



Implementasi Sistem Klasifikasi Batik Menggunakan MobileNet dengan Integrasi Chatbot Retrieval Augmented Generation

Rizki Purnomo Pratama*, Donny Avianto

Fakultas Sains dan Teknologi, Informatika, Universitas Teknologi Yogyakarta, Sleman, Indonesia

Email: ^{1,*}rizkipurnomo914@gmail.com, ²donny@uty.ac.id

Email Penulis Korespondensi: rizkipurnomo914@gmail.com

Abstrak—Sebagai warisan budaya Indonesia yang mendapat pengakuan dari UNESCO, batik memiliki berbagai motif yang sarat akan nilai filosofis. Namun, pengetahuan publik tentang motif batik masih terbatas. Studi ini mengembangkan sistem klasifikasi batik berbasis *mobile* yang mengintegrasikan MobileNetV2 dengan chatbot *Retrieval-Augmented Generation (RAG)* untuk pengalaman belajar interaktif. Sistem memungkinkan pengguna mengidentifikasi pola batik melalui pengenalan gambar dan memperoleh informasi detail via AI percakapan. Penelitian ini mengadopsi MobileNetV2 atas pertimbangan efisiensinya pada perangkat mobile. Model ini mampu mencapai keseimbangan yang optimal antara akurasi dan kinerja komputasi. Model dilatih pada dataset seimbang berisi 5000 gambar mencakup lima motif (Parang, Truntum, Kawung, Mega mendung, dan Merak), mencapai akurasi *training* 98.97% dan akurasi *testing* 96.8%. Chatbot RAG yang diorkestrasi menggunakan LangChain dan Qdrant meningkatkan interaksi dengan mengambil informasi relevan dari basis pengetahuan, memastikan *respons* faktual kontekstual tentang sejarah, filosofi, dan makna budaya batik. React Native dipilih sebagai framework pengembangan aplikasi untuk mendukung pengoperasian lintas platform. Implementasi ini berkontribusi pada pelestarian warisan budaya dengan membuat pengetahuan batik lebih aksesibel melalui teknologi modern, menggabungkan *computer vision* dan *natural language processing* dalam platform terpadu.

Kata Kunci: Batik; MobileNet; Retrieval Augmented Generation; Aplikasi Mobile; Deep Learning

Abstract—As an Indonesian cultural heritage recognized by UNESCO, batik features various motifs laden with philosophical values, yet public knowledge about batik patterns and their significance remains limited. This study presents a mobile-based batik classification system integrating MobileNetV2 architecture with a Retrieval-Augmented Generation (RAG) chatbot to provide interactive learning experiences, enabling users to identify batik patterns through image recognition while obtaining detailed information via conversational AI. This study adopts MobileNetV2 considering its efficiency on mobile devices. This model achieves an optimal balance between accuracy and computational performance. Model was trained on a balanced dataset of 5,000 images covering five pattern classes (Parang, Truntum, Kawung, Mega mendung, and Merak), achieving training accuracy of 98.97% and testing accuracy of 96.8%. The RAG-based chatbot, orchestrated using LangChain and Qdrant, enhances user interaction by retrieving relevant information from a curated knowledge base, ensuring contextual factual responses about batik's history, philosophy, and cultural significance. React Native was adopted as the development framework to ensure cross-platform operability. This implementation contributes to cultural heritage preservation by making batik knowledge more accessible through modern technology, combining computer vision and natural language processing in a unified platform.

Keywords: Batik; MobileNet; Retrieval Augmented Generation; Mobile Application; Deep Learning

1. PENDAHULUAN

Sebagai warisan budaya Indonesia, batik menyimpan kekayaan nilai dari segi estetis, historis, maupun filosofis yang sangat berharga[1]. Pengakuan global terhadap karya seni ini terealisasi melalui pencatatannya oleh UNESCO sebagai Warisan Budaya Takbenda Dunia[1]. Pengakuan tersebut telah diberikan secara resmi sejak tanggal 2 Oktober 2009[2]. Sebagai ekspresi budaya yang hidup, batik telah dipertahankan selama generasi di berbagai wilayah Indonesia, dengan akar tradisi terkuat berada di Pulau Jawa[2]. Batik menjadi kebanggaan nasional berkat kekayaan ekspresinya yang unik, terlihat dari beragam corak, pola, serta pilihan warna yang mencerminkan identitas budaya Indonesia[2]. Setiap motif batik mengandung simbolisme yang sarat makna dan menjadi perwujudan jati diri budaya masyarakatnya[1]. Simbolisme yang terkandung dalam batik ini berakar kuat pada sistem kepercayaan masyarakat agraris masa lampau, di mana nilai-nilai animisme dan dinamisme sangat mendominasi pandangan hidup mereka[1]. Motif batik juga dipandang sebagai warisan generasional yang terintegrasi erat dengan tradisi serta nilai-nilai kearifan lokal[1]. Warisan budaya batik ini diibaratkan sebagai lukisan dua dimensi dengan kain sebagai media utama[1]. Motif batik terdiri dari banyak ragamnya, salah satunya, batik Yogyakarta, memperlihatkan pola tekstur yang penuh detail[1]. Motif Batik Kawung menjadi busana eksklusif bagi sultan dan keluarga kerajaan di bawah pemerintahan Sri Sultan Hamengkubuwono VII. Namun setelahnya, motif Batik Kawung diperbolehkan untuk digunakan masyarakat secara umum sejak Sultan Hamengkubuwono IX memegang takhta[3]. Motif Batik Parang memiliki pola berupa rangkaian huruf “S” yang tersusun saling berkaitan. Pola ini membentuk pola diagonal miring menyerupai lereng gunung. Motif ini mencerminkan dinamika hidup dan kekuatan yang berkelanjutan[1].

Di tengah pesatnya perkembangan teknologi digital di era globalisasi, pelestarian batik menghadapi tantangan serius, perkembangan teknologi yang semakin marak menjadi salah satu masalah yang bisa mengikis perhatian dan kepedulian generasi muda terhadap kearifan lokal[4]. Minat anak muda masa kini terhadap batik cenderung rendah karena mereka lebih menyerap budaya timur dan barat[5]. Hal ini dilatarbelakangi oleh kurangnya pengetahuan generasi muda terhadap kearifan lokal di daerahnya bahkan, anak muda saat ini merasa malu dan tidak mau memakai batik[6]. Masalah yang muncul adalah kesulitan dalam mengidentifikasi motif batik, pengenalan motif batik secara manual memerlukan waktu[7]. Hal ini mengakibatkan kurangnya minat dalam memahami dan mengapresiasi keanekaragaman batik. Solusi



yang diusulkan berupa pengembangan aplikasi *mobile* berbasis model *deep learning* efisien untuk mengidentifikasi pola batik. Untuk menyampaikan informasi budaya secara interaktif, sistem ini akan diintegrasikan ke dalam *chatbot* yang menggunakan pendekatan *Retrieval-Augmented Generation (RAG)*, sehingga responsnya bersifat lebih faktual.

