JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 7 Nomor 3 Juni 2026, Page 337-345
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

Klasifikasi Citra Patung Dewa dan Dewi Menggunakan Support Vector Machine Berbasis Fitur CNN dan BoVW

Komang Budiarta[1], Gede Herdian Setiawan[2], Rosalia Hadi[3], Ketut Gus Oka Ciptahadi[4], I Gusti Ngurah Ady Kusuma[5]

[1,3,4)]Sistem Informasi, Intitut Teknologi dan Bisnis STIKOM Bali, Bali, Indonesia
[2,5)]Sistem Komputer, Intitut Teknologi dan Bisnis STIKOM Bali, Bali, Indonesia

Email: [1]komang_budiarta@stikom-bali.a.id, [2]herdian@stikom-bali.a.id, [3]rosa@stikom-bali.a.id, [4]okaciptahadi@stikom-bali.a.id, [5]ady_kusuma@stikom-bali.a.id
Email Penulis Korespondensi: [1]komang budiarta@stikom-bali.a.id

Abstrak− Patung dewa dan dewi dalam tradisi Hindu Bali memiliki keragaman bentuk dan detail visual yang tinggi, sehingga menimbulkan tantangan dalam proses pengenalan dan klasifikasi citra digital. Perbedaan karakteristik visual antar patung mendorong perlunya metode ekstraksi fitur yang mampu merepresentasikan informasi citra secara optimal. Penelitian ini bertujuan membandingkan kinerja ekstraksi fitur Bag of Visual Words (BoVW) dan Convolutional Neural Network (CNN) pada klasifikasi citra patung dewa dan dewi menggunakan Support Vector Machine (SVM). Dataset terdiri dari 150 citra patung yang dibagi menjadi data latih dan data uji. Proses penelitian meliputi preprocessing citra, ekstraksi fitur menggunakan BoVW dan CNN, serta klasifikasi menggunakan SVM. Evaluasi performa dilakukan menggunakan metrik akurasi, precision, dan recall. Hasil pengujian menunjukkan bahwa kombinasi CNN–SVM mencapai akurasi sebesar 84,44% dengan 38 prediksi benar, lebih tinggi dibandingkan BoVW–SVM yang memperoleh akurasi 77,78% dengan 35 prediksi benar. CNN menunjukkan kemampuan yang lebih baik dalam merepresentasikan fitur visual kompleks, terutama pada kelas Brahma, Ganesha, dan Siwa dengan nilai recall sebesar 88,9%. Namun, kelas Saraswati masih menjadi kelas yang paling sulit dikenali dengan nilai recall 44,4% meskipun memiliki precision yang tinggi. Penelitian ini menyimpulkan bahwa ekstraksi fitur berbasis CNN lebih efektif dibandingkan BoVW dalam mendukung klasifikasi citra patung dewa dan dewi yang memiliki kompleksitas visual tinggi.

Kata Kunci: Visi Komputer, Objek Warisan Budaya, Pengenalan Objek, SVM

Abstract− Statues of gods and goddesses in the Balinese Hindu tradition possess high diversity in form and visual detail, posing challenges in the process of digital image recognition and classification. The differences in visual characteristics between statues drive the need for feature extraction methods capable of representing image information optimally. This study aims to compare the performance of Bag of Visual Words (BoVW) and Convolutional Neural Network (CNN) feature extraction in the classification of deity statue images using a Support Vector Machine (SVM). The dataset consists of 150 statue images divided into training and testing data. The research process includes image preprocessing, feature extraction using BoVW and CNN, and classification using SVM. Performance evaluation was conducted using accuracy, precision, and recall metrics. The test results indicate that the CNN–SVM combination achieved an accuracy of 84.44% with 38 correct predictions, which is higher than BoVW–SVM that obtained an accuracy of 77.78% with 35 correct predictions. CNN demonstrated superior ability in representing complex visual features, particularly for the Brahma, Ganesha, and Shiva classes with a recall value of 88.9%. However, the Saraswati class remained the most difficult to recognize with a recall value of 44.4%, despite having high precision. This study concludes that CNN-based feature extraction is more effective than BoVW in supporting the image classification of gods and goddesses statues that possess high visual complexity.

Keywords: Computer Vision, Cultural Heritage Objects, Object Recognition, SVM

1. PENDAHULUAN

Patung pada zaman dahulu dibuat untuk kepentingan keagamaan, terutama pada era Hindu dan Budha, sebagai bentuk penghormatan terhadap Dewa atau tokoh teladan. Patung-patung ini termasuk kategori religius dan biasanya ditempatkan di bangunan atau tempat suci. Di Bali, pelestarian budaya dilakukan dengan menggabungkan seni tradisional dan teknologi. Pemahaman terhadap berbagai jenis patung, khususnya di kalangan masyarakat Bali, cukup umum, namun berbeda bagi masyarakat luar Bali. Wisatawan yang mengunjungi pura di Bali akan menemui patung Dewa dan Dewi dengan variasi bentuk, meski terdapat kesamaan tertentu. Oleh karena itu, diperlukan klasifikasi untuk mengenali citra patung tersebut.

Machine Learning merupakan pendekatan yang efektif untuk klasifikasi citra, termasuk citra patung budaya, melalui beberapa tahapan utama. Tahap awal adalah preprocessing, di mana citra dibersihkan, distandarisasi, dan disiapkan untuk analisis lebih lanjut. Selanjutnya, ekstraksi fitur dilakukan untuk menyeleksi informasi visual penting dari citra, sehingga representasi objek menjadi lebih deskriptif dan informatif. Tahap akhir adalah klasifikasi, di mana digunakan untuk mengelompokkan citra ke dalam kategori yang sesuai. Sehingga

klasifikasi citra merupakan salah satu tantangan utama dalam bidang computer vision, terutama pada objek yang memiliki variasi bentuk dan detail visual kompleks. Ada dua pendekatan populer yang digunakan dalam ekstraksi variasi bentuk yakni BoVW dan CNN. BoVW adalah teknik tradisional yang mengekstraksi fitur lokal dari citra, membentuk kode visual, kemudian diklasifikasikan menggunakan algoritma seperti SVM [1]. Model BoVW banyak digunakan karena

This is an open access article under the CC–BY-SA license

Terakreditasi SINTA 5 SK :72/E/KPT/2024
Submitted: **19/10/2025**; Accepted: **29/10/2026**; Published: **16/06/2026**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 7 Nomor 3 Juni 2026, Page 337-345
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

kemampuannya menangkap ciri lokal melalui histogram visual words, khususnya pada dataset dengan pola yang relatif berulang [2].

