual dapat memengaruhi kemampuan model dalam melakukan klasifikasi.

Penelitian lain mengkaji ketahanan dan kinerja model dalam menghadapi perubahan karakteristik citra. Ranupadma et al. [14] menerapkan *transfer learning* dan augmentasi pada CNN untuk klasifikasi gambar *AI-generated* yang mengalami gangguan *noise*. Model *pre-trained* memperoleh akurasi sebesar 0,83 pada data tanpa gangguan, kemudian menurun menjadi 0,75 pada *Gaussian noise* dan 0,63 pada *salt and pepper noise*. Kaluge et al. [15] membandingkan CNN, ResNet-50 *from scratch*, dan ResNet-50 berbasis *transfer learning*. Hasilnya menunjukkan bahwa CNN berbasis tekstur dan ResNet-50 *from scratch* memiliki performa yang kompetitif, sedangkan *transfer learning* tidak memberikan peningkatan yang signifikan. Shadri et al. [16] membandingkan MobileNet, MobileNetV2, ResNet50, dan *Custom CNN*, dengan MobileNetV2 memperoleh performa terbaik berupa akurasi pengujian sebesar 88%. Arini et al. [17] membandingkan InceptionV3, VGG16, dan DenseNet121, dengan InceptionV3 memperoleh akurasi tertinggi sebesar 84,26%. Temuan tersebut menunjukkan bahwa performa klasifikasi tidak hanya dipengaruhi oleh arsitektur, tetapi juga oleh karakteristik data dan kondisi citra yang digunakan. Selain penggunaan model secara individual, pendekatan *ensemble learning* telah digunakan untuk menggabungkan beberapa model dalam proses klasifikasi. Kurniawan et al. [18] menggabungkan CNN, InceptionV3, dan InceptionResNetV2 menggunakan *Soft Voting* dan memperoleh akurasi 96,02%, lebih tinggi dibandingkan model individual. Mildawati dan Prasetyaningrum [19] menerapkan *Ensemble Soft Voting* dengan VGG16 dan EfficientNetV2B0 pada klasifikasi citra MRI tumor otak dan meningkatkan akurasi dari 98,73% menjadi 99,49%. Hasil tersebut menunjukkan bahwa penggabungan beberapa model melalui *Soft Voting* dapat memberikan peningkatan kinerja klasifikasi dibandingkan penggunaan model tunggal. Namun, kedua penelitian tersebut diterapkan pada domain yang berbeda dari klasifikasi ilustrasi digital.

Berdasarkan penelitian terdahulu, masih terdapat kesenjangan dalam klasifikasi citra *Generative AI*, khususnya pada ilustrasi digital dengan karakteristik artistik dan kemiripan visual tinggi antara karya manusia dan hasil AI. Bird dan Lotfi [12] menggunakan *Convolutional Neural Network* (CNN) untuk membedakan citra nyata dan citra sintetis, sedangkan Shadri et al. [16] membandingkan beberapa arsitektur *Deep Learning*, dengan *MobileNetV2* memperoleh akurasi pengujian sebesar 88%. Ranupadma et al. [14] menunjukkan bahwa gangguan pada citra dapat menurunkan performa klasifikasi. Di sisi lain, Kurniawan et al. [18] menggabungkan CNN, *InceptionV3*, dan *InceptionResNetV2* menggunakan *Soft Voting*, sedangkan Mildawati dan Prasetyaningrum [19] menerapkan *Ensemble Soft Voting* pada klasifikasi citra MRI tumor otak. Kedua penelitian terakhir menunjukkan potensi pendekatan *ensemble*, tetapi diterapkan pada domain citra yang berbeda dari ilustrasi digital bergaya anime dan manga. Namun, berdasarkan penelitian terdahulu yang dikaji, penerapan kombinasi MobileNetV2, ResNet50, dan DenseNet121 menggunakan Soft Voting untuk klasifikasi ilustrasi digital bergaya anime dan manga belum dibahas secara khusus. Selain itu, evaluasi model menggunakan data internal, data eksternal, dan sampel individu diperlukan untuk mengamati perbedaan kinerja model pada beberapa kondisi pengujian. Berdasarkan penelitian-penelitian tersebut, penelitian ini berfokus pada penerapan kombinasi *MobileNetV2*, *ResNet50*, dan *DenseNet121* menggunakan *Soft Voting* untuk membedakan *Real Art* dan *AI-Generated Art* pada ilustrasi digital bergaya anime dan manga. Evaluasi dilakukan pada data internal, data eksternal, dan sampel individu untuk mengamati kinerja model pada kondisi pengujian yang berbeda.

Penelitian ini menerapkan *Ensemble Deep Learning* dengan menggabungkan *MobileNetV2*, *ResNet50*, dan *DenseNet121* menggunakan metode *Soft Voting* untuk membedakan *Real Art* dan *AI-Generated Art* pada ilustrasi digital. Ketiga arsitektur dipilih berdasarkan karakteristik yang berbeda. *MobileNetV2* memiliki karakteristik arsitektur yang ringan dan efisien [5], *DenseNet121* memanfaatkan mekanisme *dense connectivity* yang memungkinkan penggunaan kembali informasi fitur dari lapisan sebelumnya [7], sedangkan *ResNet50* menggunakan *residual connection* yang mendukung proses pembelajaran pada jaringan dengan kedalaman tinggi [8]. Pemilihan ketiga arsitektur tersebut tidak didasarkan pada asumsi bahwa kombinasi ini merupakan kombinasi paling optimal, tetapi pada perbedaan karakteristik arsitektur yang dapat dimanfaatkan dalam pendekatan *ensemble*. Hasil pelatihan juga menunjukkan karakteristik pembelajaran yang berbeda pada ketiga arsitektur. *MobileNetV2* dan *DenseNet121* menunjukkan *validation accuracy* yang lebih tinggi dan relatif stabil, sedangkan *ResNet50* memiliki akurasi validasi yang lebih rendah dengan kurva

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. 
This work is licensed under a Creative Commons Attribution 4.0 International License.

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 6, October 2026 | Hal 2469-2481
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i6.1334

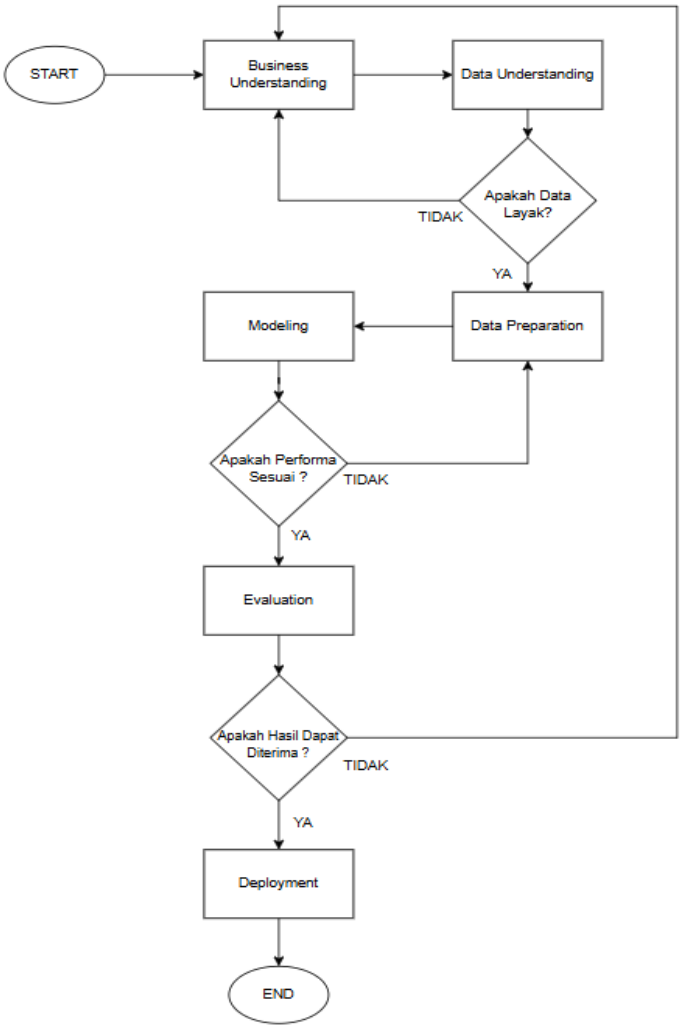
pembelajaran yang lebih berfluktuasi. Perbedaan karakteristik tersebut menjadi pertimbangan dalam mengevaluasi penggabungan prediksi ketiga model melalui metode *Soft Voting*.