Klasifikasi motif batik memanfaatkan sebuah model *Convolutional Neural Network (CNN)*, sebagai salah satu pendekatan dalam ranah *deep learning* sebagai pendekatan utamanya. Dalam CNN, alur pemrosesan umumnya dimulai dengan lapisan konvolusi, dilanjutkan oleh lapisan *pooling*, dan diakhiri dengan lapisan *fully connected* untuk menghasilkan prediksi[2]. Untuk implementasi pada perangkat *mobile*, efisiensi model adalah pertimbangan utama karena keterbatasan sumber daya komputasi. Efisiensi dan akurasi MobileNetV2 diperoleh berkat penggunaan *Depthwise Separable Convolution* sebagai pengganti konvolusi konvensional[1]. MobileNetV2 menunjukkan keseimbangan yang lebih baik antara akurasi dan efisiensi dibandingkan ResNet50, sebagaimana dilaporkan dalam salah satu penelitian yang membandingkan arsitektur jaringan saraf konvolusional ringan untuk aplikasi pada perangkat terbatas sumber daya[1].

Adapun, untuk memberikan lapisan informasi kontekstual yang akurat dan interaktif, diperlukan integrasi dengan sistem percakapan yang Pendekatan *Retrieval-Augmented Generation (RAG)* merupakan solusi yang memadukan mekanisme pencarian informasi (*retriever*) dengan kekuatan generatif dari *Large Language Model (LLM)*[8]. Sistem RAG bekerja dengan mengambil dokumen atau potongan informasi yang paling relevan dari basis pengetahuan eksternal sebelum menghasilkan *respon*. RAG juga mampu mengurangi masalah *hallucination* (halusinasi) atau jawaban yang tidak akurat dan pengetahuan yang usang pada model generatif konvensional dengan mendasarkan *output* pada informasi tambahan[9].

Pengembangan sistem klasifikasi motif batik telah menjadi fokus dalam berbagai penelitian, dimulai dari implementasi arsitektur CNN kustom. Sebagai contoh, sebuah *prototipe* web dikembangkan yang mampu mencapai akurasi pelatihan 93,18% [10]. Penelitian lain menerapkan CNN kustom pada dataset yang lebih besar (3.880 gambar, 14 kelas) dan berhasil memperoleh akurasi pengujian 91,24%[11]. Upaya lain merancang arsitektur *Multi-Layer CNN* dengan beragam ukuran filter (3x3, 5x5, 7x7) untuk mengekstrak fitur unik, yang berhasil mencapai akurasi *overall* 90,88% [12].

Penggunaan *transfer learning* menunjukkan peningkatan kinerja yang signifikan, terutama dengan arsitektur MobileNetV2 yang efisien untuk aplikasi *mobile*. Beberapa penelitian melaporkan pencapaian akurasi 100% menggunakan MobileNet [13] dan MobileNetV2 [14]. Penelitian pada batik Lampung menggunakan MobileNet juga menunjukkan hasil yang kompetitif dengan akurasi mencapai 98,00%[15]. Pendekatan gabungan yang unik, dengan menggabungkan ekstraksi fitur GLCM dengan MobileNetV2, juga menghasilkan akurasi sangat tinggi sebesar 99%[1], sementara penelitian lain yang juga menggunakan MobileNetV2 mencatat akurasi 92% [16]. Arsitektur *transfer learning* lainnya juga terbukti ampuh; VGG16 ditemukan mencapai akurasi 100% pada data validasi [17]. Pada dataset yang lebih menantang dengan 20 kelas motif, *Vision Transformer (ViT)* terbukti sangat efektif dengan akurasi 96% [18], mengungguli EfficientNetB0 yang mencapai 82,5% pada jumlah kelas yang sama [2]. Secara paralel, penelitian di bidang *Natural Language Processing (NLP)* telah berfokus pada pengembangan *chatbot* yang faktual dan kontekstual. Penelitian awal yang mengoptimalkan *chatbot* informasi akademik menggunakan teknik NLP dasar memperoleh skor *User Acceptance Test (UAT)* sebesar 88,6% [19], sementara penelitian lain yang menggunakan arsitektur BERT untuk informasi *handphone* berhasil mencapai akurasi *Intent Recognition* 87% [20].

Pendekatan *state-of-the-art* saat ini adalah RAG, yang dirancang untuk mengatasi halusinasi LLM dengan menghubungkannya ke basis pengetahuan eksternal. Evaluasi sistem RAG sering menggunakan metrik seperti *Faithfulness* (ketaatan pada fakta) yang menilai konsistensi faktual dari jawaban yang dihasilkan relatif terhadap konteks yang diberikan, serta *Answer Relevancy* yang menilai seberapa relevan jawaban tersebut terhadap *prompt* pengguna[9]. RAG telah terbukti berhasil di berbagai domain di Indonesia. Untuk layanan akademik, sebuah penelitian mencapai akurasi 100% pada set pertanyaan uji menggunakan model LLaMA dan Mistral [21], sementara penelitian lain yang mengevaluasi proposal tugas akhir mencapai 80% *Answer Correctness*[8]. Studi lain yang menerapkan RAG pada *chatbot* layanan akademik menggunakan Gemini 2.5 Pro mencatat skor *Faithfulness* (ketaatan pada fakta) sebesar 0.9100 [22]. Sebuah studi perbandingan juga mengkonfirmasi bahwa RAG (*RAGAs Score* 0.72) lebih unggul daripada *finetuning* murni untuk tugas informasi universitas [23]. Di sektor komersial dan publik, sistem RAG *on-premise* (lokal) untuk menangani keluhan pelanggan PLN dilaporkan 92% jawabannya dinilai membantu oleh evaluator manusia [24]. Selain itu, implementasi RAG untuk *chatbot e-commerce* berbasis Gemini AI menghasilkan metrik kinerja sangat tinggi, termasuk *Faithfulness* 91,67% dan *Answer Relevancy* 94,21% [25].

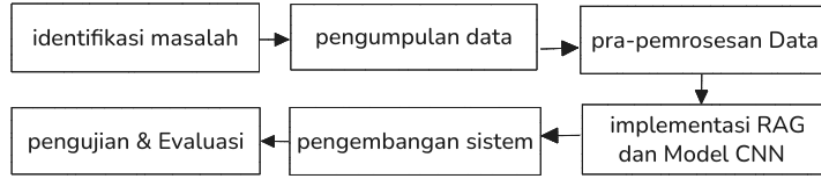
Kesenjangan utama yang teridentifikasi adalah bahwa belum ada penelitian yang mengintegrasikan klasifikasi citra batik menggunakan MobileNetV2 dengan sistem tanya jawab kontekstual berbasis *Retrieval-Augmented Generation (RAG)* dalam satu platform *mobile*. Penelitian ini mengisi celah tersebut dengan mengembangkan arsitektur sistem yang menghubungkan identifikasi visual motif batik dengan pemahaman konteks budaya dan pengetahuan faktual melalui *chatbot* RAG. Berdasarkan paparan tersebut, fokus utama dari penelitian ini adalah pada pengembangan dan mengimplementasikan sistem klasifikasi motif batik yang menggunakan arsitektur MobileNetV2 dan mengintegrasikannya dengan *chatbot* berbasis RAG untuk penyediaan informasi budaya yang akurat. Harapan yang ingin dicapai adalah menghasilkan prototipe aplikasi *mobile* yang dapat menjadi alat digital efektif untuk pelestarian dan promosi budaya batik. Model ini diharapkan mampu meningkatkan literasi budaya dan minat generasi muda terhadap batik dengan menyediakan akses informasi yang interaktif dan modern. Selanjutnya, model ini direncanakan dapat diimplementasikan ke dalam aplikasi *mobile* dengan arsitektur *client-server*.