Di sisi lain, CNN memperkenalkan pendekatan end-to-end di mana fitur diekstraksi langsung oleh lapisan konvolusi dan otomatis dioptimalkan selama pelatihan. CNN mampu mengenali pola visual hierarkis dari tingkat rendah (tepi, tekstur) ke tingkat tinggi (bentuk, objek), sehingga unggul pada dataset yang memiliki keragaman visual tinggi [3], [4]. Beberapa studi membandingkan performa CNN dengan metode klasik seperti SVM. Misalnya, penelitian pada klasifikasi penyakit daun singkong menunjukkan bahwa SVM dengan fitur manual memberikan performa stabil, sementara CNN mampu belajar pola visual otomatis [5].

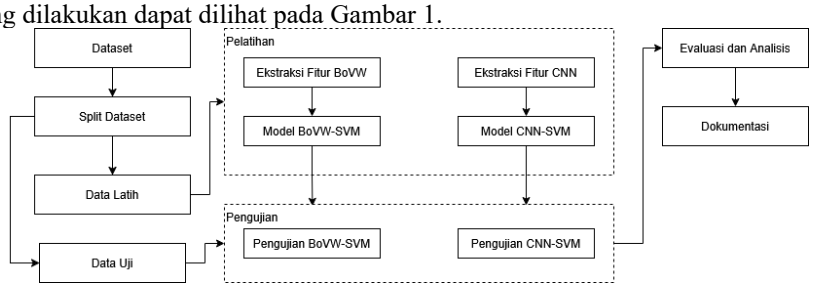
Penelitian pada klasifikasi alas kaki dan wajah manusia juga memperlihatkan CNN unggul dalam mengekstraksi fitur kompleks dibanding SVM [6], [7]. Walaupun demikian, BoVW tetap relevan sebagai baseline penting, terutama dengan modifikasi modern seperti Pooling-Based Bag of Visual Words, yang memperlihatkan efektivitas representasi fitur lokal pada citra kompleks [8]. Studi perbandingan antara CNN, KNN, dan SVM pada berbagai dataset citra menunjukkan bahwa evaluasi performa relatif antar metode klasik dan deep learning penting untuk memahami keunggulan dan keterbatasan masing-masing [9], [10]. Selain itu, CNN terbukti efektif mengekstraksi ciri visual dari citra grayscale dan berbagai tipe citra untuk klasifikasi objek, yang menjadi alasan kuat penggunaan CNN sebagai ekstraktor fitur utama sebelum klasifikasi oleh SVM.

Meskipun berbagai penelitian telah dilakukan dalam bidang klasifikasi citra menggunakan Support Vector Machine (SVM) maupun pendekatan deep learning, penerapannya pada objek budaya lokal, khususnya citra patung Dewa dan Dewi di Bali, masih relatif terbatas. Beberapa penelitian sebelumnya menerapkan SVM untuk klasifikasi citra dengan ekstraksi fitur berbasis CNN maupun deskriptor lokal seperti SURF, yang menunjukkan kinerja cukup baik pada objek umum [11], [12]. Penelitian lain juga memanfaatkan pendekatan deep learning murni menggunakan CNN atau representasi Bag of Visual Words (BoVW) untuk pengenalan objek visual [1], [13]. Namun, sebagian besar penelitian tersebut berfokus pada objek non-budaya atau tidak secara spesifik membandingkan efektivitas metode ekstraksi fitur klasik dan fitur mendalam pada citra patung yang memiliki kompleksitas visual tinggi serta kemiripan antar kelas. Selain itu, masih sedikit penelitian yang mengevaluasi secara langsung peran CNN sebagai ekstraktor fitur yang dikombinasikan dengan SVM, serta membandingkannya dengan metode BoVW sebagai pendekatan konvensional, khususnya dalam konteks klasifikasi citra patung Dewa dan Dewi di Bali.

Berdasarkan hal tersebut, penelitian ini bertujuan untuk membandingkan kinerja ekstraksi fitur Bag of Visual Words (BoVW) dan Convolutional Neural Network (CNN) dalam klasifikasi citra patung Dewa dan Dewi di Bali menggunakan Support Vector Machine (SVM). Penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan sistem klasifikasi citra berbasis computer vision untuk mendukung pelestarian budaya, edukasi wisatawan, serta dokumentasi digital kesenian Bali.

2. METODOLOGI PENELITIAN

Penelitian ini bertujuan untuk mengevaluasi dan membandingkan dua pendekatan ekstraksi fitur untuk klasifikasi citra patung Dewa dan Dewi di Bali: Bag of Visual Words (BoVW) dan Convolutional Neural Network (CNN), yang kemudian diklasifikasikan menggunakan Support Vector Machine (SVM). Diagram alur penelitian menunjukkan urutan kerja dari persiapan dataset, pembagian data, ekstraksi fitur, pelatihan model, pengujian, hingga evaluasi hasil. Adapun alur dari penelitan yang dilakukan dapat dilihat pada Gambar 1.



Gambar 1. Diagram alur metode penelitian klasifikasi citra patung Dewa dan Dewi

Tahap pertama adalah pengumpulan citra patung Dewa dan Dewi dari berbagai sumber dengan variasi bentuk, tekstur, dan latar belakang secara langsung dari beberapa lokasi di Bali, seperti pura dan artshop, kemudian dilakukan seleksi dan pelabelan kelas secara manual berdasarkan jenis patung. Semua citra kemudian melalui preprocessing, yang mencakup penyesuaian ukuran, normalisasi pencahayaan, dan penghapusan noise, sehingga data siap diproses secara komputasional. Proses preprocessing penting untuk menjaga kualitas input dan meminimalkan variasi yang tidak relevan [14]. Setelah kualitas citra terstandarisasi, dataset kemudian dibagi secara sistematis menjadi dua kategori utama, yakni data latih (training set) dan data uji (testing set). Pembagian ini dilakukan dengan rasio tertentu untuk memungkinkan model mempelajari pola-pola visual selama fase pelatihan, sementara data uji digunakan sebagai tolok ukur untuk

This is an open access article under the CC–BY-SA license

Terakreditasi SINTA 5 SK :72/E/KPT/2024
Submitted: **19/10/2025**; Accepted: **29/10/2026**; Published: **16/06/2026**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 7 Nomor 3 Juni 2026, Page 337-345
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

mengevaluasi performa model secara objektif. Dengan memisahkan kedua dataset ini, peneliti dapat memastikan bahwa model tidak hanya menghafal data (overfitting), tetapi benar-benar mampu mengenali objek pada data yang belum pernah ditemui sebelumnya [6].