Penelitian ini menggunakan metodologi *Cross-Industry Standard Process for Data Mining* (CRISP-DM) dengan evaluasi internal menggunakan *5-Fold Stratified Cross Validation*. Evaluasi juga dilakukan pada data eksternal yang tidak dilibatkan dalam proses pelatihan dan validasi serta pada sampel individu untuk mengamati kemampuan model ketika menghadapi data di luar proses pelatihan dan validasi. Dengan demikian, penelitian ini tidak hanya mengevaluasi kinerja model pada data internal, tetapi juga mengamati perbedaan kinerja pada kondisi pengujian yang berbeda. Kontribusi penelitian ini terletak pada penerapan kombinasi *MobileNetV2*, *ResNet50*, dan *DenseNet121* melalui metode *Soft Voting* untuk klasifikasi ilustrasi digital bergaya anime dan manga, serta evaluasi kinerja model menggunakan data internal, data eksternal, dan sampel individu untuk mengamati kemampuan generalisasinya dalam membedakan *Real Art* dan *AI-Generated Art*.

## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Penelitian

Penelitian ini menerapkan metodologi *Cross-Industry Standard Process for Data Mining* (*CRISP-DM*) sebagai kerangka kerja untuk membangun model klasifikasi ilustrasi asli dan ilustrasi yang dihasilkan *Generative Artificial Intelligence* (*Generative AI*). Pemilihan *CRISP-DM* didasarkan pada tahapan yang terstruktur dalam pelaksanaan proyek *data mining* serta penggunaannya yang luas dalam proyek *data science*. Metodologi ini mencakup enam tahapan, yaitu *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation*, dan *Deployment* [20]. Alur tahapan penelitian berdasarkan *CRISP-DM* ditampilkan dalam bentuk *flowchart* pada Gambar 1.




**Gambar 1.** Flowchart Tahapan Penelitian berdasarkan *CRISP-DM*

Berdasarkan *flowchart* pada Gambar 1, tahapan penelitian dilakukan secara sistematis melalui enam tahapan *CRISP-DM*. Setiap tahapan memiliki proses dan tujuan yang berbeda, mulai dari identifikasi permasalahan, pemahaman karakteristik data, persiapan data, pembangunan model, evaluasi kinerja model, hingga pengujian model. Penjelasan masing-masing tahapan penelitian adalah sebagai berikut.

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 2471
This work is licensed under a Creative Commons Attribution 4.0 International License.

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 6, October 2026 | Hal 2469-2481
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i6.1334

a. Tahap *Business Understanding* dimulai dengan mengidentifikasi permasalahan terkait semakin sulitnya membedakan ilustrasi karya manusia dengan ilustrasi yang dihasilkan *Generative Artificial Intelligence* (*Generative AI*) pada platform seni digital Pixiv karena meningkatnya kemiripan visual. Kondisi ini menjadi tantangan dalam mengidentifikasi keaslian karya, melindungi hak cipta, dan melakukan moderasi konten digital. Berdasarkan permasalahan tersebut, penelitian ini bertujuan mengembangkan model klasifikasi yang dapat membedakan ilustrasi asli dan ilustrasi hasil *Generative AI* secara otomatis. Untuk mencapai tujuan tersebut, digunakan pendekatan *Ensemble Deep Learning* dengan mengombinasikan arsitektur *MobileNetV2*, *DenseNet121*, dan *ResNet50* melalui metode *Soft Voting*.
b. Tahap *Data Understanding* dilakukan untuk memperoleh pemahaman mengenai karakteristik, struktur, dan kualitas dataset yang digunakan dalam penelitian. Dataset berupa citra ilustrasi digital berformat *.jpg* yang berasal dari dataset publik Kaggle dan kemudian dilengkapi dengan data tambahan yang dikumpulkan melalui proses *web scraping* pada platform seni digital Pixiv. Dataset tersebut mencakup ilustrasi digital dua dimensi (2D) bergaya anime dan manga dengan beragam gaya ilustrasi, komposisi objek, serta pewarnaan. Keragaman tersebut digunakan agar model dapat mempelajari perbedaan karakteristik visual antara ilustrasi asli dan ilustrasi hasil *Generative AI*. Dalam eksperimen, sebanyak 3.500 citra digunakan untuk pelatihan dan validasi internal, yang terdiri atas 1.750 *Real Art* dan 1.750 *AI-Generated Art*, sedangkan 500 citra digunakan sebagai data uji eksternal, terdiri atas 250 *Real Art* dan 250 *AI-Generated Art*. Data eksternal tidak dilibatkan dalam proses pelatihan maupun validasi internal sehingga digunakan untuk mengevaluasi kemampuan generalisasi model terhadap data baru.
c. Pada tahap *Data Preparation*, citra diolah agar memiliki format yang seragam dan memenuhi kebutuhan input model *Deep Learning*. Pengolahan dilakukan menggunakan pustaka OpenCV dan NumPy melalui lingkungan Google Colab. Dataset terlebih dahulu diakses dari Google Drive, kemudian setiap citra dibaca dan dikonversi dari format warna Blue-Green-Red (BGR) menjadi Red-Green-Blue (RGB) agar sesuai dengan kebutuhan *MobileNetV2*, *DenseNet121*, dan *ResNet50*. Selanjutnya, ukuran seluruh citra diseragamkan menjadi 224 × 224 piksel dan nilai piksel dinormalisasi dari rentang 0–255 menjadi 0–1. Untuk menghindari terhentinya proses akibat citra yang rusak atau tidak dapat dibaca, diterapkan mekanisme *error handling* sehingga citra tersebut dapat dilewati. Pada tahap akhir, citra beserta label diubah ke dalam bentuk NumPy array, dengan citra disimpan pada variabel X_data bertipe *float32* dan label pada variabel y_data bertipe *int8*. Label 1 digunakan untuk kelas *AI-Generated Art*, sedangkan label 0 untuk kelas *Real Art*. Setelah seluruh tahapan tersebut selesai, dataset siap digunakan dalam proses *Modeling*.
d. Pada tahap *Modeling*, teknik *transfer learning* serta *fine-tuning* diterapkan menggunakan tiga arsitektur *Convolutional Neural Network* (CNN), yaitu *MobileNetV2*, *DenseNet121*, dan *ResNet50*, yang digunakan sebagai *feature extractor* dengan ukuran input 224 × 224 × 3 piksel. Pemanfaatan *pre-trained weights* ImageNet dilakukan pada setiap model, dilanjutkan dengan proses *fine-tuning* melalui pelatihan ulang 10 layer terakhir *MobileNetV2*, 15 layer terakhir *DenseNet121*, dan 10 layer terakhir *ResNet50*. Seluruh *batch normalization layer* dibekukan (*frozen*) untuk menjaga stabilitas pelatihan. Hasil ekstraksi fitur dari masing-masing model diteruskan menuju *classification layer* yang terdiri atas *Global Average Pooling 2D*, *Dense layer* dengan 128 neuron dan aktivasi ReLU serta regularisasi L2, *Dropout layer*, dan *output layer* dengan aktivasi Sigmoid untuk klasifikasi biner. Proses kompilasi model menggunakan *binary crossentropy* sebagai fungsi *loss* dan Adam sebagai *optimizer*. Learning rate sebesar 0,00002 digunakan pada *MobileNetV2* dan *DenseNet121*, sedangkan *ResNet50* menggunakan learning rate sebesar 0,00001.Parameter *clipnorm* sebesar 1,0 diterapkan khusus pada *ResNet50* untuk menjaga kestabilan proses optimasi. Ketiga arsitektur dilatih selama maksimal 50 *epoch* dengan *batch size* 16. *Early stopping* diterapkan dengan *patience* 8 *epoch* dan pengaturan *restore_best_weights=True*. *ModelCheckpoint* digunakan untuk menyimpan model berdasarkan *validation loss*. Sebelum pelatihan utama, dilakukan pengujian diagnostik menggunakan salah satu bagian dari skema 5-Fold *Stratified Cross Validation* untuk mengevaluas i konfigurasi model serta memantau kecenderungan *underfitting* dan *overfitting* melalui *learning curve*. Pada setiap *fold*, data validasi diprediksi oleh satu model dari masing-masing arsitektur. Hasil prediksi dari *MobileNetV2*, *DenseNet121*, dan *ResNet50* kemudian digabungkan menggunakan metode *Soft Voting* dengan menghitung rata-rata probabilitas prediksi dari ketiga model. Dengan demikian, ketiga arsitektur memberikan kontribusi yang sama dalam menentukan probabilitas akhir. Secara matematis, probabilitas akhir *ensemble* dihitung dengan:

$$P_{Ensemble} = \frac{P_{MobileNetV2} + P_{DenseNet121} + P_{ResNet50}}{3} \quad (1)$$

Keterangan:
$P_{Ensemble}$= probabilitas akhir hasil *ensemble*.
$P_{MobileNetV2}$ = probabilitas prediksi dari *MobileNetV2*.
$P_{DenseNet121}$ = probabilitas prediksi dari *DenseNet121*.
$P_{ResNet50}$= probabilitas prediksi dari *ResNet50*.
Hasil probabilitas *ensemble* kemudian digunakan untuk menentukan kelas berdasarkan *threshold* 0,5. Probabilitas lebih dari 0,5 dikategorikan sebagai *AI-Generated Art*, sedangkan probabilitas kurang dari atau sama dengan 0,5 dikategorikan sebagai *Real Art*.
e. Tahap *Evaluation* dilakukan untuk mengukur kinerja, konsistensi, dan kemampuan sistem *Ensemble Deep Learning* setelah menggabungkan *MobileNetV2*, *DenseNet121*, dan *ResNet50* melalui metode *Soft Voting*. Evaluasi internal dilakukan menggunakan data validasi yang dihasilkan dari skema 5-Fold *Stratified Cross Validation*. Hasil prediksi

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 2472
This work is licensed under a Creative Commons Attribution 4.0 International License.

