ISSN: 2598-8565 (media cetak)
ISSN: 2620-4339 (media online)

PENERAPAN TRANSFER LEARNING DAN EXPLAINABLE AI UNTUK PENGENALAN TOKOH WAYANG BERBASIS WEB

Muhammad Al-Husaini✉, Hen Hen Lukmana, Muhammad Sidik Asyaky, Ichsan Fauzi Rachman
Universitas Siliwangi, Tasikmalaya, Indonesia
Email: alhusaini@unsil.ac.id

DOI: https://doi.org/10.46880/jmika.Vol10No2.pp983-990

ABSTRACT

Recognizing wayang characters remains difficult because many classes share similar visual traits, capture conditions vary, and labeled data is scarce. This study develops a web-based wayang character recognition system that combines transfer learning with visual explanation for classroom use. Two lightweight backbones, EfficientNetV2-S and DeiT-Tiny, were fine-tuned on ten characters using horizontal flipping, RandAugment, and TrivialAugment, then evaluated through five-fold stratified cross-validation. Performance was measured using accuracy, macro-F1, and the Matthews Correlation Coefficient. EfficientNetV2-S reached a mean accuracy of 86.8%, macro-F1 of 0.87, and MCC of 0.85, slightly above DeiT-Tiny at 86.0%, 0.86, and 0.83. Confusion analysis identified recurring errors between visually similar pairs such as Nakula-Sadewa and Gareng-Petruk. Grad-CAM overlays were integrated into a Streamlit interface so that every prediction carries visual evidence that non-experts can inspect. Median inference latency on a mid-range CPU was 28 ms for EfficientNetV2-S and 32 ms for DeiT-Tiny, fast enough for interactive use. The system shows that accuracy, speed, and interpretability can be balanced within the constraints of cultural heritage datasets.

Keywords: Wayang, Transfer Learning, EfficientNetV2, Explainable AI, Grad-CAM.

ABSTRAK

Pengenalan tokoh wayang masih sulit dilakukan karena banyak kelas memiliki ciri visual yang mirip, kondisi pengambilan gambar bervariasi, dan data berlabel sangat terbatas. Penelitian ini membangun sistem pengenalan tokoh wayang berbasis web yang memadukan transfer learning dengan penjelasan visual untuk kebutuhan pembelajaran di kelas. Dua arsitektur ringan, yaitu EfficientNetV2-S dan DeiT-Tiny, dilatih ulang pada sepuluh tokoh menggunakan augmentasi horizontal flip, RandAugment, dan TrivialAugment, lalu dievaluasi dengan five-fold stratified cross-validation. Kinerja diukur menggunakan accuracy, macro-F1, dan Matthews Correlation Coefficient. EfficientNetV2-S memperoleh rata-rata accuracy 86,8%, macro-F1 0,87, dan MCC 0,85, sedikit di atas DeiT-Tiny dengan 86,0%, 0,86, dan 0,83. Analisis confusion matrix menunjukkan kesalahan yang berulang pada pasangan tokoh yang mirip seperti Nakula-Sadewa dan Gareng-Petruk. Overlay Grad-CAM diintegrasikan ke antarmuka Streamlit sehingga setiap prediksi disertai bukti visual yang dapat diperiksa oleh pengguna awam. Latensi inferensi median pada CPU kelas menengah adalah 28 ms untuk EfficientNetV2-S dan 32 ms untuk DeiT-Tiny, cukup cepat untuk penggunaan interaktif. Sistem ini menunjukkan bahwa akurasi, kecepatan, dan keterjelasan dapat diseimbangkan dalam keterbatasan data warisan budaya.

Kata Kunci: Wayang, Transfer Learning, EfficientNetV2, Explainable AI, Grad-CAM.

PENDAHULUAN

Wayang merupakan salah satu warisan budaya Indonesia yang paling dikenal, namun pengenalan tokohnya secara otomatis masih menjadi persoalan yang sulit. Kesulitan tersebut muncul karena banyak tokoh memiliki kemiripan bentuk wajah, mahkota, dan aksesoris, kondisi pengambilan gambar sangat beragam, serta jumlah citra berlabel yang tersedia relatif sedikit (Irawan et al., 2025; Widaningrum et al., 2025). Tiga kendala ini membuat model pengenalan citra umum tidak langsung dapat dipakai tanpa penyesuaian.

Di sisi lain, kebutuhan terhadap alat bantu pengenalan tokoh wayang justru meningkat seiring digitalisasi koleksi budaya dan pemanfaatan teknologi di ruang kelas. Guru dan siswa memerlukan perangkat yang dapat diakses dengan mudah, memberi jawaban cepat, dan dapat dipahami oleh pengguna awam (Khosravi et al., 2022; Kritandani et al., 2024). Penelitian terdahulu menunjukkan bahwa model pembelajaran mendalam mampu bekerja baik pada

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

dataset wayang tertentu, tetapi hasilnya berubah-ubah bergantung pada jumlah kelas dan keseimbangan datanya (Kusrini et al., 2022; Supriyanti & Anggoro, 2021).