Pendekatan pertama dalam ekstraksi fitur menggunakan metode Bag of Visual Words (BoVW). Metode ini bekerja dengan cara mengekstraksi deskriptor visual lokal dari setiap citra, yang kemudian dikelompokkan melalui proses clustering untuk membentuk sebuah visual vocabulary atau kamus visual dengan menggunakan persamaan (1). Setiap citra kemudian direpresentasikan kembali dalam bentuk histogram visual words yang menggambarkan frekuensi distribusi fitur lokal tersebut dihitung menggunakan persamaan (2). BoVW tetap menjadi pilihan yang relevan dalam klasifikasi citra karena efektivitasnya dalam menangkap tekstur dan pola lokal yang menjadi ciri khas dari sebuah objek. Pendekatan ini banyak digunakan dalam berbagai domain klasifikasi citra karena kemampuannya menangkap pola lokal dan tekstur citra [13].

$$\arg\min_{C} \sum_{i=1}^{N} \min_{j} ¿ \vee ¿ d_i - c_j ¿ \vee ¿^2 ¿ \quad (1)$$

$$f_{BoVW} = [h_1, h_2, \ldots, h_K] \quad (2)$$

Pendekatan kedua adalah menggunakan Convolutional Neural Network (CNN) sebagai metode ekstraksi fitur. CNN secara otomatis mempelajari pola dari data mentah melalui lapisan konvolusi pada persamaan (3) dan pooling. Berbeda dengan BoVW yang memerlukan deskriptor manual, CNN dapat mengekstraksi fitur dari level rendah hingga tinggi secara otomatis. Penelitian-penelitian terbaru menunjukkan bahwa CNN sangat efektif dalam menangkap ciri visual kompleks, termasuk tekstur dan bentuk, yang penting dalam klasifikasi citra visual [15], [16]. Dalam penelitian ini, CNN digunakan hanya sebagai ekstraktor fitur. Hasil keluaran fitur dari lapisan tertentu CNN akan di-flatten pada persamaan (4) dan dijadikan input ke SVM pada persamaan (4), sehingga disebut sebagai CNN-SVM hybrid approach. Pendekatan ini umum dipakai karena CNN unggul dalam mengekstraksi fitur tetapi SVM unggul dalam memisahkan kelas dengan margin optimal [4].

$$z_{i,j}^{(k)} = \sum_{m} \sum_{n} x_{i+m, j+n} w_{m,n}^{(k)} + b^{(k)} \quad (3)$$

$$f_{CNN} = [f_1, f_2, \ldots, f_n] \quad (4)$$

Setelah proses ekstraksi fitur dilakukan, vektor fitur yang dihasilkan dari metode Bag of Visual Words (BoVW) dan Convolutional Neural Network (CNN) digunakan sebagai masukan untuk membangun model klasifikasi berbasis Support Vector Machine (SVM). SVM adalah algoritma supervised learning yang bekerja dengan menemukan hyperplane optimal yang memisahkan kelas dengan margin terbesar, sehingga mampu menangani data berdimensi tinggi dan memberikan generalisasi yang baik [17]. Proses pelatihan SVM dilakukan dengan melakukan penyesuaian parameter, termasuk pemilihan kernel (linear, Radial Basis Function, dan polynomial), guna memperoleh performa klasifikasi yang optimal. Setelah tahap pelatihan selesai, kedua model diuji menggunakan data uji yang tidak terlibat dalam proses pembelajaran. Evaluasi performa dilakukan menggunakan metrik akurasi, presisi, dan recall untuk mengukur kemampuan model dalam mengklasifikasikan citra patung Dewa dan Dewi secara menyeluruh [18]. Hasil evaluasi ini kemudian dianalisis secara komparatif untuk menentukan metode ekstraksi fitur yang memberikan kinerja terbaik dalam konteks klasifikasi citra patung budaya Bali.

3. HASIL DAN PEMBAHASAN

3.1 Praprocessing

Dataset yang telah dikumpulkan melalui tahap akuisisi data terlebih dahulu melewati proses pemotongan atau cropping untuk memastikan bahwa hanya area informasi yang paling relevan yang akan diproses lebih lanjut oleh sistem. Mengingat gambar asli memiliki resolusi yang sangat tinggi, prosedur cropping dilakukan secara manual guna menjaga integritas fitur visual yang krusial dan menghindari hilangnya detail penting akibat pemotongan otomatis. Setelah proses tersebut, setiap citra sampel diseragamkan dimensinya melalui proses resize menjadi ukuran standar 200×250 piksel. Penyesuaian dimensi ini merupakan tahapan pra-pemrosesan yang sangat krusial karena ukuran citra yang konsisten akan sangat memengaruhi stabilitas dan akurasi pada saat fase pelatihan (training). Selain bertujuan untuk menyeragamkan format input data, reduksi dimensi ini juga berperan penting dalam menekan beban komputasi secara signifikan, sehingga

This is an open access article under the CC–BY-SA license

Terakreditasi SINTA 5 SK :72/E/KPT/2024
Submitted: **19/10/2025**; Accepted: **29/10/2026**; Published: **16/06/2026**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 7 Nomor 3 Juni 2026, Page 337-345
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

mempercepat waktu eksekusi program tanpa mengorbankan kualitas informasi yang dibutuhkan oleh algoritma. Dengan demikian, standarisasi ukuran citra menjadi 200×250 piksel yang digunakan pada model dalam mengenali pola-pola pada setiap sampel gambar secara lebih optimal dan sistematis. Adapun praprocessing dilakukan sebagai berikut:

