

BAB I

PENDAHULUAN

1.1 Latar Belakang

Batik merupakan warisan budaya asli Indonesia yang telah diakui oleh UNESCO pada tanggal 2 Oktober 2009 sebagai warisan kemanusiaan yang wajib dilestarikan. Sebagai sebuah karya seni, batik adalah pola pada kain yang dibuat dengan menggunakan lilin atau malam sebagai bahan pelapis untuk menahan warna sehingga menghasilkan motif yang khas dan bernilai tinggi. Keberagaman motif batik di setiap daerah Indonesia mencerminkan kekayaan budaya, filosofi, serta identitas masyarakat setempat, menjadikan batik bukan sekadar kain bermotif, melainkan sebuah karya seni yang sarat akan makna dan nilai historis yang mendalam (Raspati dkk., 2023)

Perkembangan sistem klasifikasi pola berbasis citra digital mendorong pemanfaatan teknologi kecerdasan buatan dalam berbagai bidang, termasuk dalam upaya digitalisasi dan pelestarian budaya daerah seperti motif batik Dayak Kalimantan. Tantangan utama dalam pengembangan sistem klasifikasi motif batik Dayak Kalimantan adalah keterbatasan jumlah dan variasi *dataset* yang tersedia.

Dataset awal yang direncanakan dalam penelitian ini mencakup 600 citra batik Dayak Kalimantan yang berasal dari lima wilayah di Kalimantan, yaitu Kalimantan Barat, Kalimantan Selatan, Kalimantan Tengah, Kalimantan Timur, dan Kalimantan Utara, dengan masing-masing wilayah diwakili oleh 120 citra. Jumlah tersebut masih sangat terbatas dan belum merepresentasikan keseluruhan keragaman visual batik Dayak yang sesungguhnya tersebar di Kalimantan. Dokumentasi motif batik Dayak belum terstruktur secara menyeluruh, sementara setiap motif memiliki variasi pola, warna, dan orientasi yang cukup beragam.

Keterbatasan data pelatihan menyebabkan model klasifikasi sering kali dilakukan dengan jumlah data yang terbatas, sehingga akurasi menurun dan model cenderung mengalami *overfitting* ketika dihadapkan pada variasi citra yang berbeda dari *dataset*. Jumlah data pelatihan yang terbatas juga dapat menyebabkan model mengalami *underfitting*, yaitu kondisi ketika model belum mampu mempelajari pola penting dalam *dataset* secara optimal sehingga kinerja klasifikasi menjadi kurang maksimal. Penelitian terdahulu menunjukkan bahwa keterbatasan *dataset* berpengaruh langsung terhadap penurunan akurasi serta meningkatnya risiko *overfitting* pada model klasifikasi citra (Andzani dkk., 2025).

Permasalahan klasifikasi citra semakin kompleks pada motif batik Dayak Kalimantan yang memiliki karakter visual khas dan memiliki detail ornamen yang beragam. Motif yang terinspirasi dari unsur flora, fauna, serta simbol budaya masyarakat Dayak memiliki pola berulang dengan variasi bentuk dan komposisi yang tidak selalu seragam. Keberagaman pola tersebut memerlukan sistem yang mampu mengenali motif secara konsisten pada berbagai kondisi sudut pengambilan gambar, pencahayaan, skala, dan kualitas dokumentasi. Penelitian sebelumnya menunjukkan bahwa variasi kondisi visual seperti perubahan intensitas cahaya dan orientasi objek dapat mempengaruhi kinerja model klasifikasi citra (Rosanti dkk., 2024). Keterbatasan jumlah dan variasi *dataset* menunjukkan bahwa diperlukan strategi yang mampu memperkaya variasi data tanpa harus melakukan pengumpulan data dalam skala besar yang memerlukan waktu dan biaya tinggi.