Arsitektur berukuran kecil seperti EfficientNetV2 dan transformer hemat data seperti DeiT menjadi pilihan yang masuk akal untuk kebutuhan ini karena menjaga kualitas prediksi sambil menekan beban komputasi, sehingga layak dipasang pada layanan web (Tan & Le, 2021; Touvron et al., 2021). Untuk penggunaan berbasis web, kecepatan dan kestabilan respons sama pentingnya dengan skor agregat, karena waktu tunggu yang panjang memutus alur pembelajaran (Mendoza et al., 2024; Nigade et al., 2024).

Ukuran dataset warisan budaya yang kecil juga menuntut metrik evaluasi yang peka terhadap ketidakseimbangan kelas. Karena itu penelitian ini memakai macro-F1 dan Matthews Correlation Coefficient (MCC) serta melaporkan ringkasan per fold untuk menekan variansi akibat pembagian data (Chicco & Jurman, 2020; Takahashi et al., 2022). Daya generalisasi ditingkatkan melalui augmentasi sederhana namun kuat dan proses fine-tuning dengan decoupled weight decay (Cubuk et al., 2020; Loshchilov & Hutter, 2019; Müller & Hutter, 2021).

Tujuan penelitian ini adalah membangun sistem pengenalan tokoh wayang berbasis web yang akurat, responsif, dan dapat dijelaskan. Kontribusi yang ditawarkan ada tiga, yaitu: (i) alur kerja lengkap dari penyiapan data sampai antarmuka Streamlit; (ii) laporan evaluasi silang dengan macro-F1, MCC, dan analisis confusion; serta (iii) visualisasi bergaya Class Activation Mapping untuk keperluan pengajaran dan analisis kesalahan.

Kebaruan penelitian ini terletak pada penggabungan tiga hal yang pada penelitian sebelumnya umumnya berdiri sendiri, yaitu arsitektur ringan yang hemat komputasi, evaluasi berbasis metrik seimbang kelas, dan penjelasan visual yang langsung tersaji di antarmuka pengguna. Dengan demikian prediksi sistem tidak hanya menghasilkan label, tetapi juga bukti visual yang dapat ditelaah guru maupun siswa.

TINJAUAN PUSTAKA

Pengenalan Citra Wayang

Penelitian pengenalan citra wayang sudah dilakukan dengan beragam pendekatan. (Widaningrum et al., 2025) memakai Principal Component Analysis untuk ekstraksi fitur dan K-Nearest Neighbor untuk klasifikasi pada tokoh Pandawa Lima. (Supriyanti & Anggoro, 2021) mengklasifikasikan tokoh Pandawa dengan convolutional neural network, sementara (Kusrini et al., 2022) Kusrini et al. (2022) menunjukkan pengaruh Gaussian filter dan pra-pemrosesan terhadap klasifikasi citra Punakawan. (Irawan et al., 2025) menggabungkan ResNet-50 dan LSTM melalui lapisan ensemble, dan (Sudiatmika et al., 2021) menerapkan Mask R-CNN untuk mengenali sekaligus melokalisasi tokoh wayang.

Lima penelitian tersebut menyisakan persoalan yang berulang, yaitu bentuk tokoh yang serupa, variasi pose dan pencahayaan, serta cakupan kelas yang belum merata (Tena et al., 2023). Hal ini menegaskan pentingnya penataan dan standardisasi data sebelum model dibangun.

Arsitektur Ringan untuk Klasifikasi Citra

EfficientNetV2 memakai progressive learning dan blok yang ringkas sehingga memberi keseimbangan antara akurasi dan kecepatan (Tan & Le, 2021). DeiT menghadirkan perhatian berbasis token dengan distilasi yang menjaga waktu proses tetap singkat (Touvron et al., 2021). Arsitektur hibrida seperti Swin Transformer juga bekerja baik pada anggaran komputasi yang ketat (Liu et al., 2021), dan survei arsitektur merangkum pertimbangan pemilihan backbone pada batas sumber daya yang lazim (Han et al., 2024; Lee et al., 2023; Papa et al., 2024).

Untuk layanan web, kecepatan tidak hanya ditentukan oleh FLOPs model, tetapi juga oleh proses input/output, tingkat konkurensi, dan kondisi jaringan. Karena itu pelaporan distribusi latensi lebih informatif daripada satu angka tunggal (Mendoza et al., 2024; Nigade et al., 2024).