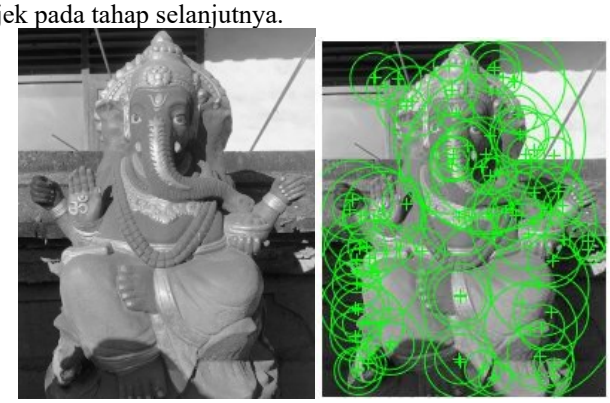


Gambar 2. (a) Gambar data sampel yang diambil (b) Hasil proses cropping

3.2 Ekstraksi Fitur

A. Fitur BoVW

Tahapan ekstraksi fitur BoVW diawali dengan melakukan konversi ruang warna citra dari format RGB (Red, Green, Blue) menjadi Grayscale atau skala abu-abu seperti pada Gambar 3(a). Secara teknis, transformasi ini dilakukan dengan mengalkulasi nilai rata-rata dari ketiga saluran warna pada setiap piksel individu. Sebagai ilustrasi, apabila sebuah piksel memiliki intensitas warna merah sebesar 100, hijau 60, dan biru 20, maka nilai tersebut dijumlahkan kemudian dibagi tiga, menghasilkan nilai intensitas tunggal sebesar 60. Proses standardisasi warna ini sangat krusial untuk menyederhanakan kompleksitas data visual sebelum memasuki tahap analisis fitur yang lebih mendalam. Setelah citra diubah menjadi skala abu-abu, langkah berikutnya adalah mengimplementasikan metode Speeded-Up Robust Features (SURF) untuk mengekstraksi titik-titik minat (interest points). Algoritma SURF bekerja dengan mendeteksi lokasi-lokasi spesifik pada gambar yang memiliki karakteristik unik dan distingtif yang mampu merepresentasikan ciri khas dari citra tersebut secara konsisten. Pada setiap titik poin yang telah berhasil diidentifikasi, sistem akan melakukan ekstraksi informasi lebih lanjut dengan menerapkan filter terhadap delapan nilai tetangga di sekitar titik pusat tersebut seperti pada Gambar 3(b). Prosedur ini bertujuan untuk membangun deskriptor fitur yang kuat dan tahan terhadap berbagai transformasi citra, sehingga data yang dihasilkan memiliki tingkat akurasi yang tinggi untuk digunakan dalam proses pengenalan atau klasifikasi objek pada tahap selanjutnya.



Gambar 3. (a) Perubahan citra ke grayscale (b) Hasil metode SURF

Berdasarkan gambar 2(b) didapatkan hasil dari interest point pada data sampel sebanyak 325500x105. Proses K-Means menggunakan jumlah fitur yang digunakan sebesar 325500x105 dan jumlah pusat yang digunakan sebesar 500. Sebagai contoh akan ditampilkan enam pusat cluster yang merupakan contoh representatif dari algoritma dalam mengidentifikasi pola unik dan kemiripan fitur di tengah kompleksitas data yang sangat luas. Melalui pendekatan ini, ribuan baris data disederhanakan menjadi kelompok-kelompok terstruktur, di mana setiap pusat cluster berfungsi merangkum karakteristik dominan yang membedakan satu segmen dengan segmen lainnya secara akurat. Dengan memetakan 500 pusat cluster tersebut, struktur data yang besar menjadi lebih mudah dikelola, memungkinkan identifikasi tren tersembunyi serta pengelompokan entitas secara sistematis guna mendukung proses analisis data yang lebih mendalam dan efisienPusat cluster ini bisa dikatakan sebagai words nantinya, atau jumlah kata pada pusat cluster. Adapun sebagian nilai dari pusat cluster dapat dilihat pada Tabel 1.

This is an open access article under the CC–BY-SA license

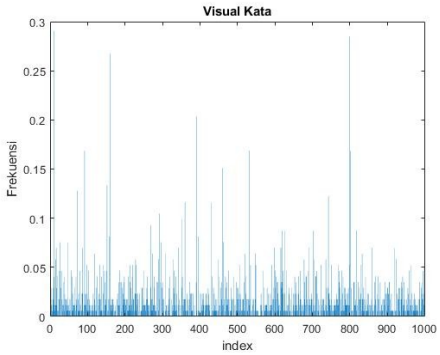
Submitted: **19/10/2025**; Accepted: **29/10/2026**; Published: **16/06/2026**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 7 Nomor 3 Juni 2026, Page 337-345
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

Tabel 1. Pusat Cluster

Pusat Cluster 1	Pusat Cluster 2	Pusat Cluster 3	Pusat Cluster 4	Pusat Cluster 5	Pusat Cluster 6
-0.0011608015	1.8028255e-05	0.0064092032	0.0056271423	0.018311942	0.0057439152
-0.0001971798	-0.0002055316	0.0030392150	0.0035760398	-0.0054929047	-0.0066170972
-0.0041838912	0.00027466496	0.010341647	0.0072904970	0.039534401	0.0094550103
0.00013874157	-0.0004827203	0.0049834820	0.0036530118	-0.021566467	-0.0099189207
0.00033527546	-0.0016375583	0.0041934792	0.0083805416	0.010952849	-0.0064663035
0.00072488212	0.0012550066	0.0042617186	0.0043228758	-0.016629113	0.0048060138
-0.0027816470	-0.0053108972	0.0078867273	0.010899192	0.055468384	-0.0097877281