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 6, October 2026 | Hal 2469-2481
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i6.1334

ensemble dibandingkan dengan label sebenarnya untuk memperoleh nilai *accuracy*, *precision*, *recall*, dan *F1-score*. Confusion matrix digunakan untuk mengetahui bentuk kesalahan klasifikasi yang terdiri atas *true positive* (TP), *true negative* (TN), *false positive* (FP), dan *false negative* (FN). Selain itu, kemampuan diskriminasi model dianalisis menggunakan *Receiver Operating Characteristic* (ROC) dan *Area Under the Curve* (AUC).

f. Tahap *Deployment* dilakukan melalui pengujian model menggunakan dataset eksternal yang tidak dilibatkan dalam proses pelatihan maupun validasi. Dataset eksternal diproses menggunakan *ImageDataGenerator*, kemudian nilai piksel dinormalisasi ke rentang 0–1 dan ukuran citra diseragamkan menjadi 224 × 224 piksel sesuai dengan kebutuhan input model. Sistem *Soft Voting Ensemble* yang terdiri atas 15 model digunakan untuk menghasilkan prediksi pada data eksternal. Pada setiap *fold*, probabilitas dari tiga model, yaitu *MobileNetV2*, *DenseNet121*, dan *ResNet50*, digabungkan dengan menghitung rata-rata probabilitas. Hasil tersebut kemudian digunakan untuk menentukan kelas berdasarkan *threshold* 0,5. Probabilitas lebih dari 0,5 dikategorikan sebagai *AI-Generated Art*, sedangkan probabilitas kurang dari atau sama dengan 0,5 dikategorikan sebagai *Real Art*. Selain dataset uji eksternal, dilakukan simulasi pengujian terhadap 10 citra individu untuk mengamati kemampuan sistem dalam mengklasifikasikan ilustrasi dengan karakteristik visual serupa. Hasil prediksi dievaluasi menggunakan *Classification Report*, *Confusion Matrix*, ROC Curve, dan AUC. Hasil klasifikasi pada pengujian citra individu juga dikelompokkan menjadi tiga kategori, yaitu prediksi benar dan yakin, prediksi benar tetapi ragu, serta prediksi salah.

### 2.2 Objek Penelitian

Objek penelitian ini adalah citra ilustrasi digital dua dimensi (2D) dengan gaya anime dan manga yang diklasifikasikan ke dalam dua kelas, yaitu *Real Art* dan *AI-Generated Art*. Citra yang digunakan memiliki keragaman dalam gaya ilustrasi, susunan objek, dan pewarnaan. Penelitian ini berfokus pada penerapan Ensemble Deep Learning dengan menggabungkan arsitektur MobileNetV2, ResNet50, dan DenseNet121 menggunakan metode Soft Voting untuk membedakan ilustrasi karya manusia dan ilustrasi yang dihasilkan oleh Generative Artificial Intelligence (Generative AI) berdasarkan karakteristik visual citra.

Dataset penelitian berasal dari dataset publik Kaggle berjudul "*Pixiv Illustration Dataset of AI/Real Art"* yang dipublikasikan oleh Tagamit dan bersumber dari platform seni digital Pixiv, kemudian diperluas dengan data tambahan hasil *web scraping* pada platform seni digital Pixiv. Seluruh citra yang digunakan berformat *.jpg*. Data penelitian dibagi menjadi data internal dan data eksternal. Data internal terdiri atas 3.500 citra yang digunakan untuk proses pelatihan dan validasi menggunakan skema 5-Fold *Stratified Cross Validation*, dengan komposisi 1.750 citra *Real Art* dan 1.750 citra *AI-Generated Art*. Sementara itu, data eksternal terdiri atas 500 citra, yaitu 250 citra *Real Art* dan 250 citra *AI-Generated Art*. Data eksternal tidak digunakan dalam proses pelatihan maupun validasi dan digunakan untuk menguji kemampuan generalisasi model terhadap data yang belum digunakan sebelumnya. Dengan demikian, objek penelitian berfokus pada ilustrasi digital *Real Art* dan *AI-Generated Art* dengan karakteristik visual anime dan manga, sedangkan data eksternal digunakan sebagai data pengujian untuk melihat kemampuan model dalam melakukan klasifikasi pada citra di luar proses pelatihan dan validasi.

## 3. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil implementasi metode *Ensemble Deep Learning* yang menggabungkan *MobileNetV2*, *ResNet50*, dan *DenseNet121* menggunakan metode *Soft Voting* untuk mengklasifikasikan ilustrasi asli dan ilustrasi hasil *Generative Artificial Intelligence* (*Generative AI*). Hasil penelitian meliputi proses pelatihan model, evaluasi performa *ensemble* pada data internal, pengujian menggunakan data eksternal, serta pembahasan yang mengaitkan hasil penelitian dengan penelitian terdahulu guna menunjukkan kontribusi penelitian yang dilakukan.

### 3.1 Hasil Pelatihan Model

Penerapan tiga arsitektur *convolutional neural network* (CNN), mencakup MobileNetV2, ResNet50, dan DenseNet121, digunakan dalam mengoperasikan pelatihan. Tahap awal sebelum pengujian dilakukan dengan memisahkan *dataset* ke dalam data internal untuk pelatihan dan validasi, serta data eksternal untuk evaluasi akhir. Secara keseluruhan, dataset terdiri atas 4.000 citra ilustrasi digital yang dibagi secara seimbang ke dalam kelas *AI-Generated Art* dan *Real Art*. Sebanyak 3.500 citra digunakan sebagai data internal untuk pelatihan dan validasi, sedangkan 500 citra digunakan sebagai data eksternal untuk mengevaluasi kemampuan generalisasi model. Rincian pembagian dataset ditampilkan pada Tabel 1.

**Tabel 1.** Rincian Pembagian Jumlah Data Citra Pelatihan dan Pengujian

| Jenis Dataset | AI-Generated Art (AI) | Real Art (Real) | Total Citra |
|---|---|---|---|
| Dataset Internal (Pelatihan & Validasi) | 1.750 | 1.750 | 3.500 |
| Dataset Eksternal (Folder Pengujian /test) | 250 | 250 | 500 |
| Total Kumulatif Data | 2.000 | 2.000 | 4.000 |