Augmentasi dan Strategi Pelatihan

Kebijakan augmentasi yang mudah disetel seperti RandAugment (Cubuk et al., 2020) dan TrivialAugment (Müller & Hutter, 2021) terbukti meningkatkan kinerja dengan mengubah tampilan citra tanpa mengubah identitas semantiknya. Pendekatan ini berguna untuk koleksi warisan budaya yang berukuran kecil. Pada tahap fine-tuning, optimizer AdamW membantu menjaga kestabilan konvergensi melalui weight decay yang terpisah dari gradien, baik untuk lapisan bergaya transformer maupun CNN modern (Loshchilov & Hutter, 2019; Papa et al., 2024).

Explainable AI dan Metrik Evaluasi

Metode saliency yang bersifat class-discriminative menampilkan wilayah citra yang mendasari keputusan model sehingga dapat diperiksa oleh pakar maupun pembelajar (Wang et al., 2020; Zhang et al., 2024). Dalam konteks warisan budaya,

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

overlay yang dapat dijelaskan membantu proses dokumentasi, penelaahan, dan pembelajaran karena logika model menjadi terlihat dan dapat didiskusikan (Ardiyasa, 2020; Khosravi et al., 2022).

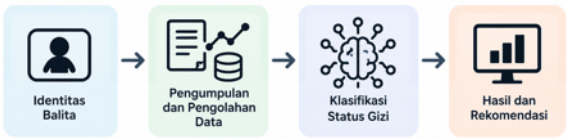
Di luar akurasi top-1, macro-F1 menyeimbangkan precision dan recall tiap kelas, sedangkan MCC memberi ringkasan menyerupai korelasi yang peka terhadap seluruh isi confusion matrix (Chicco & Jurman, 2020). Pelaporan berbasis fold dan ringkasan yang menyertakan sebaran dinilai lebih dapat dipercaya pada koleksi kecil (Takahashi et al., 2022; Wang et al., 2020).

Celah Penelitian

Penelitian terdahulu umumnya menonjolkan salah satu aspek saja. Studi pengenalan wayang berfokus pada peningkatan akurasi tanpa melaporkan latensi atau menyediakan penjelasan visual (Irawan et al., 2025; Supriyanti & Anggoro, 2021). Studi arsitektur ringan membahas efisiensi tetapi tidak diuji pada domain wayang yang kelasnya saling mirip (Tan & Le, 2021; Touvron et al., 2021). Studi explainable AI menyediakan metode saliency namun jarang diintegrasikan ke sistem siap pakai untuk kelas. Penelitian ini mengisi celah tersebut dengan menyatukan ketiganya dalam satu alur kerja yang diuji secara silang dan dipasang pada antarmuka web.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif eksperimental dengan alur kerja dari penyiapan data sampai penerapan model pada antarmuka web. Alur lengkapnya ditunjukkan pada Gambar 1.



Gambar 1. Alur Kerja dari Dataset Sampai Penerapan Sistem

Data dan Kelas

Penelitian menggunakan sepuluh tokoh wayang, yaitu Arjuna, Bima, Yudhistira, Nakula, Sadewa, Srikandi, Semar, Gareng, Petruk, dan Bagong berasal dari dataset Kaggle (Achmad, n.d.). Tokoh-tokoh ini dipilih karena mudah dikenali publik, namun kemiripan antaranggota kelompok, seperti pada Pandawa dan Punakawan, membuat pembedaannya tetap menantang (Irawan et al., 2025; Kusrini et al., 2022).

Evaluasi memakai five-fold stratified cross-validation. Pada skema five-fold yang konsisten, setiap putaran menggunakan satu dari lima fold 20% data sebagai data uji, sedangkan empat fold lain 80% menjadi data pengembangan (Takahashi et al., 2022). Seluruh citra yang berasal dari sumber yang sama ditempatkan pada fold yang sama untuk mencegah kebocoran data. Keseimbangan kelas diperiksa ulang setelah setiap pembagian, dan hasil dilaporkan beserta sebarannya dalam bentuk rata-rata dan simpangan baku.

Pra-pemrosesan dan Augmentasi

Seluruh citra diubah ukurannya menjadi 224 x 224 piksel agar sesuai dengan kebutuhan data pra-pelatihan. Standardisasi dilakukan dengan menyesuaikan rata-rata dan variansi tiap kanal mengikuti protokol masing-masing backbone.

Untuk menekan overfitting dan meningkatkan generalisasi, diterapkan augmentasi berupa horizontal flipping, RandAugment, dan TrivialAugment. Ketiganya menghadirkan variasi warna, geometri, dan tekstur yang membantu pada kasus tokoh dengan bentuk dan ornamen serupa (Cubuk et al., 2020; Müller & Hutter, 2021).