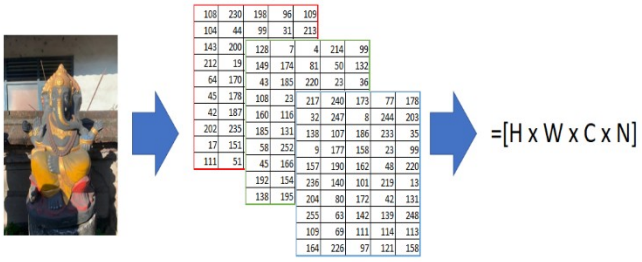
Proses kuantisasi fitur dimulai setelah penentuan 500 pusat cluster melalui algoritma K-Means, yang kemudian berfungsi sebagai kamus kata visual (*visual vocabulary*) untuk merepresentasikan setiap citra sampel dalam bentuk histogram. Setiap fitur lokal yang diekstraksi dari citra dipetakan ke pusat cluster terdekat sehingga membentuk distribusi frekuensi kata visual yang merepresentasikan karakteristik visual masing-masing sampel. Panjang histogram yang dihasilkan mengikuti jumlah pusat cluster yang telah ditentukan, di mana setiap nilai histogram mencerminkan intensitas kemunculan pola visual tertentu di dalam citra. Representasi ini memungkinkan citra yang memiliki kompleksitas visual tinggi untuk ditransformasikan ke dalam bentuk numerik yang terstruktur dan dapat diproses lebih lanjut oleh algoritma klasifikasi. Berdasarkan visualisasi histogram pada Gambar 4, terlihat bahwa setiap kelas patung memiliki pola distribusi kata visual yang berbeda, meskipun terdapat beberapa kemiripan pada kelas tertentu. Hal ini menunjukkan bahwa metode BoVW mampu menangkap ciri-ciri lokal seperti tekstur dan kontur permukaan patung, namun kurang efektif dalam merepresentasikan hubungan spasial antar fitur. Temuan ini sejalan dengan teori BoVW yang menekankan frekuensi fitur lokal tanpa mempertimbangkan struktur global objek, sehingga performanya cenderung menurun pada objek dengan detail visual kompleks dan variasi bentuk yang tinggi.



Gambar 4. Hasil Fitur BoVW

B. Fitur CNN

Tahapan awal dalam arsitektur Convolutional Neural Network (CNN) dimulai pada input layer, yang berfungsi untuk mentransformasi data citra mentah ke dalam struktur numerik berupa matriks atau array empat dimensi. Format tensor ini disusun berdasarkan parameter H×W×C×N, di mana setiap komponen memiliki peran spesifik: H (Height) merepresentasikan tinggi piksel, W (Width) adalah lebar piksel, C (Channel) menunjukkan kedalaman warna citra—seperti tiga saluran untuk RGB atau satu saluran untuk grayscale serta N yang menyatakan jumlah total sampel foto dalam satu batch yang diinputkan secara simultan. Proses pengubahan dimensi ini sangat fundamental karena memungkinkan model untuk memproses informasi spasial dan fitur visual secara paralel dan terstruktur. Dengan menyusun data ke dalam format array empat dimensi tersebut, mesin dapat mengenali pola piksel secara akurat melalui operasi konvolusi pada tahap berikutnya. Detail mengenai representasi visual dari mekanisme kerja pada input layer ini dapat diamati secara lebih komprehensif pada ilustrasi yang tersaji di Gambar 5.



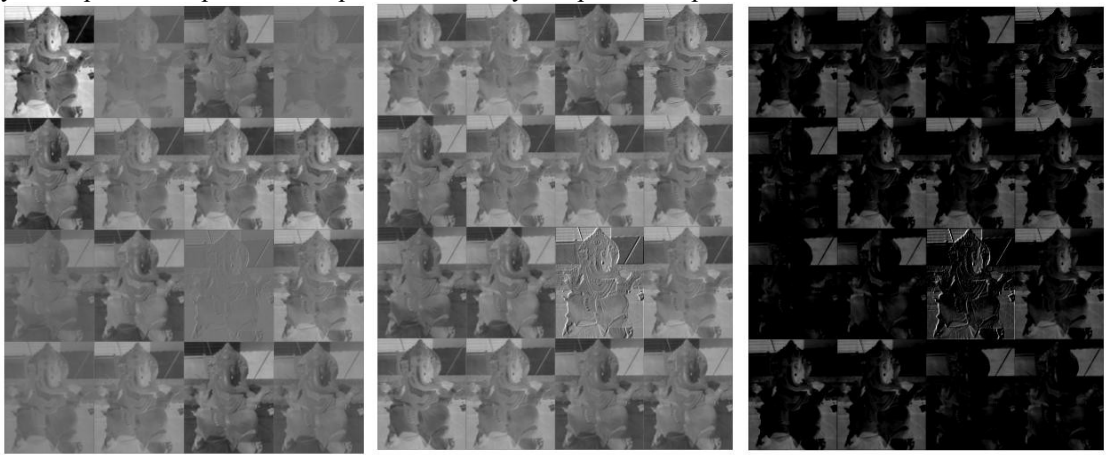
Gambar 1. Input Layer

This is an open access article under the CC–BY-SA license

Submitted: **19/10/2025**; Accepted: **29/10/2026**; Published: **16/06/2026**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 7 Nomor 3 Juni 2026, Page 337-345
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

Setelah melewati input layer, data citra memasuki serangkaian hidden layer yang diawali oleh convolutional layer. Pada tahapan ini, mesin melakukan ekstraksi fitur dengan mengaplikasikan filter atau kernel untuk menghasilkan feature map yang merepresentasikan pola unik dari citra input. Guna memastikan stabilitas dan mempercepat konvergensi selama proses pelatihan, diterapkan Batch Normalization yang berfungsi menormalisasi aktivasi pada setiap lapisan. Selanjutnya, digunakan fungsi aktivasi Rectified Linear Unit (ReLU) untuk memperkenalkan sifat non-linearitas ke dalam model dengan mengubah nilai input negatif menjadi nol. Kombinasi ketiga komponen ini memungkinkan CNN untuk mentransformasi data mentah menjadi representasi fitur yang lebih abstrak dan diskriminatif, sehingga mampu mengidentifikasi karakteristik visual yang kompleks dengan tingkat akurasi yang lebih optimal pada lapisan-lapisan berikutnya. Adapun beberapa hasil dari proses hidden layer dapat dilihat pada Gambar 6.