2. METODOLOGI PENELITIAN

2.1 Alur Penelitian

Penelitian yang dikembangkan dalam artikel ini dilaksanakan melalui lima tahapan sistematis. Gambar 1 menyajikan diagram alur yang merinci langkah-langkah yang diambil.



Gambar 1. Alur Penelitian

Langkah awal dalam penelitian ini adalah mengidentifikasi permasalahan sebagai dasar pengembangan studi. Tahap kedua melakukan pengumpulan data dan literatur review untuk memahami kondisi dan teori yang relevan. Tahap ketiga melatih model CNN MobileNet V2 sebagai model klasifikasi utama. Tahap keempat mengimplementasikan sistem RAG dengan *Vector database* dan *LLM* untuk meningkatkan kemampuan sistem dalam memberikan informasi kontekstual. Tahap terakhir adalah pengujian dan evaluasi untuk mengukur performa sistem yang telah dikembangkan.

2.2 Pengumpulan Data

2.1.1 Model Klasifikasi

Dataset lima jenis motif batik tradisional Indonesia dijadikan kelas dalam dataset penelitian ini, seperti yang dapat dilihat pada Gambar 2.



Gambar 2. Dataset Penelitian

Dataset mencakup lima kategori motif batik yaitu Truntum, Mega mendung, Parang, Kawung, dan Merak. Setiap kelas motif memiliki karakteristik visual yang berbeda dari aspek pola geometris dan komposisi warna yang memungkinkan model untuk melakukan diferensiasi antar kelas. Dataset ini digunakan untuk melatih dan menguji model MobileNet V2 dalam melakukan klasifikasi motif batik berdasarkan fitur visual yang diekstrak dari citra input. Pembagian data dari semua dataset ditampilkan pada Tabel 1.

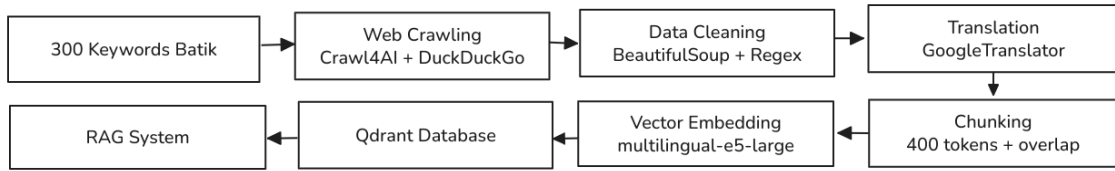
Tabel 1. Pembagian Dataset

Kelas Motif Batik	Data Training	Data Validation	Data Testing	Total Data per Kelas
Kawung	600	200	200 (42 primer)	1,000
Truntum	600	200	200 (42 primer)	1,000
Mega mendung	600	200	200 (42 primer)	1,000
Parang	600	200	200 (42 primer)	1,000
Merak	600	200	200	1,000

Studi ini mengandalkan data yang bersumber dari tiga sumber dengan total 5.000 citra motif batik. Dataset primer merupakan hasil pengumpulan manual yang terdiri dari 168 citra dan seluruhnya digunakan sebagai bagian dari data *testing*. Dataset sekunder pertama diperoleh dari Kaggle Dataset "Batik Keraton" yang berkontribusi sebanyak 1.799 citra dan digunakan untuk data *training*, *validation*, dan *testing*. Dataset sekunder kedua berasal dari Kaggle Dataset "Motif Batik Indonesia Fix" dengan 3.033 citra yang diseleksi secara parsial berdasarkan label kategori yang relevan, kemudian didistribusikan untuk keperluan *training*, *validation*, dan *testing*. Kombinasi ketiga sumber dataset ini bertujuan untuk meningkatkan variabilitas data dan generalisasi model dalam mengenali berbagai variasi visual dari setiap kelas motif batik yang diteliti.

2.1.2 Crawling dan preprocessing data untuk RAG

Rincian proses tahapan untuk mengumpulkan dan mempersiapkan data bagi sistem RAG divisualisasikan dalam Gambar 3.

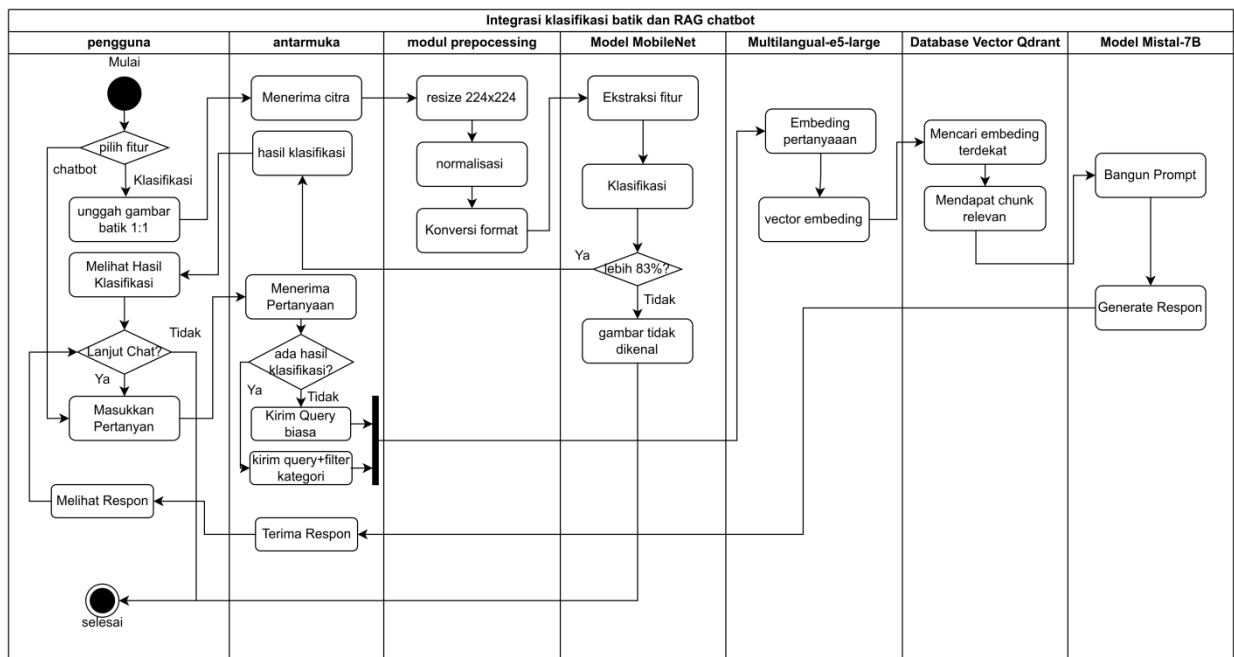


Gambar 3. Crawling dan preprocessing

Tahap pertama adalah pengumpulan data menggunakan metode *web crawling* melalui *library* Crawl4AI dan DuckDuckGo berdasarkan 300 kata kunci terkait batik seperti web iwarebatik. Tahap kedua merupakan *data cleaning* menggunakan BeautifulSoup dan Regex untuk menghilangkan elemen HTML dan karakter tidak relevan. Tahap ketiga adalah translasi menggunakan GoogleTranslator untuk menyeragamkan bahasa data. Tahap keempat adalah *chunking* dengan parameter 400 *tokens* per *chunk* dan *overlap* untuk membagi dokumen menjadi segmen optimal menghasilkan 2748 *chunks*. Tahap kelima adalah *vector embedding* menggunakan model *multilingual-e5-large* untuk mengonversi teks menjadi *vektor*. Tahap akhir adalah penyimpanan vektor ke Qdrant Database yang menghasilkan total 2.748 *chunks* untuk mendukung proses *retrieval* dalam sistem RAG.