Kebijakan pra-pemrosesan dan augmentasi dijaga tetap sama pada seluruh percobaan. Dengan parameter, algoritma, dan pengaturan yang seragam, perbedaan hasil yang teramati dapat diatribusikan pada variasi data dan arsitektur, bukan pada perbedaan perlakuan citra. Konsistensi ini sekaligus memudahkan peneliti lain mengulang percobaan yang sama.

Model

Dua backbone pra-terlatih di-fine-tune, yaitu EfficientNetV2-S dan DeiT-Tiny. EfficientNetV2-S memakai pelatihan bertahap dan blok yang ringan untuk perangkat bergerak sehingga memberi keseimbangan akurasi dan kecepatan yang cocok untuk aplikasi web interaktif (Tan & Le, 2021). DeiT-Tiny memanfaatkan perhatian berbasis token dan distilasi untuk mencapai akurasi serupa dengan ukuran model kecil dan waktu proses yang memadai pada CPU (Touvron et al., 2021).

Pemilihan dua arsitektur ini memungkinkan perbandingan antara lokalitas konvolusi dan self-attention global. Dari perbandingan tersebut dapat ditelaah apakah detail halus pada citra wayang, seperti bentuk wajah atau tepi mahkota, lebih baik dikenali melalui wilayah lokal atau melalui interaksi antar token, dengan prosedur pelatihan yang dijaga identik.

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Pelatihan

Pelatihan memakai optimizer AdamW (Loshchilov & Hutter, 2019) dengan weight decay 1e-4 dan label smoothing sebesar 0,1. Pembaruan parameter pada AdamW mengikuti persamaan (1).

$$\theta_{t+1} = \theta_t - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t}+\varepsilon} - \eta\lambda\theta_t \qquad (1)$$

Pada Persamaan (1), θ_t menunjukkan parameter pada langkah t, η adalah learning rate, $\hat{m}_t$ dan $\hat{v}_t$ merupakan estimasi momen pertama dan kedua yang telah dikoreksi bias, ε konstanta stabilitas numerik, serta λ koefisien weight decay. Suku terakhir merepresentasikan penyusutan bobot yang terpisah dari gradien. Dengan definisi ini, Persamaan (1) menyatakan pembaruan AdamW secara ringkas.

Label smoothing sebesar 0,1 berarti label target dicampur dengan 10% distribusi seragam. Cara ini menurunkan keyakinan berlebih pada prediksi dan membantu generalisasi ketika data latih belum sepenuhnya mewakili kondisi nyata. Untuk mengendalikan overfitting sekaligus biaya komputasi, pelatihan dihentikan lebih awal apabila macro-F1 pada data validasi tidak lagi membaik setelah sejumlah epoch tertentu.

Metrik Evaluasi

Kinerja diukur dengan accuracy, macro-F1, dan Matthews Correlation Coefficient. Macro-F1 menangani ketidakseimbangan kelas dengan merata-ratakan precision dan recall seluruh kelas. MCC memberi ringkasan confusion matrix yang memperhitungkan seluruh sel matriks, bukan hanya prediksi yang benar (Chicco & Jurman, 2021). MCC dihitung dengan persamaan (2).

$$\mathrm{MCC} = \frac{N\sum_k c_{kk} - \sum_k r_k c_k}{\sqrt{(N^2 - \sum_k r_k^2)(N^2 - \sum_k c_k^2)}} \qquad (2)$$

Keterangan: C adalah confusion matrix; r dan c berturut-turut adalah jumlah baris dan jumlah kolom; serta N adalah total data. Untuk memperhitungkan variabilitas sampel pada dataset kecil, hasil disajikan sebagai estimasi titik beserta sebaran antar fold dalam bentuk rata-rata dan simpangan baku.

Penerapan Sistem

Model hasil pelatihan dikonversi ke format yang efisien untuk inferensi lalu dipasang di balik antarmuka Streamlit. Antarmuka ini menyediakan empat fungsi, yaitu: (i) mengunggah citra; (ii) menampilkan prediksi teratas beserta skor keyakinan; (iii) mengaktifkan Grad-CAM sebagai penjelasan visual; dan (iv) menampilkan waktu proses permintaan terakhir.

Overlay saliency dibuat pada blok konvolusi terakhir untuk CNN, atau pada komponen penjelasan yang setara untuk transformer. Overlay disajikan sebagai heatmap semitransparan sehingga prediksi dapat ditelaah tanpa menutupi detail penting citra (Wang et al., 2020). Bobot model, langkah pra-pemrosesan, dan hash kebijakan augmentasi disertakan dalam paket penerapan agar proses inferensi dapat direproduksi persis seperti pada tahap evaluasi.