Gambar 2. (a) hasil convolution layer (b) hasil batch normalization layer (c) hasil aktivasi ReLU layer

Fase terakhir terletak pada Fully Connected Layer yang bertujuan untuk melakukan transformasi dimensi data agar informasi fitur yang telah diekstraksi dapat diklasifikasikan secara linear. Secara teknis, pada lapisan ini terjadi proses flattening, yaitu perubahan struktur data dari array empat dimensi (H×W×C×N) menjadi array dua dimensi yang lebih sederhana. Dalam transformasi ini, nilai dimensi tinggi (H), lebar (W), dan saluran (channel C) dilebur menjadi satu vektor tunggal (array pertama), sementara dimensi N yang mewakili jumlah sampel foto tetap dipertahankan sebagai array kedua. Tujuan utama dari restrukturisasi dimensi ini adalah untuk memetakan seluruh fitur spasial yang sebelumnya terpisah menjadi satu ruang representasi yang terpadu. Hasil akhir dari fully connected layer inilah yang nantinya diekstraksi sebagai fitur final citra yang siap digunakan untuk proses identifikasi maupun prediksi label kelas. Nilai-nilai numerik yang dihasilkan pada lapisan ini mencerminkan karakteristik unik dari setiap sampel dataset, yang menunjukkan kekuatan fitur tertentu terhadap klasifikasi target. Adapun hasil konkret dari kuantifikasi fitur melalui proses fully connected layer ini dapat diamati secara mendalam pada Tabel 2, yang merinci distribusi nilai fitur dari setiap citra sampel, mulai dari Fitur 1 hingga fitur-fitur selanjutnya untuk seluruh total N dataset.

Tabel 2. Layer Fully Connected

Dataset	Fitur 1	Fitur 2	Fitur 3	Fitur 4	Fitur 4
1	8.50947 5	-3.43851	-0.86505	-2.62909	-0.12264
2	10.1264 3	-4.11294	-0.08286	-0.74245	-2.6143
3	7.64484	-4.67892	0.408775	-3.54735	1.094311
4	6.73441 8	-2.27606	-1.84284	-3.22621	-1.6835
5	5.69232 9	-1.96604	-1.03814	-1.90193	-0.44523
6	6.60396	-2.58199	0.462438	-2.29742	-1.12102
7	4.78357 8	-1.68852	-0.24948	-1.07325	-0.19102
8	7.05565 3	-0.96853	-2.23329	0.052052	-2.55569
9	4.71675	-0.32507	-3.4064	-1.43849	-1.48862
10	5.23627 5	-0.70099	-1.29287	-2.29544	-0.78111

This is an open access article under the CC–BY-SA license

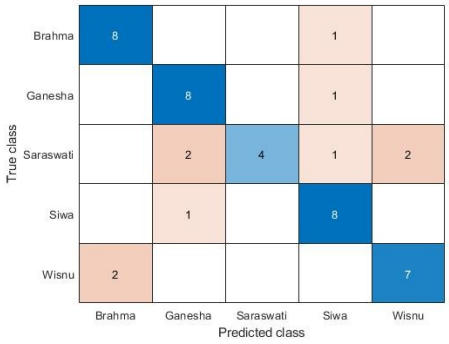
Terakreditasi SINTA 5 SK :72/E/KPT/2024
Submitted: **19/10/2025**; Accepted: **29/10/2026**; Published: **16/06/2026**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 7 Nomor 3 Juni 2026, Page 337-345
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

	4.97419				
11	2	-0.47263	-3.74596	-4.10932	0.933341
12	10.2748	-4.25573	-5.38122	-2.26722	2.069866
…	…	…	…	…	…
…	…	…	…	…	…
N	…	…	…	…	…

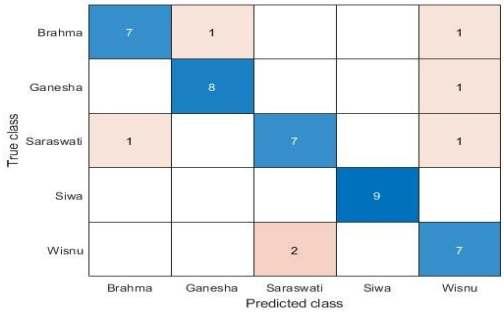
C. Klasifikasi SVM

Setelah mendapatkan representasi fitur melalui skema Bag of Visual Words (BoVW), data tersebut diinputkan ke dalam model klasifikasi untuk memvalidasi kemampuan sistem dalam mengenali pola patung dewa-dewi. Pengujian dilakukan secara sistematis dengan membandingkan label aktual pada dataset terhadap hasil prediksi yang dikeluarkan oleh model. Kinerja klasifikasi diukur menggunakan metrik Confusion Matrix untuk melihat distribusi sebaran prediksi pada setiap kelas objek. Berdasarkan hasil pengujian ini, diperoleh nilai akurasi, presisi, dan recall yang mencerminkan tingkat keberhasilan metode BoVW dalam mengklasifikasikan citra berdasarkan histogram kata visual yang telah terbentuk. Detail sebaran data dan performa klasifikasi tersebut dapat diamati secara lengkap pada Gambar 7.



Gambar 7. Confution Matriks metode SVM dengan Bag of Visual Words

Proses pengujian selanjutnya dilakukan pada fitur yang dihasilkan oleh algoritma Convolutional Neural Network (CNN) melalui lapisan fully connected. Karena fitur ini memiliki karakteristik yang lebih mendalam, model klasifikasi dievaluasi untuk melihat sejauh mana representasi spasial ini mampu meminimalkan tingkat kesalahan pengenalan. Pengujian menggunakan Confusion Matrix memungkinkan identifikasi kelas patung mana yang memiliki tingkat kemiripan tinggi sehingga berpotensi menyebabkan salah prediksi (misclassification). Hasil dari evaluasi ini memberikan gambaran objektif mengenai keandalan fitur CNN dalam mencapai nilai akurasi, presisi, dan recall yang optimal. Distribusi prediksi dan performa model secara menyeluruh untuk ekstraksi fitur CNN tersaji pada gambar di bawah ini:



Gambar 8. Confusion Matriks metode SVM dengan CNN

Berdasarkan hasil analisis dari confusion matrix yang telah diperoleh, maka dilakukan representasi nilai hasil pengujian recall dan precision secara komprehensif ke dalam bentuk tabel perbandingan untuk memberikan gambaran yang jelas mengenai kinerja sistem. Evaluasi ini merupakan tahapan krusial dalam memvalidasi keandalan model klasifikasi Support Vector Machine (SVM) yang telah diintegrasikan dengan dua metode ekstraksi fitur yang berbeda, yakni Bag of Visual Words (BoVW) dan Convolutional Neural Network (CNN). Penyajian data dalam format tabel ini bertujuan untuk mempermudah identifikasi perbedaan performa pada setiap kategori patung dewa-dewi secara mendetail, sehingga efektivitas dari masing-masing pendekatan ekstraksi fitur dapat dibandingkan secara objektif.