Proses identifikasi motif batik Dayak secara manual memiliki berbagai keterbatasan karena banyaknya pola yang memiliki kemiripan visual sehingga

sulit dibedakan secara tepat. Perbedaan kecil pada warna, detail ornamen, dan orientasi pola memerlukan pengamatan yang cermat agar identifikasi dapat dilakukan secara akurat. Penelitian terdahulu menunjukkan bahwa identifikasi motif batik secara manual memerlukan keahlian khusus, memakan waktu, dan rentan terhadap perbedaan penilaian (Muria and Dedy, 2025). Metode yang banyak digunakan dalam pendekatan *Machine Learning* untuk klasifikasi citra adalah *Convolutional Neural Network* (CNN), yang merupakan pengembangan dari *Artificial Neural Network* (ANN) dirancang khusus untuk mengolah data berbentuk citra. *Convolutional Neural Network* (CNN) bekerja melalui lapisan *convolution* untuk mengambil fitur dari gambar, kemudian menggunakan lapisan *pooling* untuk menyederhanakan ukuran data tanpa menghilangkan informasi penting, serta lapisan *fully connected* untuk menentukan hasil klasifikasi. Keunggulan *Convolutional Neural Network* (CNN) adalah kemampuannya mengenali pola secara otomatis, mulai dari fitur sederhana seperti tepi dan tekstur hingga pola yang lebih kompleks, tanpa perlu perancangan fitur secara manual (Samidin and Fadjeri, 2024).

Kinerja model *Convolutional Neural Network* (CNN) dalam klasifikasi citra dipengaruhi oleh ketersediaan *dataset* yang memadai dan bervariasi. Hasil penelitian sebelumnya menunjukkan bahwa keterbatasan jumlah dan variasi *dataset* tidak hanya berdampak pada kinerja sistem klasifikasi berbasis citra digital, tetapi juga mempengaruhi konsistensi proses identifikasi manual pada motif batik Dayak Kalimantan. Penerapan teknik augmentasi citra telah terbukti mampu meningkatkan perluasan *dataset* dan kemampuan generalisasi model, namun penelitian yang secara khusus mengkaji pengaruhnya pada klasifikasi

motif batik Dayak Kalimantan berdasarkan wilayah asal masih belum dilakukan. Penelitian lain menunjukkan bahwa perubahan citra seperti rotasi, *flipping*, translasi, dan perubahan skala mampu meningkatkan keragaman *dataset* sehingga model menjadi lebih stabil dan adaptif terhadap variasi kondisi visual (Yoo and Kang, 2023).

Penelitian yang secara khusus membahas penerapan augmentasi citra pada klasifikasi citra batik Dayak Kalimantan berdasarkan wilayah asal masih terbatas dan belum dikaji secara mendalam, padahal citra batik Dayak memiliki variasi bentuk, warna, dan detail ornamen yang beragam sehingga membutuhkan *dataset* yang lebih bervariasi agar model mampu melakukan klasifikasi secara lebih optimal. Masih terbatasnya penelitian yang membahas penerapan teknik augmentasi pada klasifikasi citra batik Dayak Kalimantan menunjukkan adanya celah penelitian pada aspek pengaruh penerapan teknik augmentasi terhadap perluasan *dataset* dan peningkatan kinerja klasifikasi citra batik Dayak Kalimantan berbasis wilayah asal. Augmentasi citra memanfaatkan bahasa pemrograman Python untuk menghasilkan variasi data secara otomatis sehingga dapat mendukung peningkatan kinerja model klasifikasi.

Penelitian ini difokuskan pada satu permasalahan utama yaitu keterbatasan jumlah dan variasi *dataset* motif batik Dayak Kalimantan dalam pengembangan sistem klasifikasi citra digital, yang berpotensi menyebabkan model mengalami *overfitting*, menurunkan kemampuan generalisasi, serta menghasilkan tingkat akurasi klasifikasi yang tidak optimal ketika dihadapkan pada data baru. Jumlah dan variasi *dataset* yang belum memadai menyebabkan variasi pola, warna, dan orientasi motif yang digunakan dalam pengembangan sistem menjadi kurang

beragam. Akibatnya, sistem menunjukkan kinerja tidak maksimal ketika mengklasifikasikan citra dengan kondisi sudut pengambilan gambar, pencahayaan, dan skala yang berbeda. Penelitian ini bertujuan untuk menerapkan teknik augmentasi gambar guna memperluas jumlah dan variasi *dataset* batik Dayak Kalimantan tanpa mengubah karakteristik pola utama motif batik Dayak, serta membandingkan hasil klasifikasi citra pada Orange Data Mining sebelum dan sesudah augmentasi data berdasarkan nilai *AUC*, *CA*, *F1-score*, *Precision*, dan *Recall*.