HASIL DAN PEMBAHASAN

Kinerja Keseluruhan Antar Fold

EfficientNetV2-S unggul tipis atas DeiT-Tiny dalam accuracy dan MCC pada seluruh fold yang diuji, sedangkan macro-F1 keduanya setara pada fold 1. Accuracy EfficientNetV2-S berkisar 85,8% sampai 87,6%, macro-F1 antara 0,86 dan 0,88, serta MCC antara 0,83 dan 0,86. DeiT-Tiny mencatat accuracy 85,1% sampai 86,6%, macro- F1 0,85 sampai 0,86, dan MCC 0,82 sampai 0,84. Rentang yang sempit pada ketiga metrik menunjukkan kemampuan generalisasi yang stabil untuk kedua model. Rinciannya disajikan pada Tabel 1.

Tabel 1. Kinerja per Fold

Fold	EffNetV2-S			DeiT-Tiny		
	Acc (%)	Macro-F1	MCC	Acc (%)	Macro-F1	MCC
1	86,1	0,86	0,84	85,9	0,86	0,83
2	87,5	0,88	0,86	86,6	0,86	0,84
3	85,8	0,86	0,83	85,1	0,85	0,82
4	86,9	0,87	0,85	86,2	0,86	0,83
5	87,6	0,88	0,86	86,3	0,86	0,84

Ringkasan rata-rata dan simpangan baku antar fold disajikan pada Tabel 2. Rata-rata accuracy kedua backbone berada pada kisaran 86% sampai 87% dengan simpangan baku yang kecil, yaitu tidak lebih dari 0,8 poin persen untuk accuracy dan tidak lebih dari 0,01 untuk macro-F1 maupun MCC. Angka ini menguatkan temuan pada Tabel 1 bahwa kedua pendekatan konsisten antar fold, dengan keunggulan tipis pada EfficientNetV2-S.

Tabel 2. Rata-rata dan Simpangan Baku Antar Fold

Model	Accuracy (%) ± SD	Macro-F1 ± SD	MCC ± SD
EffNetV2-S	86,8 ± 0,8	0,87 ± 0,01	0,85 ± 0,01
DeiT-Tiny	86,0 ± 0,6	0,86 ± 0,00	0,83 ± 0,01

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

ISSN: 2598-8565 (media cetak)
ISSN: 2620-4339 (media online)

Selisih yang kecil antara kedua model menunjukkan bahwa kapasitas model dan mekanisme pencegahan overfitting sudah seimbang dengan jumlah data yang tersedia. Dengan kata lain, perbaikan pada sisi data, baik kualitas maupun cakupannya, berpeluang memberi dampak lebih besar daripada mengganti arsitektur. Nilai MCC yang bertahan pada kisaran 0,82 sampai 0,86 juga menandakan bahwa kinerja tidak hanya terkonsentrasi pada kelas mayoritas.

Perilaku Per Kelas

Tabel 3 menyajikan precision, recall, dan F1 tiap tokoh pada satu fold representatif menggunakan EfficientNetV2-S. Skor antarkelas relatif merata, dengan sebagian besar F1 berada pada rentang 0,85 sampai 0,90.

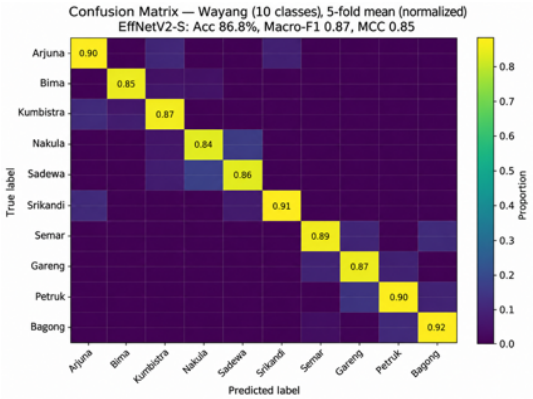
Tabel 3. Precision, Recall, dan F1 per Kelas (Fold Representatif, EffNetV2-S)

Kelas	Precision	Recall	F1-score
Arjuna	0,91	0,89	0,90
Bima	0,90	0,88	0,89
Yudhistira	0,88	0,87	0,88
Nakula	0,85	0,84	0,85
Sadewa	0,86	0,84	0,85
Srikandi	0,89	0,88	0,88
Semar	0,90	0,89	0,90
Gareng	0,88	0,87	0,88
Petruk	0,89	0,85	0,87
Bagong	0,90	0,88	0,89

Arjuna dan Semar memperoleh F1 dan recall tertinggi (0,90 dan 0,89) karena bentuk dan atribut keduanya paling khas, dengan precision tertinggi dicapai Arjuna (0,91). Nakula dan Sadewa justru memiliki recall terendah, yang sejalan dengan kemiripan visual keduanya. Kelompok Punakawan secara umum bekerja baik, meskipun Petruk menunjukkan recall sedikit lebih rendah dengan precision yang tetap tinggi. Pola ini mengindikasikan kesalahan berupa tokoh yang terlewat, bukan tokoh yang tertukar secara meluas. Temuan per kelas tersebut mengarahkan prioritas penambahan data pada pasangan tokoh yang mirip, terutama Nakula dan Sadewa serta beberapa pasangan Pandawa lainnya.