Melalui tabulasi data pada tabel 3, setiap parameter penilaian seperti precision yang menunjukkan ketepatan prediksi dan recall yang menggambarkan keberhasilan model dalam mengenali kembali data dari suatu kelas, dijabarkan secara sistematis. Struktur tabel yang teratur memungkinkan peneliti untuk mengamati tren keberhasilan atau kegagalan

This is an open access article under the CC–BY-SA license

Terakreditasi SINTA 5 SK :72/E/KPT/2024
Submitted: **19/10/2025**; Accepted: **29/10/2026**; Published: **16/06/2026**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 7 Nomor 3 Juni 2026, Page 337-345
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

klasifikasi pada kelas-kelas tertentu, seperti tingkat presisi yang sangat tinggi namun diikuti dengan nilai recall yang rendah pada objek spesifik. Dengan demikian, data yang disajikan tidak hanya sekadar angka statistik, melainkan menjadi dasar analisis dalam menentukan metode ekstraksi fitur mana yang paling optimal untuk mendukung pengenalan citra berbasis kearifan lokal Bali.

Tabel 3. Hasil Precision dan Recall

Kelas	BoVW-SVM		CNN - SVM	
	Precision	Recall	Precision	Recall
Brahma	80%	88,9%	80,0%	88,9%
Ganesha	72,7%	88,9%	72,7%	88,9%
Saraswati	100%	44,4%	100%	44,4%
Siwa	72,7%	88,9%	72,7%	88,9%
Wisnu	77,8%	77,8%	77,8%	77,8%

Berdasarkan data yang tersaji pada tabel perbandingan tersebut, terlihat bahwa kedua metode ekstraksi fitur, baik *Bag of Visual Words* (BoVW) maupun *Convolutional Neural Network* (CNN), menghasilkan nilai performa yang identik dalam proses klasifikasi menggunakan SVM untuk setiap kategori patung dewa-dewi. Fenomena unik ini menunjukkan bahwa pada dataset yang diuji, kedua metode tersebut mampu mencapai titik optimal yang sama dalam memetakan fitur ke dalam ruang klasifikasi, di mana kelas Saraswati mencatatkan nilai presisi sempurna sebesar 100% namun dengan nilai *recall* terendah di angka 44,4%. Sebaliknya, kelas Brahma, Ganesha, dan Siwa menunjukkan konsistensi yang kuat dengan nilai *recall* mencapai 88,9%, yang mengindikasikan bahwa model memiliki sensitivitas yang sangat tinggi dalam mengenali kembali objek-objek tersebut meskipun terdapat variasi pada citra input. Secara keseluruhan, hasil ini menggambarkan bahwa sistem memiliki ketepatan prediksi yang sangat baik pada kelas tertentu, namun masih menghadapi tantangan dalam mendeteksi seluruh sampel pada kelas Saraswati, terlepas dari jenis ekstraksi fitur yang diterapkan. Pengujian juga dilakukan berdasarkan tingkat akurasi dengan memberikan perbedaan performa yang signifikan antara kedua metode tersebut.

Tabel 4. Hasil Pengujian SVM terhadap Ektraksi Fitur

Metode	Total Sampel	Prediksi Benar	Akurasi (%)
BoVW - SVM	45	35	77,78%
CNN - SVM	45	38	84,44%

Penggunaan fitur yang diekstraksi melalui Convolutional Neural Network (CNN) terbukti memberikan hasil yang lebih unggul dengan akurasi sebesar 84,44%, atau berhasil memprediksi 38 citra dengan benar dari total 45 sampel. Nilai ini lebih tinggi dibandingkan metode BoVW yang hanya mencapai akurasi 77,78%. Peningkatan sebesar 6,66% ini menunjukkan bahwa fitur mendalam (deep features) yang dihasilkan oleh arsitektur CNN lebih mampu merepresentasikan karakteristik visual patung dewa-dewi di Bali secara lebih spesifik dan diskriminatif dibandingkan dengan metode histogram kata visual konvensional pada BoVW.

4. KESIMPULAN

Berdasarkan hasil penelitian dan pembahasan dapat disimpulkan bahwa penggunaan metode Convolutional Neural Network (CNN) sebagai teknik ekstraksi fitur memberikan performa yang lebih unggul dibandingkan dengan metode Bag of Visual Words (BoVW). Secara kuantitatif, integrasi CNN dengan klasifikasi Support Vector Machine (SVM) berhasil mencapai tingkat akurasi sebesar 84,44% dengan total 38 prediksi benar dari 45 sampel yang diuji. Capaian ini melampaui metode BoVW-SVM yang hanya mampu meraih akurasi sebesar 77,78% dengan 35 prediksi benar. Selisih peningkatan akurasi sebesar 6,66% ini menegaskan bahwa fitur mendalam yang dihasilkan CNN mampu merepresentasikan karakteristik visual patung secara lebih diskriminatif, khususnya pada objek dengan kompleksitas bentuk dan detail ornamen yang tinggi. Hasil ini menunjukkan bahwa pendekatan dengan ekstraksi fitur CNN lebih sesuai diterapkan pada klasifikasi objek budaya. Penelitian ini masih memiliki keterbatasan pada jumlah dan variasi data citra yang digunakan, sehingga kemampuan model dalam mengenali perbedaan visual yang sangat halus belum sepenuhnya optimal. Oleh karena itu, penelitian selanjutnya disarankan untuk memperluas dataset, menerapkan teknik augmentasi data, serta mengeksplorasi arsitektur CNN yang lebih dalam atau pendekatan klasifikasi end-to-end guna meningkatkan kinerja sistem secara keseluruhan.

REFERENCES

[1] Y. Huang, K. Inoue, and D. Li, "Pooling-Based Bag of Visual Words for Geographic Image Classification," *Journal of the Institute of Industrial Applications Engineers*, vol. 13, no. 1, pp. 1–10, Jan. 2025, doi: 10.12792/jiiae.13.1.