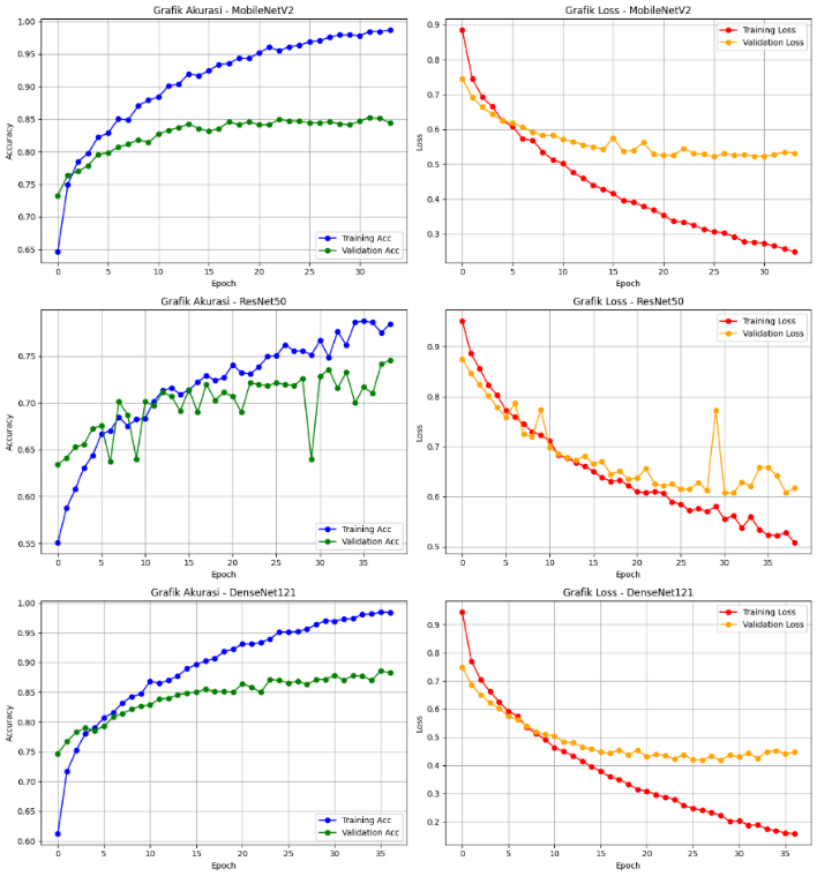
Berdasarkan Tabel 1, pembagian dataset dilakukan secara seimbang antara kelas *AI-Generated Art* dan *Real Art* pada data internal maupun eksternal. Komposisi tersebut bertujuan memastikan setiap kelas memiliki proporsi yang sama

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 2473
This work is licensed under a Creative Commons Attribution 4.0 International License.

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 6, October 2026 | Hal 2469-2481
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i6.1334

selama pelatihan dan pengujian sehingga model tidak cenderung mempelajari salah satu kelas secara dominan. Selain itu, evaluasi objektif mengenai kemampuan generalisasi model pada data asing dilakukan dengan menggunakan data eksternal yang tidak dilibatkan dalam proses pelatihan maupun validasi.

Setelah dataset dipisahkan, eksekusi pelatihan model dijalankan menggunakan tiga arsitektur *convolutional neural network* (CNN), yaitu MobileNetV2, ResNet50, serta DenseNet121 berlandaskan skema *5-fold stratified cross validation*. Melalui skema ini, skenario pelatihan dan validasi dapat dilangsungkan pada beberapa faksi data demi menghasilkan model yang lebih stabil. Pemantauan progres kinerja tiap arsitektur sepanjang pelatihan memanfaatkan *learning curve*, yang mencatat pergerakan *training accuracy*, *validation accuracy*, *training loss*, dan *validation loss* pada tiap *epoch*. Tampilan visual kurva pembelajaran dari ketiga arsitektur ini dituangkan pada Gambar 2.




**Gambar 2.** Kurva Pembelajaran (*Learning Curve*) MobileNetV2, ResNet50, dan DenseNet121

Berdasarkan Gambar 2, ketiga model menunjukkan karakteristik pembelajaran yang berbeda selama proses pelatihan. *MobileNetV2* menyelesaikan pelatihan dalam 34 *epoch* (0–33) dengan *training accuracy* sekitar 0,99 dan *validation accuracy* 0,85. Nilai *training loss* mencapai sekitar 0,25, sedangkan *validation loss* sebesar 0,53. *ResNet50* dilatih selama 39 *epoch* (0–38) dengan *training accuracy* sekitar 0,79 dan *validation accuracy* 0,74, serta menghasilkan *training loss* sekitar 0,51 dan *validation loss* sekitar 0,62. Model ini juga menunjukkan kurva pelatihan yang lebih berfluktuasi dibandingkan model lainnya. Sementara itu, *DenseNet121* menyelesaikan pelatihan dalam 37 *epoch* (0–36) dengan *training accuracy* sekitar 0,98 dan *validation accuracy* 0,88, serta menghasilkan *training loss* sekitar 0,16 dan *validation loss* sekitar 0,45. Hasil tersebut menunjukkan bahwa *DenseNet121* memiliki akurasi validasi tertinggi dan konvergensi yang lebih stabil. Secara umum, *MobileNetV2* dan *DenseNet121* menunjukkan performa pelatihan yang lebih baik dan *validation accuracy* yang relatif stabil, sedangkan *ResNet50* memiliki akurasi lebih rendah dan *validation loss* yang lebih berfluktuasi. Meskipun demikian, karakteristik pembelajaran *ResNet50* yang berbeda tetap memberikan variasi prediksi yang bermanfaat dalam pembentukan sistem *Ensemble Learning*.

Secara keseluruhan, ketiga arsitektur memiliki karakteristik pembelajaran yang saling melengkapi. *MobileNetV2* dan *DenseNet121* memberikan performa validasi yang lebih tinggi, sedangkan *ResNet50* memberikan variasi prediksi pada proses klasifikasi. Perbedaan tersebut menjadi dasar penerapan *Soft Voting Ensemble* untuk menggabungkan kemampuan ketiga model dan memperoleh performa klasifikasi yang lebih optimal pada tahap evaluasi berikutnya.

### 3.2 Hasil Evaluasi Model

Setelah pelatihan selesai, model *Ensemble Deep Learning* yang mengintegrasikan *MobileNetV2*, *ResNet50*, dan *DenseNet121* melalui metode *Soft Voting* selanjutnya diuji menggunakan 3.500 citra pada data internal. Pengujian ini bertujuan mengetahui kinerja model dalam membedakan kelas *AI-Generated Art* dan *Real Art* berdasarkan beberapa

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. 
This work is licensed under a Creative Commons Attribution 4.0 International License.

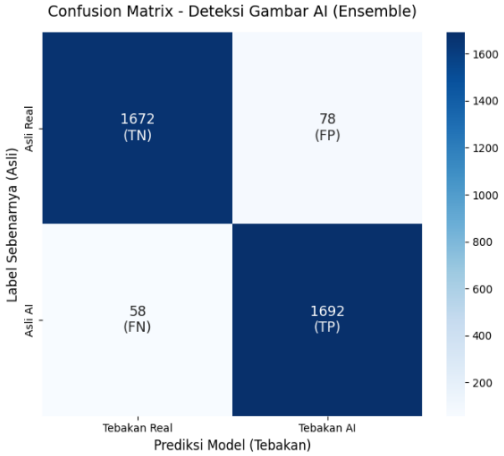
BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 6, October 2026 | Hal 2469-2481
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i6.1334

indikator, yaitu *precision*, *recall*, *F1-score*, *accuracy*, serta *Receiver Operating Characteristic–Area Under Curve* (*ROC-AUC*). Penggunaan berbagai metrik tersebut memberikan gambaran yang lebih menyeluruh mengenai kemampuan model dalam melakukan klasifikasi pada data validasi internal. Hasil *classification report* dari sistem ditampilkan pada Gambar 3.

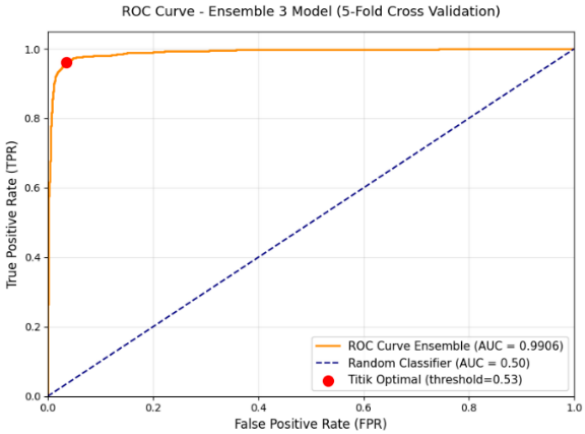
```
TABEL PERFORMA MODEL (BAB 4):
============================================================
              precision    recall  f1-score   support

    Real (0)       0.97      0.96      0.96      1750
      AI (1)       0.96      0.97      0.96      1750

    accuracy                           0.96      3500
   macro avg       0.96      0.96      0.96      3500
weighted avg       0.96      0.96      0.96      3500

ROC-AUC Score : 0.9906 (99.06%)
============================================================
```

**Gambar 3.** Classification Report Sistem Ensemble pada Validasi Internal

Hasil evaluasi internal yang ditampilkan pada Gambar 3 memperlihatkan bahwa sistem *Ensemble Deep Learning* mampu memberikan performa klasifikasi yang tinggi. Nilai *accuracy* yang diperoleh mencapai 96,00%, dengan *ROC-AUC Score* sebesar 0,9906 (99,06%). Pada kelas *Real (0)*, sistem menghasilkan *precision* 0,97, *recall* 0,96, serta *F1-score* 0,96. Sementara itu, kelas *AI (1)* memperoleh *precision* 0,96, *recall* 0,97, dan *F1-score* 0,96. Nilai *macro average* dan *weighted average* yang sama-sama mencapai 0,96 mengindikasikan bahwa model memberikan performa yang relatif merata pada kedua kelas. Kinerja tersebut diperoleh melalui mekanisme *Soft Voting* yang mengintegrasikan probabilitas prediksi dari *MobileNetV2*, *ResNet50*, dan *DenseNet121* sebelum menentukan hasil klasifikasi akhir. Mekanisme ini memungkinkan kontribusi ketiga arsitektur digunakan secara bersamaan, sehingga ketergantungan terhadap satu model dapat dikurangi dan perbedaan kesalahan antarmodel dapat saling dikompensasi. Dengan hasil tersebut, evaluasi kemudian dilanjutkan menggunakan *Confusion Matrix* dan kurva *Receiver Operating Characteristic* (*ROC*) sebagaimana ditampilkan pada Gambar 4 dan Gambar 5.