2.2 Perancangan Sistem

Arsitektur sistem yang dikembangkan terintegrasi seperti yang ditunjukkan melalui *activity* diagram pada Gambar 4.

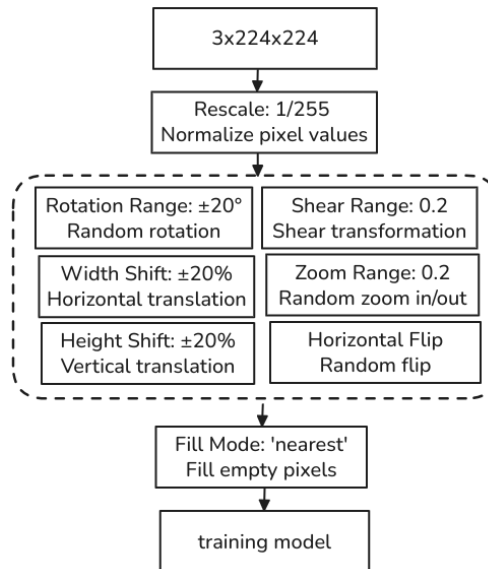


Gambar 4. Activity Diagram Sistem

Sistem chatbot klasifikasi batik dirancang dengan arsitektur modular yang mengintegrasikan *computer vision* dan NLP melalui lima komponen utama sebagaimana ditunjukkan pada Gambar 4. Antarmuka pengguna menyediakan dua modalitas interaksi yaitu klasifikasi citra batik dan *query* tekstual melalui *chatbot*. Citra input mengalami *preprocessing* berupa *resize* 224x224 piksel dan normalisasi sebelum diklasifikasikan menggunakan model MobileNet berbasis *depthwise separable convolutions* untuk identifikasi kategori batik. Sistem mengimplementasikan RAG dengan memanfaatkan model *Multilingual-e5-large* untuk transformasi *query* menjadi *vector embedding*, kemudian *database vector* Qdrant melakukan *similarity search* menggunakan *cosine similarity* untuk mengambil *chunk* dokumen relevan. Proses *retrieval* dapat dilakukan secara *conditional* dengan filter kategori hasil klasifikasi atau *unconditional* untuk *query* umum. Konteks dokumen hasil *retrieval*, *template prompt*, dan *query* pengguna dikombinasikan sebagai input bagi model Mistral-7B untuk menghasilkan *respons*. Integrasi antar komponen dilakukan secara *asynchronous* melalui API. Pendekatan RAG digunakan untuk melengkapi *knowledge model* bahasa dengan informasi dari *external knowledge base* yang tersimpan dalam *database vector*.

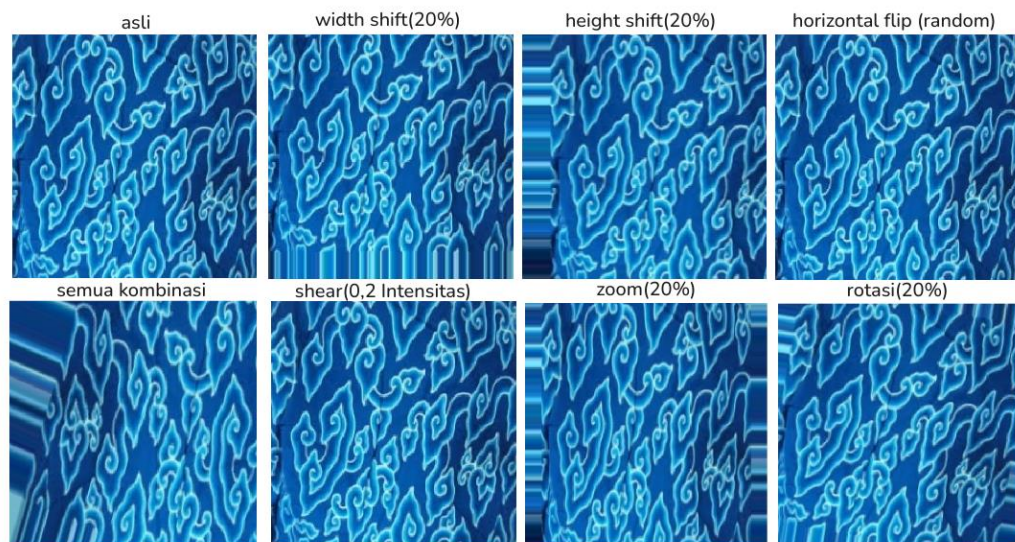
2.2 Augmentasi Citra

Augmentasi Citra meningkatkan kualitas dan variabilitas dataset seperti yang ditunjukkan pada Gambar 5.



Gambar 5.Augmentasi Citra

Citra diubah menjadi dimensi tetap $3 \times 224 \times 224$ piksel. Selanjutnya, nilai piksel masing-masing *channel* dikonversi ke skala $[0,1]$ dengan cara membaginya oleh 255. Untuk meningkatkan variasi data pelatihan, dilakukan augmentasi secara acak. Parameter augmentasi yang diterapkan mencakup beberapa transformasi. Gambar-gambar dirotasi secara acak dalam rentang 20 derajat dan juga mengalami pergeseran horizontal serta vertikal hingga 20%. Selain itu, transformasi shear diterapkan dengan rentang 0.2, dan gambar juga diperbesar (*zoom*) dengan faktor 0.2, serta pencerminan horizontal. Nilai piksel yang hilang akibat transformasi diisi menggunakan mode '*nearest*'. Beberapa hasil dari teknik augmentasi data divisualisasikan pada Gambar 6.

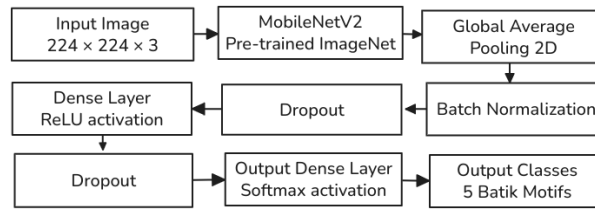


Gambar 6. Contoh Augmentasi Citra Batik

Gambar 6 menunjukkan contoh hasil augmentasi data yang diimplementasikan pada citra batik untuk meningkatkan variabilitas dataset *training*. Delapan jenis transformasi geometris diterapkan meliputi: citra *original* tanpa modifikasi, *width shift* (pergeseran lebar 20%), *height shift* (pergeseran tinggi 20%), *horizontal flip* (pencerminan horizontal acak), kombinasi seluruh transformasi, *shear* (rotasi 2 derajat intensitas), *zoom* (pembesaran 20%), dan *rotate* (rotasi 20%). Setiap transformasi menghasilkan variasi visual berbeda dari citra batik yang sama, memungkinkan model mempelajari invariansi terhadap perubahan geometris dan orientasi. Implementasi augmentasi ini bertujuan untuk meningkatkan *robustness* model terhadap variasi input dan mencegah *overfitting* dengan memperkaya representasi fitur visual yang dapat diekstrak selama proses *training*.