Struktur Confusion Matrix

Nilai terbesar pada confusion matrix terkumpul di diagonal utama, sesuai dengan metrik agregat pada Tabel 1 dan Tabel 2. Pengelompokan di luar diagonal paling kentara terjadi pada saudara Pandawa dan sebagian pasangan Punakawan, khususnya Petruk dan Gareng. Kesalahan ini umumnya muncul ketika bentuk wajah, mahkota, atau aksesoris tampak serupa pada kondisi citra yang kurang ideal.

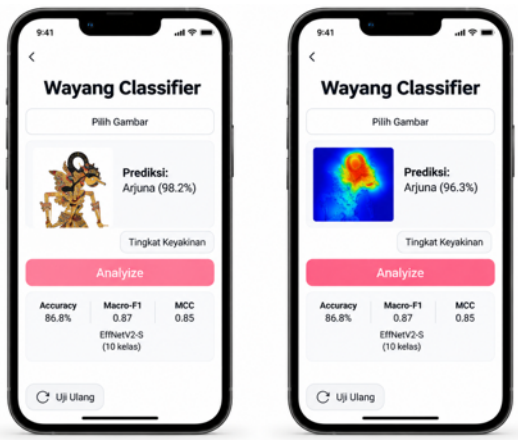


Gambar 2. Confusion Matrix EfficientNetV2-S pada Fold Representatif

Dari pola tersebut dapat ditarik dua langkah perbaikan. Pertama, memperkaya data pada pasangan bermasalah dengan variasi pencahayaan, sudut pandang, dan kasus tertutup sebagian. Kedua, menerapkan penyesuaian pada tingkat fitur atau pelatihan yang difokuskan pada wilayah tertentu agar bagian paling membedakan mendapat penekanan lebih besar.

Penjelasan Visual dan Analisis Kesalahan

Overlay Grad-CAM membantu memahami dasar prediksi, baik ketika hasilnya benar maupun salah. Pada klasifikasi yang benar, perhatian model umumnya tertuju pada puncak mahkota, bentuk wajah, dan aksesoris khas. Pada kasus yang keliru atau berselisih tipis, heatmap cenderung menyebar atau berpindah ke latar belakang. Pola ini konsisten dengan kesalahan yang terlihat pada Gambar 2.



Gambar 3. Antarmuka Streamlit dan overlay Grad-CAM

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Dalam pengujian sistem, tampilan bukti visual ini berguna untuk dua hal, yaitu memeriksa kesalahan prediksi positif maupun negatif, dan menyusun petunjuk pembelajaran singkat bagi siswa. Gabungan antara angka pada Tabel 1 sampai Tabel 3 dengan pemeriksaan visual pada Gambar 2 dan Gambar 3 membentuk prosedur analisis kesalahan yang mudah diikuti di ruang kelas. Guru dan siswa dapat melihat secara langsung bagian citra mana yang memengaruhi keputusan model, sehingga proses pembelajaran mesin yang semula abstrak menjadi bahan diskusi yang konkret. Penilaian interpretabilitas memerlukan desain yang memisahkan tampilan heatmap dari kualitas penjelasan. Pada versi saat ini, Gambar 3 merupakan bukti bahwa antarmuka menampilkan heatmap, bukan bukti kesetiaan terhadap keputusan model. Pengujian dengan pakar dan metrik kuantitatif merupakan pekerjaan lanjutan, bukan hasil eksperimen yang telah dilakukan

Latensi Inferensi

Tabel 4 menunjukkan waktu proses satu citra pada CPU kelas menengah. EffNetV2-S memerlukan waktu median sekitar 28 ms dengan persentil ke-95 sekitar 41 ms, sedangkan DeiT-Tiny sedikit lebih lambat pada median 32 ms dan persentil ke-95 sekitar 46 ms. Waktu ini cukup cepat untuk penggunaan interaktif melalui antarmuka Streamlit dan masih menyisakan ruang untuk penyiapan data serta perenderan tampilan.

Tabel 4. Ringkasan Latensi pada CPU Kelas Menengah

Model	Latensi Median (ms)	Latensi p95 (ms)	Batch Size	Resolusi
EffNetV2-S	28	41	1	224 × 224
DeiT-Tiny	32	46	1	224 × 224

Selisih yang kecil antara median dan persentil ke-95 menandakan waktu eksekusi yang relatif stabil tanpa pencilan yang mengganggu. Kestabilan ini penting pada situasi dengan tuntutan waktu nyata, seperti presentasi atau kegiatan belajar di kelas, karena jeda yang terlalu lama memutus alur penyampaian materi.