This is an open access article under the CC–BY-SA license

Terakreditasi SINTA 5 SK :72/E/KPT/2024
Submitted: **19/10/2025**; Accepted: **29/10/2026**; Published: **16/06/2026**

[2] Gina Annisa and Ninuk Wiliani, "Perbandingan Model CNN dan SVM untuk Klasifikasi Jenis Footwear pada Dataset Alas Kaki Berbasis Citra," *Teknomatika: Jurnal Informatika dan Komputer*, vol. 18, no. 1, pp. 28–36, Jun. 2025, doi: 10.30989/teknomatika.v18i1.1564.

[3] A. Gusviani, I. Handayani, and U. T. Yogyakarta, "PERBANDINGAN SUPPORT VECTOR MACHINE DAN CONVOLUTIONAL NEURAL NETWORK UNTUK KLASIFIKASI PENYAKIT DAUN SINGKONG," *Journal of Information Technology and Computer Science (INTECOMS)*, vol. 8, no. 6, 2025.

[4] Y. I. Royan, P. Pramono, and A. A. K. Asri, "Performance Comparison of Convolutional Neural Networks (CNN) and Support Vector Machine (SVM) Algorithms in Human Face Classification," *G-Tech: Jurnal Teknologi Terapan*, vol. 9, no. 3, pp. 1544–1553, Jul. 2025, doi: 10.70609/g-tech.v9i3.7384.

[5] Q. A'yuni and B. Hendrik, "Literature Review: Analisis Komparatif Algoritma CNN, KNN, dan SVM untuk Klasifikasi Penyakit Kelapa Sawit," *Journal of Education Research*, vol. 5, no. 4, pp. 6589–6596, 2024.

[6] N. Arifin and C. N. Insani, "Comparative Analysis of CNN, SVM, Decision Tree, Random Forest, and KNN for Maize Leaf Disease Detection Using Color and Texture Feature Extraction," *Jurnal Teknik Informatika (Jutif)*, vol. 6, no. 5, pp. 3572–3586, Oct. 2025, doi: 10.52436/1.jutif.2025.6.5.5128.

[7] Diah Putri Kartikasari, Fiqri Dian Priyatna Sinaga, Tiara Ayu Triarta Tambak, Zahra Humaira Kudadiri, and M. Khalil Gibran, "Ekstraksi Fitur Citra Grayscale dengan Convolutional Neural Networks," *Jurnal Teknik Informatika dan Teknologi Informasi*, vol. 5, no. 1, pp. 198–205, May 2025, doi: 10.55606/jutiti.v5i1.5175.

[8] Ò. Lorente, I. Riera, and A. Rana, "Image Classification with Classic and Deep Learning Techniques," May 2021, [Online]. Available: http://arxiv.org/abs/2105.04895

[9] S. Vijh, S. Kumar, and M. Saraswat, "New bag-of-feature for histopathology image classification using reinforced cat swarm algorithm and weighted Gaussian mixture modelling," *Complex and Intelligent Systems*, vol. 8, no. 6, pp. 5027–5046, Dec. 2022, doi: 10.1007/s40747-022-00726-5.

[10] A. Rehman, S. Saleem, U. G. Khan, S. Jabeen, and M. O. Shafiq, "Scene Recognition by Joint Learning of DNN from Bag of Visual Words and Convolutional DCT Features," *Applied Artificial Intelligence*, vol. 35, no. 9, pp. 623–641, 2021, doi: 10.1080/08839514.2021.1881296.

[11] Y. Amrozi, D. Yuliati, A. Susilo, N. Novianto, and R. Ramadhan, "Klasifikasi Jenis Buah Pisang Berdasarkan Citra Warna dengan Metode SVM," *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, vol. 11, no. 3, pp. 394–399, Dec. 2022, doi: 10.32736/sisfokom.v11i3.1502.

[12] R. Yohanes and M. Ezat Al Rivan, "Klasifikasi Jenis Kanker Kulit Menggunakan CNN - SVM," *Jurnal Algoritme*, vol. 2, no. 2, pp. 133–144, 2022.

[13] C. Valarmathi, S. John, and J. Thangaraj, "The Combination of Feature Extraction and Classification by Bag of Visual Words to Detect Breast Cancer for Improved Accuracy," 2024.

[14] Y. Wang, "Research on Image Classification and Recognition Technology Based on Machine Learning," *Applied Mathematics and Nonlinear Sciences*, vol. 9, no. 1, Jan. 2024, doi: 10.2478/amns-2024-1514.

[15] X. Zhao, L. Wang, Y. Zhang, X. Han, M. Deveci, and M. Parmar, "A review of convolutional neural networks in computer vision," *Artif. Intell. Rev.*, vol. 57, no. 4, Apr. 2024, doi: 10.1007/s10462-024-10721-6.

[16] Mozes Hasian Veltin Sinaga, Muhammad Albirra, and Muhammad Fajar Sidiq, "Klasifikasi Gambar Pemandangan dengan Kecerdasan Buatan Berbasis CNN," *Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi)*, 2024, Accessed: Jan. 30, 2026. [Online]. Available: https://doi.org/10.35870/jti k.v8i2.1424.

[17] A. I. Kushartanto, F. Fauziah, and R. T. Aldisa, "Comparison of CNN and SVM Methods on Web-based Skin Disease Classification Process," *Sinkron : : Jurnal dan Penelitian Teknik Informatika*, vol. 8, no. 2, pp. 778–788, Mar. 2024, doi: 10.33395/sinkron.v8i2.13349.

[18] D. Lanro, M. Tambunan, V. Siahaan, O. Dalimunthe, M. T. Ilham, and W. P. Manullang, "Implementasi Klasifikasi Gambar Daun Pepaya dan Daun Sirih Menggunakan Metode Convolution Neural Network (CNN)," *Jurnal Sains Dan Teknologi (JSIT)*, vol. 5, no. 3, pp. 369–377, 2025, doi: 10.47233/jsit.v5i2.3685.

This is an open access article under the CC–BY-SA license

Terakreditasi SINTA 5 SK :72/E/KPT/2024
Submitted: **19/10/2025**; Accepted: **29/10/2026**; Published: **16/06/2026**