2.3 Fine-Tuning Mobilenet

Arsitektur model klasifikasi menggunakan MobileNetV2 yang telah dimodifikasi untuk klasifikasi lima motif batik seperti yang ditunjukkan pada Gambar 7.

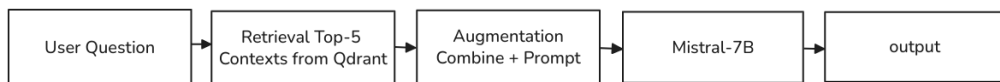


Gambar 7. Fine-Tuning Arsitektur

Input citra berukuran $224 \times 224 \times 3$ diproses melalui *MobileNetV2 base model* yang telah di-*pretrained* pada ImageNet sebagai *feature extractor*. *Output* dari *base model* direduksi menggunakan *Global Average Pooling 2D* untuk menghasilkan vektor fitur satu dimensi. Proses normalisasi dilakukan melalui *Batch Normalization* untuk menstabilkan distribusi aktivasi, kemudian dilanjutkan dengan *dense layer* menggunakan aktivasi ReLU untuk transformasi fitur non-linear. Regularisasi *Dropout* diterapkan untuk mencegah *overfitting* dengan men-deaktivasi neuron. *Layer output* menggunakan *Dense Layer* dengan aktivasi Softmax yang menghasilkan distribusi probabilitas untuk lima kelas motif batik. Arsitektur ini mengadopsi pendekatan *transfer learning* untuk meningkatkan performa klasifikasi dengan memanfaatkan representasi fitur yang telah dipelajari dari dataset ImageNet.

2.4 RAG

Alur kerja sistem RAG dalam penelitian ini mengikuti *pipeline* standar yang terdiri dari lima tahapan berurutan seperti yang ditunjukkan pada Gambar 8.



Gambar 8. RAG

Tahap pertama adalah *user question* dimana pengguna menginputkan pertanyaan terkait motif batik melalui antarmuka sistem. Tahap kedua adalah proses *retrieval* yang mengambil top-5 konteks paling relevan dari Qdrant *Database* berdasarkan *similarity search* antara vektor pertanyaan dan vektor dokumen yang tersimpan. Tahap ketiga adalah *augmentation* yang menggabungkan konteks yang diperoleh dengan *template prompt* terstruktur untuk membentuk *input* bagi model bahasa. Tahap keempat adalah proses *generation* menggunakan model Mistral-7B yang menghasilkan respon berdasarkan konteks yang telah diaugmentasi. Tahap terakhir adalah *output* yang menampilkan jawaban final kepada pengguna berupa penjelasan deskriptif mengenai motif batik yang ditanyakan.

2.5 Metode Evaluasi

2.5.1 Evaluasi Model Klasifikasi

Pengukuran performa model menggunakan empat komponen utama dalam *confusion matrix* yang ditunjukkan pada Gambar 9.

True Positif	False Negatif
False Positif	True Negatif

Gambar 9. Confusion Matrix

Evaluasi performa model klasifikasi MobileNetV2 dilakukan menggunakan metrik utama yaitu akurasi. Akurasi mengukur proporsi prediksi yang benar terhadap total prediksi yang dilakukan [18]. Evaluasi performa model akan dilakukan dengan menghitung akurasi berdasarkan Persamaan (1) dibawah ini.

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{1}$$

True Positive (TP) merepresentasikan jumlah observasi positif yang diidentifikasi dengan tepat oleh model. Sebaliknya, *True Negative (TN)* adalah jumlah observasi negatif yang juga diidentifikasi dengan tepat. Kesalahan prediksi terjadi dalam dua bentuk: *False Positive (FP)*, yaitu ketika observasi negatif keliru dilabeli positif, dan *False Negative (FN)*, yaitu ketika observasi positif keliru dilabeli negatif [12].

2.5.2 Evaluasi RAG

Menilai kinerja sistem RAG merupakan tugas yang kompleks, karena metrik tradisional sering kali gagal menangkap sifatnya yang multi-dimensi [24], [25]. Evaluasi sistem RAG dilakukan menggunakan 4 metode evaluasi dari *framework*



RAGAs (*Retrieval Augmented Generation Assessment*) [26]. Metrik *faithfulness* mengukur konsistensi faktual antara jawaban yang dihasilkan dengan konteks yang di-*retrieve*, dihitung menggunakan Persamaan (2) dibawah ini.

$$Faithfulness = \frac{|Verified\ Statements|}{|Total\ Statements|} \quad (2)$$

dimana *Verified Statements* adalah pernyataan dalam jawaban yang didukung oleh konteks, dan *Total Statements* adalah jumlah total pernyataan dalam jawaban.

Metrik *answer relevancy* mengukur relevansi jawaban terhadap pertanyaan pengguna berdasarkan *cosine similarity*, dihitung menggunakan Persamaan (3) dibawah ini.

$$AR = \frac{1}{n} \sum_{i=1}^n sim(q, q_i) \quad (3)$$

dimana n adalah jumlah pertanyaan yang di-*generate* dari jawaban, $sim(q, q_i)$ adalah *cosine similarity* antara *embedding* pertanyaan asli (q) dan *embedding* pertanyaan yang di-*generate* (q_i)

Metrik *context relevance* mengukur relevansi konteks yang di-*retrieve* terhadap pertanyaan, dihitung menggunakan Persamaan (4) dibawah ini.

$$Context\ Relevance = \frac{\text{Number of sentences retrieved}}{\text{Overall number of sentence in } c(q)} \quad (4)$$

dimana *Relevant Sentences* adalah jumlah kalimat relevan dalam konteks yang di-*retrieve*, dan *Total Sentences* adalah jumlah total kalimat dalam konteks.

Metrik *context utilization* mengukur seberapa efektif sistem memanfaatkan konteks yang di-*retrieve* untuk menghasilkan jawaban, dihitung menggunakan Persamaan (5) untuk menghitung *context utilization* dan Persamaan (6) untuk menghitung *precision* [27].

$$Context\ Utilization@K = \frac{\sum_{k=1}^K (Precision@k \times v_k)}{\text{Sum of relevant items within the } K \text{ highest-ranked}} \quad (5)$$

$$Precision@k = \frac{\text{true positives@k}}{(\text{true positives@k} + \text{false positives@k})} \quad (6)$$

dimana K adalah jumlah total *chunks* (potongan teks) di dalam konteks, dan v_k adalah indikator relevansi (bernilai 1 jika relevan atau 0 jika tidak) pada peringkat K .