**Gambar 4.** Confusion Matrix Sistem Ensemble pada Validasi Internal




**Gambar 5.** ROC Curve Sistem Ensemble pada Validasi Internal

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 2475
This work is licensed under a Creative Commons Attribution 4.0 International License.

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 6, October 2026 | Hal 2469-2481
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i6.1334

Berdasarkan Gambar 4 dan Gambar 5, model *Ensemble Deep Learning* menunjukkan kemampuan klasifikasi yang sangat baik pada data validasi internal. Dari 1.750 citra *Real Art*, sebanyak 1.672 citra berhasil diklasifikasikan sebagai *Real* (*True Negative*) dengan ketepatan 95,54%, sedangkan 78 citra (4,46%) diprediksi sebagai *AI* (*False Positive*). Pada kelas *AI-Generated Art*, sebanyak 1.692 dari 1.750 citra berhasil diklasifikasikan sebagai *AI* (*True Positive*) dengan ketepatan 96,69%, sementara 58 citra (3,31%) diprediksi sebagai *Real* (*False Negative*). Hasil tersebut diperkuat oleh kurva *Receiver Operating Characteristic* (*ROC*) dengan *Area Under Curve* (*AUC*) sebesar 0,9906 (99,06%) dan *threshold* optimal 0,53 yang menunjukkan keseimbangan antara sensitivitas dan spesifisitas. Dengan demikian, *Soft Voting Ensemble* memiliki kemampuan diskriminasi yang sangat tinggi dalam membedakan *AI-Generated Art* dan *Real Art*. Secara keseluruhan, *Soft Voting Ensemble* yang menggabungkan *MobileNetV2*, *ResNet50*, dan *DenseNet121* menghasilkan *accuracy* 96,00% dan *AUC* 99,06%, serta menunjukkan performa klasifikasi yang akurat dan seimbang pada kedua kelas. Hasil ini menjadi dasar untuk menguji kemampuan generalisasi model pada dataset eksternal yang belum digunakan selama pelatihan maupun validasi.

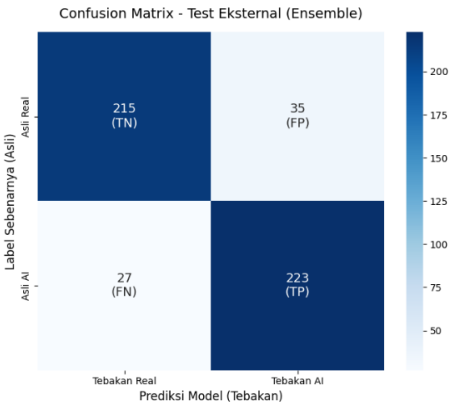
### 3.3 Hasil Pengujian Eksternal

Tahap akhir berupa pengujian eksternal dijalankan demi menilai keandalan generalisasi sistem *Soft Voting Ensemble* terhadap data yang terbebas dari rangkaian pelatihan maupun validasi silang internal. Kombinasi 15 model komputasi, yang mencakup masing-masing lima model MobileNetV2, ResNet50, dan DenseNet121, diuji memakai *dataset* eksternal sejumlah 500 citra, terbagi seimbang atas 250 citra kategori *AI-Generated Art* dan 250 citra *Real Art*. Seluruh data diambil dari direktori pengujian */test* sebagai *unseen data*, sehingga performa yang dihasilkan sanggup mencerminkan daya klasifikasi model pada data di luar proses pembelajaran terdahulu. Tampilan *classification report* dari pengujian eksternal ini dituangkan pada Gambar 6.

```
============================================================
          HASIL AKHIR TEST EKSTERNAL (15 MODEL)
============================================================
AKURASI FINAL : 87.60%
ROC-AUC SCORE : 0.9509 (95.09%)

Laporan Klasifikasi Final:
              precision    recall  f1-score   support

    Real (0)       0.89      0.86      0.87       250
      AI (1)       0.86      0.89      0.88       250

    accuracy                           0.88       500
   macro avg       0.88      0.88      0.88       500
weighted avg       0.88      0.88      0.88       500

============================================================
```

**Gambar 6.** Classification Report Sistem Ensemble pada Pengujian Eksternal

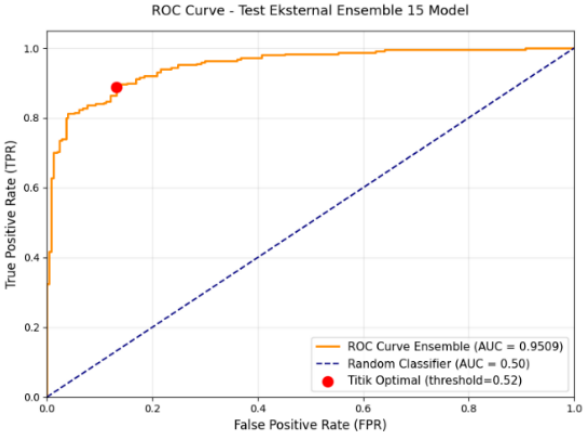
Hasil pengujian pada Gambar 6 memperlihatkan kemampuan model *ensemble deep learning* dalam meraih *accuracy* sebesar 87,60% dan *ROC-AUC* sebesar 0,9509 (95,09%) menggunakan data eksternal. Pada klasifikasi kelas *Real (0)* berjumlah 250 citra, diperoleh *precision* 0,89, *recall* 0,86, dan *F1-score* 0,87, sedangkan untuk kelas *AI (1)* dengan 250 citra, diperoleh *precision* 0,86, *recall* 0,89, dan *F1-score* 0,88. Perolehan nilai *macro average* serta *weighted average* yang identik pada angka 0,88 menandakan performa model yang seimbang di kedua kelas. Penurunan hasil dibanding evaluasi internal disebabkan oleh variasi karakteristik antara data pembelajaran dan data eksternal di luar proses pelatihan maupun validasi. Analisis tebakan secara komprehensif dilakukan menggunakan *confusion matrix* untuk mengevaluasi ketepatan klasifikasi di tiap kategori, serta *receiver operating characteristic* (ROC) untuk menilai daya diskriminasi antara *Real Art* dan *AI-Generated Art* pada bermacam *threshold*. Pemaparan analisis tersebut dituangkan dalam Gambar 7 dan Gambar 8.




**Gambar 7.** Classification Report Sistem Ensemble pada Pengujian Eksternal

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. 
This work is licensed under a Creative Commons Attribution 4.0 International License.

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 6, October 2026 | Hal 2469-2481
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i6.1334




**Gambar 8.** Classification Report Sistem Ensemble pada Pengujian Eksternal

Hasil *Confusion Matrix* pada Gambar 7 dan Gambar 8 memperlihatkan bahwa sistem *Ensemble Deep Learning* tetap mampu melakukan klasifikasi dengan baik ketika diterapkan pada data eksternal yang tidak digunakan dalam proses pelatihan maupun validasi. Dari keseluruhan 250 citra *Real Art*, sebanyak 215 citra dapat dikenali dengan tepat sebagai *Real* (*True Negative*) dengan tingkat ketepatan 86,00%, sementara 35 citra atau 14,00% diklasifikasikan secara keliru sebagai *AI* (*False Positive*). Pada kelompok *AI-Generated Art*, sistem berhasil mengidentifikasi 223 dari 250 citra sebagai *AI* (*True Positive*) dengan ketepatan 89,20%, sedangkan 27 citra atau 10,80% keliru dikategorikan sebagai *Real* (*False Negative*). Hasil tersebut memperlihatkan bahwa kemampuan sistem dalam mengenali kedua kelas relatif seimbang, dengan tingkat kesalahan yang masih terkendali pada data eksternal. Kurva *Receiver Operating Characteristic* (*ROC*) pada Gambar 8 turut memperlihatkan kemampuan sistem dalam memisahkan kelas *Real Art* dan *AI-Generated Art*. Perolehan nilai *Area Under Curve* (*AUC*) sebesar 0,9509 atau 95,09% menunjukkan bahwa model memiliki tingkat diskriminasi yang sangat baik ketika diuji menggunakan data eksternal. Analisis kurva tersebut juga menghasilkan *threshold* optimal sebesar 0,52 berdasarkan jarak kesalahan terpendek, sehingga nilai tersebut dapat digunakan sebagai batas probabilitas yang relatif optimal untuk membedakan kedua kelas. Secara keseluruhan, meskipun terjadi penurunan performa dibandingkan hasil validasi internal, *Soft Voting Ensemble* masih mampu mempertahankan kemampuan deteksi yang baik ketika menghadapi data baru.