1.2 Rumusan Masalah

Berdasarkan latar belakang penelitian, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Apakah penerapan teknik augmentasi gambar dapat memperluas jumlah dan variasi *dataset* batik Dayak Kalimantan tanpa mengubah karakteristik pola utama motif batik Dayak?
2. Bagaimana perbandingan hasil klasifikasi citra batik Dayak Kalimantan pada Orange Data Mining sebelum dan sesudah augmentasi data berdasarkan nilai *AUC*, *CA*, *F1-score*, *Precision*, dan *Recall*?

1.3 Tujuan Penelitian

Tujuan penelitian ini adalah:

1. Menerapkan teknik augmentasi gambar untuk memperluas jumlah dan variasi *dataset* batik Dayak Kalimantan tanpa mengubah karakteristik pola utama motif batik Dayak.

2. Membandingkan hasil klasifikasi citra batik Dayak Kalimantan pada Orange Data Mining sebelum dan sesudah augmentasi data berdasarkan nilai AUC, CA, *F1-score*, *Precision*, dan *Recall*.

1.4 Ruang Lingkup

Penelitian ini berfokus pada penerapan teknik augmentasi gambar untuk memperluas *dataset* batik Dayak Kalimantan dalam sistem klasifikasi pola berbasis citra digital. Ruang lingkup materi mencakup pengolahan citra digital, teknik augmentasi gambar, serta pemanfaatan teknologi *Machine Learning* berbasis *Convolutional Neural Network* (CNN) dalam pengenalan pola batik.

Ruang lingkup secara rinci untuk penelitian ini adalah sebagai berikut:

1. Jenis data berupa citra digital batik Dayak Kalimantan yang dikelompokkan menjadi 5 kelas berdasarkan wilayah asal, yaitu Kalimantan Barat, Kalimantan Selatan, Kalimantan Tengah, Kalimantan Timur, dan Kalimantan Utara, dengan masing-masing kelas berisi 120 citra yang memiliki variasi warna, bentuk, dan pola ornamen khas.
2. Sumber data berupa citra batik Dayak Kalimantan yang diperoleh dari *platform kaggle* dan *dataset* penelitian (Dzulqarnain dkk., 2025) yang bersifat legal dan bebas copyright, tanpa melakukan pengumpulan data lapangan secara langsung.
3. Teknik augmentasi yang diterapkan meliputi rotasi, *flip horizontal*, *flip vertical*, *zoom*, dan *color shift* menggunakan bahasa pemrograman Python dengan memanfaatkan *library keras ImageDataGenerator*, tanpa mengubah pola utama motif batik Dayak Kalimantan.

4. Metode klasifikasi dilakukan dengan memanfaatkan ekstraksi fitur citra berbasis *Convolutional Neural Network* (CNN) melalui fitur *Image Embedding* pada Orange Data Mining. Alur kerja meliputi *import dataset* citra yang telah diperluas menggunakan bahasa pemrograman Python, proses *Image Embedding* untuk ekstraksi fitur citra, pembentukan model klasifikasi menggunakan algoritma *Machine Learning* yaitu *Neural Network*, pembagian data *training* dan *testing*, serta evaluasi kinerja model melalui *Test and Score* berdasarkan nilai AUC, CA, *F1-score*, *Precision*, dan *Recall*.

5. Model CNN pralatih yang digunakan dalam proses *Image Embedding* adalah *SqueezeNet (local)*, yang dipilih karena bersifat ringan dan dapat dijalankan

secara *offline* tanpa memerlukan koneksi internet maupun spesifikasi perangkat keras yang tinggi. Metode pembagian data yang digunakan adalah *Stratified 5-fold Cross Validation* dengan proporsi 80% data latih dan 20% data uji pada setiap *fold*, sehingga seluruh data mendapat kesempatan digunakan sebagai data pengujian secara bergantian dan hasil evaluasi menjadi lebih representatif.