3. HASIL DAN PEMBAHASAN

3.1 Performa Model Klasifikasi

Hasil evaluasi performa model *machine learning* yang telah dikembangkan menggunakan 3000 data *training*, 1000 data validasi, dan 1000 data *testing* berdasarkan metrik akurasi ditunjukkan pada Tabel 2.

Tabel 2. Pengujian Akurasi Model

LR Stage1	LR Stage2	Drop1	Drop2	Dense1	Dense2	Batch	Train Acc	Test Acc
0.001	0.0001	0.4	0.3	128	128	32	0.9863	0.967
0.001	0.00005	0.4	0.2	128	128	32	0.9803	0.96
0.001	0.00005	0.3	0.3	128	128	32	0.9847	0.964
0.001	0.0001	0.3	0.3	256	256	32	0.9897	0.968
0.001	0.00005	0.3	0.2	128	128	32	0.9837	0.959

Tabel 2 menampilkan hasil eksperimen *hyperparameter* tuning pada arsitektur *neural network* dengan variasi *learning rate* bertahap (LR Stage1 dan Stage2), *dropout regularization* (Drop1 dan Drop2), jumlah unit *dense layer* (Dense1 dan Dense2), serta ukuran *batch*. Hasil evaluasi lima konfigurasi menunjukkan bahwa model terbaik diperoleh menggunakan *learning rate* sebesar 0.001 untuk tahap pertama dan 0.0001 untuk tahap kedua, *dropout rate* 0.3 pada kedua *layer*, *dense layer* dengan ukuran 256 *nodes* masing-masing, dan *batch size* 32, yang menghasilkan akurasi 98.97% untuk *training* serta 96.8% untuk *testing*. Hasil ini mengindikasikan bahwa kombinasi *regularization* yang moderat dengan arsitektur yang lebih dalam memberikan *trade-off* terbaik antara kapasitas model dan kemampuan generalisasi.

3.2 Performa RAG menggunakan RAGAs

RAGAs (*Retrieval Augmented Generation Assessment*) adalah *framework* untuk mengevaluasi sistem RAG yang mengukur kualitas dari proses pengambilan informasi (*retrieval*) dan pembuatan jawaban (*generation*). *Framework* ini menggunakan beberapa metrik utama seperti *faithfulness* untuk mengukur konsistensi jawaban dengan konteks yang diambil, *answer relevancy* untuk menilai relevansi jawaban terhadap pertanyaan, *context precision* dan *context recall* untuk mengukur akurasi dan kelengkapan konteks yang di-*retrieve*, serta *context relevancy* untuk mengevaluasi kesesuaian konteks dengan pertanyaan. RAGAs memungkinkan evaluasi sistematis terhadap sistem RAG tanpa



memerlukan data berlabel secara manual, sehingga memudahkan identifikasi kelemahan dan optimalisasi performa sistem. Pada Tabel 3 menunjukkan contoh pengujian menggunakan RAGs yang telah dilakukan.

Tabel 3. Contoh pengujian RAG

Pertanyaan (Query)	Expected Output (Ground Truth)	Real Output (Nemotron-12B)	Real Output (Mistral-7B)
Siapa yang dipercaya menciptakan motif Batik Truntum?	Motif ini diciptakan oleh Kanjeng Ratu Kencana (atau Ratu Beruk), yang merupakan permaisuri dari Paku Buwono III Surakarta.	Kanjeng Ratu Kencana dipercaya menciptakan motif Batik Truntum. Informasi ini disebutkan dalam...	Motif batik ini diciptakan oleh Kanjeng Ratu Kencana (Permaisuri Ingkang Sinuhun Sri Susuhunan Pakubuwana III dari Surakarta).
Apa makna di balik nama 'Truntum'?	Truntum memiliki arti 'timbul kembali' atau bersemi, yang melambangkan cinta kasih (katresnan) suami istri yang bersemi kembali.	Nama "Truntum" berasal dari kata "tuntum" yang berarti "bersemi kembali" atau "tumbuh lagi", melambangkan cinta yang...	Nama "Truntum" berasal dari kata tuntum (bersemi kembali), melambangkan cinta yang tumbuh lagi.

Analisis ini difokuskan pada evaluasi performa *pipeline Retrieval-Augmented Generation (RAG)* melalui 26 skenario pengujian yang telah ditentukan. Kumpulan data evaluasi, yang didefinisikan sebagai variabel *raw_dataset*, mencakup pasangan "query" (pertanyaan) dan "answer" (jawaban *ground truth*). Metodologi pengujian ini dirancang secara spesifik untuk melakukan analisis komparatif performa RAG pada *database* pengetahuan batik dengan menggunakan dua model bahasa besar (LLM) yang berbeda: Nemotron-12B (nvidia/nemotron-nano-12b-v2-v1) dan Mistral-7B (mistralai/mistral-7b-instruct;free). Hasil eksperimen yang telah dilakukan ditampilkan pada Tabel 4 dibawah ini.

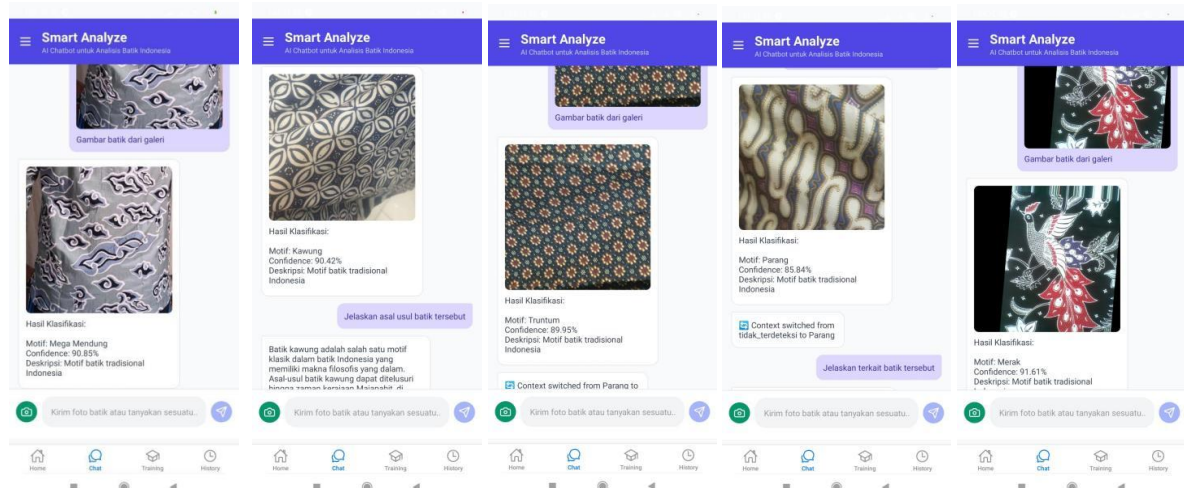
Tabel 4. Hasil Eksperimen

Model	Faithfulness	Answer Relevancy	Context Relevance	Context Utilization	Avg All Metrics
Mistral-7B	0.956044	0.877081	0.9991	0.785897	0.904755
Nemotron-12B	0.85989	0.826321	0.961538	0.790759	0.859627