Hasil pengujian tersebut menunjukkan bahwa sistem *Soft Voting Ensemble* memperoleh *accuracy* sebesar 87,60% dan nilai *AUC* sebesar 0,9509 atau 95,09%, dengan performa yang relatif seimbang pada kelas *Real Art* dan *AI-Generated Art*. Walaupun terjadi penurunan *accuracy* dari 96,00% pada validasi internal, model tetap mampu mengklasifikasikan sebagian besar citra dengan tingkat kesalahan yang relatif terkendali. Penggabungan *MobileNetV2*, *ResNet50*, dan *DenseNet121* melalui metode *Soft Voting* menunjukkan kemampuan generalisasi yang cukup baik ketika model diterapkan pada data eksternal dengan karakteristik yang berbeda dari data pelatihan.

### 3.4 Pengujian Individu Data Uji Baru

Sebagai bentuk validasi tambahan untuk melihat kemampuan sistem secara praktis di luar evaluasi numerik pada dataset eksternal, dilakukan pengujian individu menggunakan 10 sampel citra baru yang diambil dari platform Pixiv. Sampel tersebut terdiri atas 5 citra ilustrasi karya manusia yang dipilih dari daftar *Illustrations Daily Rankings* tanpa label *AI-generated* dan 5 citra hasil kecerdasan buatan yang diambil dari kategori dengan tag "AI生成" (*AI-generated*). Pengujian ini dilakukan untuk mengamati kemampuan sistem dalam memberikan prediksi pada citra yang memiliki karakteristik visual yang beragam serta memiliki tingkat kemiripan antara ilustrasi karya manusia dan ilustrasi hasil AI. Sampel visual dari kesepuluh citra yang digunakan dalam pengujian individu disajikan pada Tabel 2.

**Tabel 2.** Sampel Citra Ilustrasi Karya Manusia dan Buatan AI pada Pengujian Individu

| NO | Citra Asli Karya Ilustrator Manusia (Daily Rankings) | Citra Asli Hasil Kecerdasan Buatan (*Tag AI-generated*) |
|---|---|---|
| 1 | | |

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 2477
This work is licensed under a Creative Commons Attribution 4.0 International License.

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 6, October 2026 | Hal 2469-2481
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i6.1334

| NO | Citra Asli Karya Ilustrator Manusia (Daily Rankings) | Citra Asli Hasil Kecerdasan Buatan (*Tag AI-generated*) |
|---|---|---|
| 2 | | |
| 3 | | |
| 4 | | |
| 5 | | |

Selanjutnya, setiap sampel citra diuji secara individual menggunakan sistem *Ensemble Deep Learning* untuk memperoleh hasil klasifikasi beserta tingkat probabilitas prediksinya. Pengujian ini dilakukan untuk melihat tingkat keyakinan sistem terhadap keputusan yang diberikan pada masing-masing citra, baik pada karya ilustrator manusia maupun ilustrasi yang dihasilkan oleh AI. Hasil prediksi dan tingkat probabilitas sistem terhadap sampel citra karya ilustrator manusia disajikan pada Tabel 3, sedangkan hasil prediksi terhadap sampel citra buatan AI disajikan pada Tabel 4.

**Tabel 3.** Hasil Prediksi dan Tingkat Probabilitas Sistem terhadap Sampel Citra Karya Ilustrator

| No | Kategori Asli (Sumber) | Prediksi Sistem | Keyakinan Probabilitas | Keterangan Status |
|---|---|---|---|---|
| 1. | Asli Ilustrator | MANUSIA | 77,10% | Benar (Model Yakin) |
| 2. | Asli Ilustrator | MANUSIA | 94,80% | Benar (Model Yakin) |
| 3. | Asli Ilustrator | MANUSIA | 51,50% | Benar (Ragu) |
| 4. | Asli Ilustrator | MANUSIA | 69,90% | Benar (Ragu) |
| 5. | Asli Ilustrator | MANUSIA | 87,00% | Benar (Model Yakin) |

**Tabel 4.** Hasil Prediksi dan Tingkat Probabilitas Sistem terhadap Sampel Citra Buatan AI

| No | Kategori Asli (Sumber) | Prediksi Sistem | Keyakinan Probabilitas | Keterangan Status |
|---|---|---|---|---|
| 1. | Asli AI | BUATAN AI | 79,80% | Benar (Model Yakin) |
| 2. | Asli AI | BUATAN AI | 56,50% | Benar (Ragu) |

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. 
This work is licensed under a Creative Commons Attribution 4.0 International License.

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 6, October 2026 | Hal 2469-2481
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i6.1334

| No | Kategori Asli (Sumber) | Prediksi Sistem | Keyakinan Probabilitas | Keterangan Status |
|---|---|---|---|---|
| 3. | Asli AI | MANUSIA | 73,40% | Salah (Kesalahan Prediksi) |
| 4. | Asli AI | MANUSIA | 66,70% | Salah (Kesalahan Prediksi) |
| 5. | Asli AI | BUATAN AI | 63,20% | Benar (Model Yakin) |

Hasil evaluasi pada Tabel 3 dan Tabel 4 menunjukkan bahwa sistem *Ensemble Deep Learning* berhasil melakukan klasifikasi dengan benar terhadap 8 dari 10 citra yang diuji, sehingga diperoleh *accuracy* sebesar 80,00%. Pada kategori *Real Art*, seluruh 5 citra karya manusia berhasil diklasifikasikan dengan tepat dengan tingkat keberhasilan 100,00% dan rata-rata keyakinan prediksi sebesar 76,06%. Sementara itu, pada kategori *AI-Generated Art*, sistem berhasil mengklasifikasikan 3 dari 5 citra dengan benar dengan tingkat keberhasilan 60,00% dan rata-rata keyakinan sebesar 66,50%. Dua citra AI salah diklasifikasikan sebagai karya manusia, sehingga menghasilkan tingkat kesalahan 40,00%. Hasil ini menunjukkan bahwa sistem lebih konsisten dalam mengenali ilustrasi karya manusia dibandingkan ilustrasi hasil AI pada pengujian individu. Kesalahan klasifikasi pada dua citra AI menunjukkan bahwa ilustrasi hasil *Generative AI* tertentu masih memiliki karakteristik visual yang menyerupai karya ilustrator manusia. Kemiripan tersebut dapat terlihat pada tekstur garis, pencahayaan, saturasi warna, dan detail anatomi sehingga batas visual kedua kelas menjadi lebih sulit dibedakan. Meskipun demikian, sistem tetap memperoleh akurasi 80,00% pada sampel pengujian individu. Hasil tersebut menunjukkan bahwa penggabungan *MobileNetV2*, *ResNet50*, dan *DenseNet121* melalui *Soft Voting* dapat memberikan prediksi yang relatif stabil pada citra dengan karakteristik visual kompleks dan kemiripan tinggi antara karya manusia dan hasil AI.

### 3.5 Pembahasan

Hasil penelitian menunjukkan bahwa penerapan *Ensemble Deep Learning* menggunakan *MobileNetV2*, *ResNet50*, dan *DenseNet121* dengan metode *Soft Voting* mampu memberikan performa klasifikasi yang baik dalam membedakan *Real Art* dan *AI-Generated Art*. Sistem memperoleh *accuracy* sebesar 96,00% dan *ROC-AUC* sebesar 99,06% pada evaluasi internal, sedangkan pengujian eksternal memperoleh *accuracy* sebesar 87,60% dan *ROC-AUC* sebesar 95,09%. Meskipun performa model mengalami penurunan saat diuji pada data eksternal, kemampuan dalam memisahkan kedua kelas tetap terjaga. Hal ini ditunjukkan oleh nilai AUC yang masih tinggi, sehingga model tetap mampu membedakan data secara baik meskipun data tersebut tidak digunakan pada tahap pelatihan dan validasi.