Pada perangkat dengan sumber daya terbatas, kecepatan masih dapat ditingkatkan melalui kuantisasi int8 yang mengubah format bilangan titik mengambang menjadi bilangan bulat 8 bit, atau melalui pruning yang menghapus parameter dengan kontribusi kecil. Keduanya berpotensi menurunkan akurasi dalam derajat yang bergantung pada data dan tugasnya, sehingga perlu divalidasi ulang pada data uji yang sulit sebelum digunakan.

Keterbatasan

Sistem ini masih menyimpan sejumlah keterbatasan. Jumlah kelas terbatas pada sepuluh tokoh sehingga belum mencakup keragaman wayang secara utuh. Kesalahan pada pasangan tokoh yang mirip belum sepenuhnya teratasi dan masih bergantung pada ketersediaan data tambahan. Selain itu, skor keyakinan model belum dikalibrasi, sehingga angka keyakinan yang tinggi tidak selalu mencerminkan peluang benar yang sepadan. Karena itu sistem ini diposisikan sebagai alat bantu, bukan pengganti penelaahan ahli.

KESIMPULAN

Penelitian ini menghasilkan sistem pengenalan tokoh wayang berbasis web yang memadukan transfer learning dengan penjelasan visual. Pada sepuluh kelas tokoh, EfficientNetV2-S mencapai rata-rata accuracy 86,8%, macro-F1 0,87, dan MCC 0,85 melalui five-fold stratified cross-validation, sedikit di atas DeiT-Tiny yang memperoleh 86,0%, 0,86, dan 0,83. Latensi inferensi median sebesar 28 ms dan 32 ms pada CPU kelas menengah menunjukkan keduanya layak dipakai secara interaktif.

Analisis confusion matrix memperlihatkan kesalahan terkonsentrasi pada pasangan tokoh yang mirip, yaitu Nakula dengan Sadewa serta Petruk dengan Gareng. Overlay Grad-CAM pada antarmuka Streamlit membuat setiap prediksi disertai bukti visual yang dapat ditelaah guru dan siswa, sehingga analisis kesalahan dapat dijadikan bahan pembelajaran.

Selisih kinerja yang tipis antara kedua arsitektur menunjukkan bahwa perbaikan data lebih menjanjikan daripada penggantian model. Penelitian lanjutan diarahkan pada perluasan cakupan kelas, penambahan citra untuk subkelas yang mirip beserta variasi pencahayaan, pose, dan oklusi, serta kalibrasi skor keyakinan agar angka keyakinan lebih selaras dengan peluang benar. Sistem otomatis seperti ini perlu menyatakan batas kemampuannya secara terbuka dan tidak menggantikan kurasi ahli, sehingga kerja sama dengan komunitas budaya tetap diperlukan untuk menjaga ketepatan representasi.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada komunitas sumber terbuka serta para peneliti terdahulu yang dataset dan kajiannya menjadi rujukan dalam penelitian ini.

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

DAFTAR PUSTAKA

Ardiyasa, I. P. (2020). Wayang (puppet) character transformation model in augmented reality. *Lekesan: Interdisciplinary Journal of Asia Pacific Arts, 3*(2), 54–63. https://doi.org/10.31091/lekesan.v3i2.1141

Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics, 21*(1), 1–13. https://doi.org/10.1186/s12864-019-6413-7

Chicco, D., & Jurman, G. (2021). An ensemble learning approach for enhanced classification of patients with hepatitis and cirrhosis. *IEEE Access*, 9, 24485–24498. https://doi.org/10.1109/ACCESS.2021.3057.196

Cubuk, E. D., Zoph, B., Shlens, J., & Le, Q. V. (2020). RandAugment: Practical automated data augmentation with a reduced search space. *Advances in Neural Information Processing Systems, 33*, 18613–18624.

Han, K., Wang, Y., Guo, J., & Wu, E. (2024). ParameterNet: Parameters are all you need for large-scale visual pretraining of mobile networks. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,* 15751–15761. https://doi.org/10.1109/CVPR52733.2024.01491

Irawan, C., Rachmawanto, E. H., & Hadi, H. P. (2025). Ensemble learning layer for wayang recognition using CNN-based ResNet-50 and LSTM. *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control, 10*(1). https://doi.org/10.22219/kinetik.v10i1.2053

Khosravi, H., Shum, S. B., Chen, G., Conati, C., Tsai, Y.-S., Kay, J., Knight, S., Martinez-Maldonado, R., Sadiq, S., & Gašević, D. (2022). Explainable Artificial Intelligence in education. Computers and Education: *Artificial Intelligence, 3*, 100074. https://doi.org/https://doi.org/10.1016/j.caeai.2022.100074

Kritandani, W., Aryani, R., & Rakasiwi, T. (2024). A report review: Artificial intelligence and the future of teaching and learning. *International Research Education Journal, 6*(2), 245–253. https://doi.org/10.17977/um043v6i2p245-253