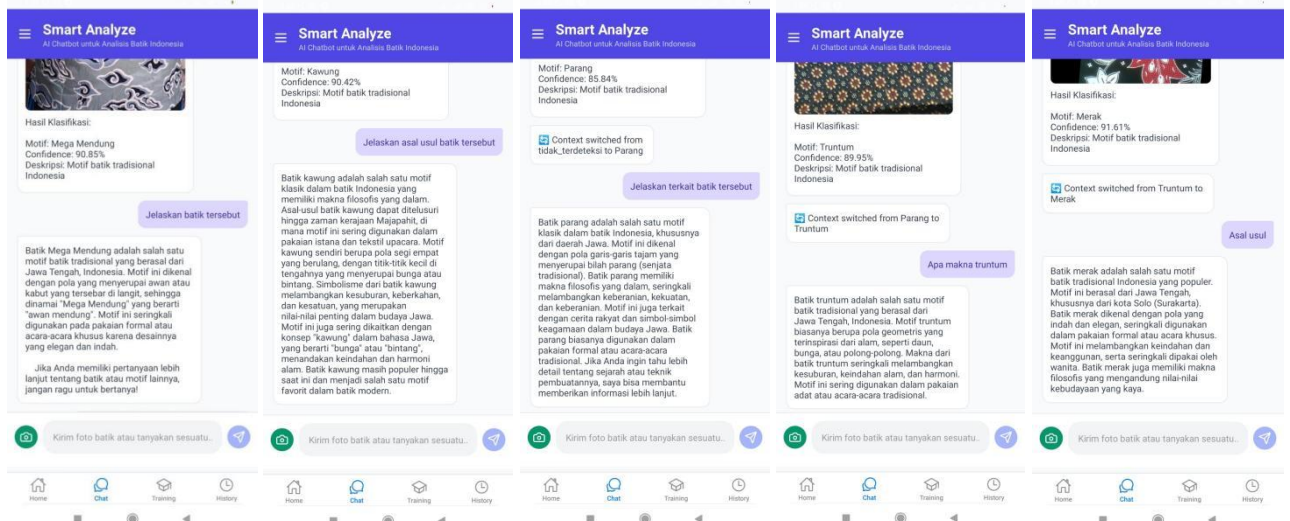
Tabel 4 menunjukkan hasil evaluasi performa dua model LLM menggunakan metrik RAGs. Model Mistral-7B menunjukkan performa dengan rata-rata metrik sebesar 0.904755, lebih tinggi dibandingkan Nemotron-12B yang memperoleh rata-rata 0.859627. Mistral-7B unggul pada metrik *faithfulness* (0.956044), *answer relevancy* (0.877081), *context relevance* (0.9991), dan *context utilization* (0.785897). Sementara itu, Nemotron-12B hanya unggul pada metrik *context relevance* (0.961538). Berdasarkan evaluasi ini, Mistral-7B dipilih sebagai model LLM untuk sistem RAG karena menunjukkan konsistensi dan akurasi yang lebih baik dalam menghasilkan jawaban yang relevan dan sesuai konteks.

3.3 Implementasi Sistem

Implementasi sistem telah diintegrasikan dalam aplikasi *mobile* dengan antarmuka sesuai arsitektur sistem yang diusulkan. Pada Gambar 10 menunjukkan contoh gambar hasil pengenalan batik yang dilakukan dan Gambar 11 menunjukkan contoh hasil implementasi *chatbot*.



Gambar 10. Contoh Pengenalan Motif Batik oleh Sistem



Gambar 11. Contoh Hasil Chatbot

Sistem menyediakan fungsi upload gambar dari galeri yang diproses melalui model MobileNet V2 untuk klasifikasi motif batik. Hasil klasifikasi ditampilkan secara terstruktur meliputi nama motif yang terdeteksi, nilai *confidence score*, dan deskripsi singkat. Sistem dilengkapi *interface* input teks yang memungkinkan pengguna mengajukan pertanyaan kepada sistem RAG untuk memperoleh informasi detail mengenai motif yang teridentifikasi. Navigasi aplikasi menggunakan *bottom navigation* bar untuk memudahkan akses ke berbagai fitur sistem. Implementasi ini menunjukkan integrasi yang efektif antara modul klasifikasi citra menggunakan *deep learning* dan modul *conversational AI* berbasis RAG dalam satu platform *mobile* yang responsif.

3.4 Pengujian

Hasil pengujian blackbox terhadap fungsionalitas utama sistem yang telah diimplementasikan ditunjukkan pada Tabel 5.

Tabel 5. Pengujian

Tes Scenario	Expected Result	Actual Result	Status
Klasifikasi melalui upload galeri	Confidence $\geq 83\%$	teridentifikasi lebih 83%	Berhasil
Query makna motif melalui RAG	memberikan jawaban relevan	Jawaban sesuai konteks	Berhasil

Pengujian dilakukan pada dua skenario utama untuk memvalidasi performa sistem secara keseluruhan. Skenario pertama menguji fungsi klasifikasi melalui upload galeri dengan ekspektasi *confidence score* minimal 83%, dan hasil pengujian menunjukkan lebih dari 83% dengan status berhasil. Skenario kedua menguji fungsi *query* makna motif melalui sistem RAG dengan ekspektasi sistem memberikan jawaban relevan, dan hasil pengujian menunjukkan jawaban yang dihasilkan sesuai dengan konteks pertanyaan dengan status berhasil. Hasil pengujian *blackbox* mengindikasikan bahwa kedua modul utama sistem berfungsi sesuai kriteria yang ditetapkan sebelumnya.

4. KESIMPULAN

Penelitian ini telah berhasil mengembangkan sistem klasifikasi batik berbasis mobile yang mengintegrasikan arsitektur MobileNetV2 dengan chatbot *Retrieval-Augmented Generation (RAG)* untuk memberikan solusi komprehensif dalam identifikasi dan pembelajaran motif batik Indonesia. Model MobileNetV2 mencapai akurasi *training* 98.97% dan akurasi *testing* 96.8% dalam mengklasifikasikan lima kategori motif batik yaitu Parang, Truntum, Kawung, Mega mendung, dan Merak. Konfigurasi optimal model diperoleh dengan *learning rate* 0.001 untuk *stage* pertama dan 0.0001 untuk *stage* kedua, *dropout rate* 0.3 untuk kedua *layer*, serta *batch size* 32. Hasil pengujian *blackbox* menunjukkan sistem mampu melakukan klasifikasi melampaui target minimal yang ditetapkan. Integrasi *chatbot* berbasis RAG menggunakan *framework* LangChain dan Qdrant *vector database* memberikan kontribusi dalam aspek edukatif sistem. Evaluasi menggunakan metrik RAGs menunjukkan model Mistral-7B menghasilkan performa dengan rata-rata skor 0.904755, unggul dalam *faithfulness* (0.956044), *answer relevancy* (0.877081), dan *context relevance* (0.991). Hal ini mengindikasikan kemampuan sistem memberikan jawaban relevan dan sesuai koInteks berdasarkan 2.748 chunks pengetahuan dari 300 kata kunci terkait batik. Implementasi sistem dalam aplikasi mobile menggunakan *react native* mendemonstrasikan integrasi antara modul klasifikasi citra dan *conversational AI* dalam platform yang *user-friendly*. Keterbatasan penelitian mencakup cakupan dataset yang masih terbatas pada lima motif batik utama dan sistem RAG yang bergantung pada kualitas basis pengetahuan yang memerlukan pembaruan berkala. Untuk penelitian selanjutnya, disarankan untuk memperluas dataset mencakup lebih banyak variasi motif batik dari berbagai daerah, mengimplementasikan pendekatan multimodal yang mengintegrasikan input citra dan teks secara bersamaan untuk



meningkatkan akurasi klasifikasi dan kualitas respons sistem, serta mengembangkan mekanisme *feedback loop* yang memungkinkan pengguna memberikan koreksi untuk peningkatan performa berkelanjutan.