Penurunan *accuracy* dari 96,00% pada evaluasi internal menjadi 87,60% pada pengujian eksternal menunjukkan adanya perbedaan karakteristik data yang belum pernah digunakan sebelumnya. Meskipun demikian, kesalahan klasifikasi masih relatif terkendali, dengan 35 citra *Real Art* diklasifikasikan sebagai *AI-Generated Art* dan 27 citra *AI-Generated Art* sebagai *Real Art*. Pada pengujian individu, model memperoleh *accuracy* 80,00% dan berhasil mengenali seluruh lima citra *Real Art*, tetapi dua dari lima citra *AI-Generated Art* masih salah diklasifikasikan. Hal ini menunjukkan bahwa ilustrasi AI dengan karakteristik visual yang menyerupai karya manusia masih menjadi tantangan utama dalam klasifikasi. Penggunaan *Soft Voting* memungkinkan probabilitas prediksi dari *MobileNetV2*, *ResNet50*, dan *DenseNet121* digabungkan untuk menentukan kelas akhir. Pendekatan ini memanfaatkan karakteristik pembelajaran yang berbeda dari setiap model sehingga prediksi tidak hanya bergantung pada satu arsitektur. Hasil evaluasi internal dengan *accuracy* 96,00% dan *ROC-AUC* 99,06% menunjukkan bahwa kombinasi tersebut mampu menghasilkan klasifikasi yang kuat, sedangkan *ROC-AUC* 95,09% pada data eksternal menunjukkan bahwa kemampuan diskriminasi tetap terjaga ketika model menghadapi data baru. Dengan demikian, *Soft Voting* berkontribusi dalam menghasilkan prediksi yang lebih stabil pada klasifikasi ilustrasi digital.

Hasil penelitian menunjukkan kinerja yang cukup kompetitif dibandingkan dengan penelitian sebelumnya dalam mendeteksi citra *AI-generated*. Shadri et al. [16] memperoleh *accuracy* sebesar 88,00% menggunakan *MobileNetV2*, sedangkan Aria et al. [10] mencapai 96% dengan EfficientNet-B0 dan 98% dengan *Vision Transformer* (ViT). Maulana et al. [13] juga memperoleh *accuracy* validasi sebesar 96,70% menggunakan ResNet-50 dengan *Squeeze-and-Excitation Block*. Penelitian ini menghasilkan *accuracy* 96,00% pada evaluasi internal dan 87,60% pada pengujian eksternal. Perbedaan hasil menunjukkan bahwa performa model dipengaruhi oleh arsitektur, karakteristik dataset, dan kondisi pengujian. Kontribusi penelitian ini terletak pada penerapan *Ensemble Deep Learning* untuk membedakan ilustrasi digital *Real Art* dan *AI-Generated Art* dengan menggabungkan *MobileNetV2*, *ResNet50*, dan *DenseNet121* melalui *Soft Voting*. Berbeda dari penelitian terdahulu yang menggunakan model individual atau pendekatan ensemble pada domain citra berbeda, penelitian ini berfokus pada ilustrasi digital dengan karakteristik artistik dan kemiripan visual tinggi antara karya manusia dan hasil AI. Evaluasi menggunakan data internal, data eksternal, dan sampel individu memungkinkan kinerja model diamati pada beberapa kondisi pengujian. Meskipun beberapa citra AI yang menyerupai karya manusia masih salah diklasifikasikan, *Soft Voting Ensemble* menunjukkan potensi untuk klasifikasi ilustrasi digital.

## 4. KESIMPULAN

Penelitian ini menunjukkan bahwa pendekatan *Ensemble Deep Learning* dengan menggabungkan *MobileNetV2*, *DenseNet121*, dan *ResNet50* melalui metode *Soft Voting* dapat digunakan untuk melakukan klasifikasi ilustrasi digital menjadi *Real Art* dan *AI-Generated Art* pada ilustrasi bergaya anime dan manga. Penggunaan ketiga arsitektur dengan karakteristik pembelajaran yang berbeda memungkinkan probabilitas prediksi dari masing-masing model digabungkan untuk memperoleh keputusan klasifikasi akhir. Pada evaluasi internal menggunakan 5-Fold *Stratified Cross Validation*,

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 2479
This work is licensed under a Creative Commons Attribution 4.0 International License.

sistem memperoleh *accuracy* sebesar 96,00% dan *ROC-AUC* sebesar 99,06%, sedangkan pengujian pada data eksternal memperoleh *accuracy* sebesar 87,60% dan AUC sebesar 0,9509. Perbedaan kinerja tersebut menunjukkan adanya penurunan sebesar 8,40% ketika model menghadapi data yang tidak digunakan dalam proses pelatihan dan validasi. Meskipun demikian, nilai AUC yang tetap tinggi menunjukkan bahwa model masih memiliki kemampuan diskriminasi yang baik pada data eksternal. Hasil tersebut menjadi temuan penting bahwa kinerja tinggi pada evaluasi internal belum sepenuhnya menggambarkan kemampuan generalisasi model terhadap ilustrasi dengan karakteristik visual yang berbeda. Hal tersebut semakin terlihat pada pengujian 10 sampel individu yang menghasilkan *accuracy* sebesar 80,00%, dengan seluruh sampel *Real Art* berhasil diklasifikasikan dengan benar, sedangkan dua dari lima sampel *AI-Generated Art* mengalami kesalahan klasifikasi. Kesalahan tersebut menunjukkan bahwa ilustrasi hasil *Generative AI* yang memiliki karakteristik visual sangat menyerupai karya manusia masih menjadi tantangan bagi sistem klasifikasi. Dengan demikian, kontribusi penelitian ini tidak hanya terletak pada penerapan *Soft Voting Ensemble* untuk klasifikasi ilustrasi digital bergaya anime dan manga melalui kombinasi *MobileNetV2*, *DenseNet121*, dan *ResNet50*, tetapi juga pada evaluasi model menggunakan data internal, data eksternal, dan sampel individu untuk mengamati perbedaan kinerja serta kemampuan generalisasinya. Evaluasi pada beberapa kondisi pengujian tersebut memberikan gambaran mengenai kemampuan model dalam membedakan ilustrasi karya manusia dan hasil *Generative AI*, sekaligus menunjukkan bahwa hasil evaluasi internal belum sepenuhnya mencerminkan kinerja pada data dengan karakteristik berbeda. Keterbatasan penelitian terutama terletak pada kemampuan generalisasi model terhadap variasi ilustrasi yang belum terwakili secara optimal dalam data pelatihan. Oleh karena itu, penelitian selanjutnya dapat diarahkan pada penambahan data dari berbagai platform ilustrasi dan generator AI untuk meningkatkan keragaman karakteristik visual yang dipelajari model, serta melakukan analisis khusus terhadap citra yang salah diklasifikasikan untuk mengidentifikasi karakteristik visual yang menyebabkan kesalahan prediksi. Hasil analisis tersebut dapat menjadi dasar dalam menentukan strategi *preprocessing*, augmentasi data, atau pemilihan arsitektur yang lebih sesuai. Pengujian menggunakan arsitektur lain seperti *Vision Transformer* dan *EfficientNet* juga dapat dilakukan sebagai pembanding untuk mengetahui pengaruh karakteristik arsitektur terhadap kemampuan generalisasi dalam membedakan ilustrasi manusia dan hasil *Generative AI*.

# 5. PERNYATAAN PENULIS

### 5.1 Kontribusi Penulis

Konseptualisasi: I. H. P. dan Y. H.;
Metodologi: I. H. P.;
Perangkat Lunak: I. H. P.;
Validasi: I. H. P. dan Y. H.;
Analisis Formal: I. H. P.;
Investigasi: I. H. P.;
Sumber Daya: I. H. P.;
Kurasi Data: I. H. P.;
Penulisan Draf Awal: I. H. P.;
Penulisan, Peninjauan, dan Penyuntingan: I. H. P. dan Y. H.;
Visualisasi: I. H. P.;
Seluruh penulis telah membaca dan menyetujui versi manuskrip yang telah diterbitkan.

### 5.2 Ketersediaan Data

Data yang disajikan dalam penelitian ini tersedia berdasarkan permintaan kepada penulis korespondensi.

### 5.3 Pendanaan

Penelitian ini tidak menerima pendanaan khusus dari lembaga pemerintah, swasta, maupun sponsor lainnya. Seluruh proses penelitian dan penulisan manuskrip dilakukan secara mandiri oleh penulis.

### 5.4 Pernyataan Dewan Peninjau Institusional

Tidak Berlaku (*Not applicable*).