Kusrini, K., Yudianto, M. R. A., & Al Fatta, H. (2022). The effect of Gaussian filter and data preprocessing on the classification of Punakawan puppet images with the convolutional neural network algorithm. *International Journal of Electrical and Computer Engineering, 12*(4), 3752–3761. https://doi.org/10.11591/ijece.v12i4.pp3752-3761

Lee, J., Mukhanov, L., Molahosseini, A. S., Minhas, U., Hua, Y., del Rincon, J., Dichev, K., Hong, C.-H., & Vandierendonck, H. (2023). Resource-Efficient Convolutional Networks: A Survey on Model-, Arithmetic-, and Implementation-Level Techniques. *ACM Comput. Surv., 55*(13s). https://doi.org/10.1145/3587095

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., & Guo, B. (2021). Swin Transformer: Hierarchical vision transformer using shifted windows. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 9992–10002.

Loshchilov, I., & Hutter, F. (2019). Decoupled weight decay regularization. *Proceedings of the International Conference on Learning Representations (ICLR).*

Mendoza, D., Romero, F., & Trippel, C. (2024). Model selection for latency-critical inference serving. *Proceedings of the Nineteenth European Conference on Computer Systems*, 1016–1038.

Müller, S. G., & Hutter, F. (2021). TrivialAugment: Tuning-free yet state-of-the-art data augmentation. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 774–782.

Nigade, V., Bauszat, P., Bal, H., & Wang, L. (2024). Inference serving with end-to-end latency SLOs over dynamic edge networks. Real-Time Systems, 60(2). https://doi.org/10.1007/s11241-024-09418-4

Papa, L., Russo, P., Amerini, I., & Zhou, L. (2024). A survey on efficient vision transformers: Algorithms, techniques, and performance benchmarking. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 46*(12), 7682–7700. https://doi.org/10.1109/TPAMI.2024.3392941

Sudiatmika, I. B. K., Artana, M., Utami, N. W., Putra, M. A. P., & Dewi, E. G. A. (2021). Mask R-CNN for Indonesian shadow puppet recognition and classification. *Journal of Physics: Conference Series, 1783*(1), 12032. https://doi.org/10.1088/1742-6596/1783/1/012032

Supriyanti, W., & Anggoro, D. A. (2021). Classification of Pandavas Figure in Shadow Puppet Images using Convolutional Neural Networks. *Khazanah Informatika: Jurnal Ilmu Komputer dan Informatika, 7*(1), 18–24. https://doi.org/10.23917/khif.v7i1.12484

Takahashi, K., Yamamoto, K., Kuchiba, A., & Koyama, T. (2022). Confidence interval for micro-averaged F1 and macro-averaged F1 scores. *Applied Intelligence, 52*(5), 4961–4972. https://doi.org/10.1007/s10489-021-02635-5

Tan, M., & Le, Q. V. (2021). EfficientNetV2: Smaller models and faster training. *Proceedings of Machine Learning Research, 139*, 10096–10106.

Tena, S., Hartanto, R., & Ardiyanto, I. (2023). Content-Based Image Retrieval for Traditional Indonesian Woven Fabric Images Using a

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

ISSN: 2598-8565 (media cetak)
ISSN: 2620-4339 (media online)

Modified Convolutional Neural Network Method. *Journal of Imaging, 9*(8). https://doi.org/10.3390/jimaging9080165

Touvron, H., Cord, M., Douze, M., Massa, F., Sablayrolles, A., & Jégou, H. (2021). Training data-efficient image transformers and distillation through attention. *Proceedings of Machine Learning Research, 139*, 10347–10357.

Wang, H., Wang, Z., Du, M., Yang, F., Zhang, Z., Ding, S., Mardziel, P., & Hu, X. (2020). Score-CAM: Score-Weighted Visual Explanations for Convolutional Neural Networks. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW),*111–119. https://doi.org/10.1109/CVPRW50498.2020.00020

WIdaningrum, I., Astuti, I. P., Mustikasari, D., Nurfitri, K., Az-Zahra, R. R., Vidyastari, R. I., & Selamat, A. (2025). Recognition of the Lima Pandawa Shadow Puppet characters utilizing Principal Component Analysis (PCA) for feature extraction and K-Nearest Neighbor (KNN) for classification. *Indonesian Journal of Information Systems, 8*(1), 68-77. https://doi.org/10.24002/ijis.v8i1.11032

Zhang, H., Torres, F., Sicre, R., Avrithis, Y., & Ayache, S. (2024). Opti-CAM: Optimizing saliency maps for interpretability. *Computer Vision and Image Understanding, 248,* 104101. https://doi.org/10.1016/j.cviu.2024.104101

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.