### 5.5 Pernyataan Persetujuan Berdasarkan Informasi

Penelitian ini menggunakan dataset dari Kaggle dan hasil *web scraping* tanpa melibatkan partisipasi manusia secara langsung. Oleh karena itu, pernyataan persetujuan berdasarkan informasi: Tidak Berlaku (*Not applicable*).

### 5.6 Pernyataan Konflik Kepentingan

Para penulis menyatakan bahwa mereka tidak memiliki kepentingan keuangan yang bersaing atau hubungan pribadi yang diketahui yang dapat dianggap memengaruhi pekerjaan yang dilaporkan dalam artikel ini.

### 5.7 Penggunaan AI Generatif dan Teknologi Berbantuan AI dalam Proses Penulisan

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 2480
This work is licensed under a Creative Commons Attribution 4.0 International License.

Para penulis menggunakan AI Generatif untuk meningkatkan kejelasan penulisan dalam artikel ini. Para penulis telah meninjau dan menyunting konten yang dihasilkan dengan bantuan AI serta bertanggung jawab penuh atas versi akhir publikasi.

## REFERENCES

[1] A. Arifin, A. N. Mubiin, B. A. Setyanto, and I. A. Fatahillah, "Analisis Yuridis Hak Cipta Kepada User Atas Karya Hasil Generative Artificial Intelligence Di Indonesia," *At-Tanwir Law Review*, pp. 412–427, 2025, doi: https://dx.doi.org/10.31314/atlarev.v5i2.4504.

[2] D. W. M. Sandi, P. E. P. Utomo, and U. Khaira, "Perbandingan EfficientNetV2 Dan MobileNetV3 Pada CNN Untuk Klasifikasi Gambar Asli Dan Gambar AI," *JUSS) Jurnal Sains dan Sistem Informasi*, vol. 9, no. 1, 2026, doi: https://doi.org/10.22437/juss.v9i1.58864.

[3] F. Roumpies, G. Spathoulas, and A. Kakarountas, "Synthetically Generated Media, types, detection, and the threat they lurk for users, state and private actors," *Proceedings - 28th Pan-Hellenic Conference on Progress in Computing and Informatics with International Participation, PCI 2024*, pp. 431–436, May 2025, doi: 10.1145/3716554.3716835.

[4] Rachmawati, "Kreativitas atau Plagiarisme? Kontroversi AI dalam Seni Bergaya Studio Ghibli," https://www.kompas.com/jawa-timur/read/2025/04/02/060000488/kreativitas-atau-plagiarisme-kontroversi-ai-dalam-seni-bergaya-studio?page=all. Accessed: Apr. 18, 2026. [Online]. Available: https://www.kompas.com/jawa-timur/read/2025/04/02/060000488/kreativitas-atau-plagiarisme-kontroversi-ai-dalam-seni-bergaya-studio?page=all

[5] A. N. Nafisa, E. N. D. B. Purba, F. A. A. Harahap, and N. A. Putri, "Implementasi Algoritma Convolutional Neural Network Arsitektur Model MobileNetV2 dalam Klasifikasi Penyakit Tumor Otak Glioma, Pituitary dan Meningioma," *Jurnal Teknologi Informasi, Komputer dan Aplikasinya (JTIKA)*, vol. 5, no. 1, pp. 53–61, 2023, doi: https://doi.org/10.29303/jtika.v5i1.234.

[6] O. V. Putra, M. Z. Mustaqim, and D. Muriatmoko, "Transfer Learning untuk Klasifikasi Penyakit dan Hama Padi Menggunakan MobileNetV2 Transfer Learning for Rice Disease and Pest Classification using MobileNetV2," vol. 22, pp. 562–575, 2023, doi: https://doi.org/10.33633/tc.v22i3.8516.

[7] A. Agustina, F. Yanto, E. Budianita, I. Iskandar, and F. Syafria, "Klasifikasi Penyakit Tanaman Padi Menggunakan Metode CNN Arsitektur DenseNet-121 Dan Augmentasi Data," *Journal Of Information Systems And Informatics Engineering*, vol. 8, no. 1, pp. 124–134, Jun. 2024, doi: 10.35145/joisie.v8i1.4256.

[8] S. A. Hakim *et al.*, "Klasifikasi Citra Generasi Artificial Intelligence Menggunakan Metode Fine Tuning Pada Residual Network AI Generated Image Classification Using Fine Tuning On Residual Network," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 11, no. 3, pp. 655–666, 2024, doi: 10.25126/jtiik.938118.

[9] Steven Joses, D. Yulvida, and S. Rochimah, "Pendekatan Metode Ensemble Learning untuk Prakiraan Cuaca menggunakan Soft Voting Classifier," *Journal of Applied Computer Science and Technology*, vol. 5, no. 1, pp. 72–80, Jun. 2024, doi: 10.52158/jacost.v5i1.741.

[10] M. S. A. Aria, C. Slamet, and M. D. Firdaus, "Klasifikasi Fake dan Real Menggunakan Vision Transformer dan EfficientNet-B0 pada Gambar Asli dan Generatif AI," *SMATIKA JURNAL*, vol. 15, no. 01, pp. 179–192, Jun. 2025, doi: 10.32664/smatika.v15i01.1531.

[11] Q. Yun, "Vision Transformers (ViTs) for Feature Extraction and Classification of AI-Generated Visual Designs," *IEEE Access*, vol. 13, pp. 69459–69477, 2025, doi: 10.1109/ACCESS.2025.3562130.

[12] J. J. Bird and A. Lotfi, "CIFAKE: Image Classification and Explainable Identification of AI-Generated Synthetic Images," *IEEE Access*, vol. 12, pp. 15642–15650, 2024, doi: 10.1109/ACCESS.2024.3356122.

[13] D. Maulana, N. S. Azzahrah, R. A. Syach, S. Syauri, M. Syafi'i, and C. Rozikin, "Identifikasi Konten Visual Buatan AI Dengan ResNet Dan Fine-Grained Feature Extraction," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 3S1, pp. 1248–1256, Oct. 2025, doi: 10.23960/jitet.v13i3S1.8056.

[14] N. K. R. Ranupadma, E. M. Dharma, and A. A. I. I. Paramitha, "Implementasi Transfer Learning Dan Augmentasi Pada Model CNN Untuk Meningkatkan Robustness Klasifikasi AI-Generated Image," *JATI: Jurnal Mahasiswa Teknik Informatika*, vol. 10, no. 2, pp. 2370–2377, 2026, doi: https://doi.org/10.36040/jati.v10i2.17412.

[15] H. R. V. Kaluge, N. Yudistira, and Marji, "Implementasi Resnet-50 Dalam Deteksi Anomali Tekstur Pada Gambar Yang Dihasilkan Oleh AI (Artificial Intelligence)," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 10, no. 4, pp. 1–7, Mar. 2026, [Online]. Available: http://j-ptiik.ub.ac.id

[16] M. Shadri, A. Hanafiah, and P. Velusamy, "AI-Generated Image Detection Using Convolutional Neural Network (CNN) Algorithm," *Artificial Intelligence and Language Models*, 2026, [Online]. Available: https://acspub.id/index.php/ailm/article/view/4

[17] Arini, M. Azhari, I. I. A. Fitri, and F. Fahrianto, "Performance Analysis of Transfer Learning Models for Identifying AI-Generated and Real Images," *JURNAL TEKNIK INFORMATIKA*, vol. 17, no. 2, pp. 139–152, Oct. 2024, doi: https://doi.org/10.15408/jti.v17i2.40453.

[18] I. R. Kurniawan *et al.*, "Pengaruh Penambahan Arsitektur Model dalam Klasifikasi Citra Bencana Alam Menggunakan Ensemble Learning," *JURNAL FASILKOM*, vol. 15, no. 2, pp. 280–289, 2025, doi: https://doi.org/10.37859/jf.v15i2.9103.

[19] S. Mildawati and P. T. Prasetyaningrum, "Analisis Peningkatan Performa Convolutional Neural Network Menggunakan Hyperparameter Tuning Dan Ensemble Learning Pada Klasifikasi Citra MRI Tumor Otak," *Rabit : Jurnal Teknologi dan Sistem Informasi Univrab*, vol. 11, no. 2, pp. 4769–4785, Jul. 2026, doi: https://doi.org/10.36341/rabit.v11i2.8341.

[20] A. Rianti, N. W. A. Majid, and A. Fauzi, "CRISP-DM: Metodologi Proyek Data Science," *Prosiding Seminar Nasional Teknologi Informasi dan Bisnis (SENATIB)*, pp. 107–114, 2023, doi: https://doi.org/10.47701/senatib.v3i1.

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. 
This work is licensed under a Creative Commons Attribution 4.0 International License.