REFERENCES

- [1] M. H. P. S. Ikhyia Ulummuddin, Anggraini Puspita Sari, "Klasifikasi Motif Batik Yogyakarta Menggunakan Metode GLCM Dan CNN," *J. Teknol. Terpadu*, vol. 6, no. 22, pp. 72–78, 2020, doi: <https://doi.org/10.54914/jtt.v10i2.1451>.
- [2] E. Budimansyah, D. Ariawan, R. D. Maulana, W. R. Yudha, and M. Munsarif, "Klasifikasi Motif Batik Indonesia Menggunakan Convolutional Neural Network (Cnn)," *J. Komput. Dan Teknol. Inf.*, vol. 3, no. 2, pp. 1–8, 2025, doi: 10.26714/jkti.v3i2.18627.
- [3] F. Okti Yolanda and A. Putra, "Systematic Literature Review: Eksplorasi Etnomatematika pada Motif Batik," *Prima Magistra J. Ilm. Kependidikan*, vol. 3, no. 2, pp. 188–195, Mar. 2022, doi: 10.37478/jpm.v3i2.1533.
- [4] L. Sariroh, R. N. Aulia, and O. Purnawirawan, "Systematic Literature Review: Peran Kearifan Lokal Masyarakat Indonesia dalam Melestarikan Budaya Batik di Era Revolusi Industri 4.0," in *Seminar Nasional Industri Kerajinan dan Batik*, D.04, 2024, p. D.04 1-14. [Online]. Available: <https://proceeding.batik.go.id/index.php/SNBK/article/view/251>
- [5] W. Sasmita, M. N. Muzaki, R. N. Safitri, R. Mattin, L. Lensi, and M. Alfian, "Pengembangan Produk Batik dalam Usaha Menarik Minat Anak Muda Terhadap Produk Khas Kelurahan Dandangan," *J. Pengabdian Kpd. Masy.*, vol. 3, no. 2, pp. 219–231, 2024, doi: <https://doi.org/10.55506/arch.v3i2.97>.
- [6] M. U. N. Saputra and K. B. Prasetyo, "Reproduksi Budaya Batik Milenial: Upaya Pelestarian dan Inovasi Batik Tradisional di Identik Batik Semarang," *J. Paradig. J. Sociol. Res. Educ.*, vol. 4, no. 2, pp. 126–140, Dec. 2023, doi: 10.53682/jpsre.v4i2.8046.
- [7] G. P. I. Fanani, Y. Safitri, M. A. Mu'min, S. A. Wijaya, T. S. Famuji, and N. Trisanti, "Pengenalan Citra Batik Tradisional Menggunakan Deep Learning untuk Klasifikasi Motif Daerah," *Sci. J. Comput. Sci. Inform.*, vol. 2, no. 1, pp. 1–7, 2025, doi: 10.34304/scientific.v2i1.336.
- [8] I. Fanani, "Implementasi Retrieval Augmented Generation untuk Evaluasi Proposal Tugas Akhir Mahasiswa," *J. Teknol. Komput. Dan Inform.*, vol. 3, no. 2, 2025, doi: <https://doi.org/10.59820/tekominfo.v3i2.336>.
- [9] J. P. Nayinzira and M. Adda, "SentimentCareBot: Retrieval-Augmented Generation Chatbot for Mental Health Support with Sentiment Analysis," *Procedia Comput. Sci.*, vol. 251, pp. 334–341, 2024, doi: 10.1016/j.procs.2024.11.118.
- [10] S. Arifin and N. Nurfaizah, "Klasifikasi Motif Batik Menggunakan Metode Convolutional Neural Network (CNN) Dengan Multi Class Classification," *J. Ilm. IT CIDA*, vol. 10, no. 1, p. 30, 2024, doi: 10.55635/jic.v10i1.206.
- [11] M. M. A. Wona *et al.*, "Klasifikasi Batik Indonesia Menggunakan Convolutional Neural Network (CNN)," *J. Rekayasa Teknol. Inf. JURTI*, vol. 7, no. 2, p. 172, Dec. 2023, doi: 10.30872/jurti.v7i2.13694.
- [12] D. Sinaga, C. Jatmoko, S. Suprayogi, and N. Hedriyanto, "Multi-Layer Convolutional Neural Networks for Batik Image Classification," *Sci. J. Inform.*, vol. 11, no. 2, pp. 477–484, May 2024, doi: 10.15294/sji.v11i2.3309.
- [13] H. Sastypratiwi and H. Muhandi, "Batik Recognition and Classification Using Transfer Learning and MobileNet Approach," *Int. J. Inform. Vis.*, vol. 8, no. 4, pp. 2400–2410, 2024, doi: <http://dx.doi.org/10.62527/joiv.8.4.2407>.
- [14] E. Khoirunnisa *et al.*, "Enhanced Semarang Batik Classification using MobileNetV2 and Data Augmentation," *sinkron*, vol. 9, no. 1, pp. 43–54, Jan. 2025, doi: 10.33395/sinkron.v9i1.14308.
- [15] R. Andrian, R. Taufik, D. Kurniawan, A. S. Nahri, and H. C. Herwanto, "Lampung Batik Classification Using AlexNet, EfficientNet, LeNet and MobileNet Architecture," *Int. J. Adv. Comput. Sci. Appl.*, vol. 15, no. 11, pp. 930–935, 2024, doi: 10.14569/IJACSA.2024.0151191.
- [16] R. Andriawan, M. S. Afrian, and G. E. Perkasa, "Identifikasi Jenis Rimpang Menggunakan Metode Convolutional Neural Network Berbasis Website," *Pros. Semin. Nas. Teknol. Dan Sains*, vol. 4, no. 2, pp. 273–280, 2025.
- [17] Nafidanisa, D. Aldo, and Nicolaus, "Klasifikasi Motif Batik Semarang Menggunakan Convolutional Neural Network Dengan VGG16," *E-Proceeding Eng.*, vol. 12, no. 4, p. 6814, 2025.
- [18] I. Fathurrahman, M. Djamaluddin, Z. Amri, and M. N. Wathani, "Klasifikasi Motif Batik Nusantara Menggunakan Vision Transformer (ViT) Berbasis Deep Learning," *Infotek J. Inform. Dan Teknol.*, vol. 8, no. 2, pp. 511–522, Jul. 2025, doi: 10.29408/jit.v8i2.31108.
- [19] P. T. Informatika and U. Abdurrab, "Optimasi Chatbot dengan Pemanfaatan Natural Language Processing," *J. Komput. Dan Teknol. Inf.*, vol. 10, no. 1, pp. 17–26, 2024, doi: <https://doi.org/10.35143/jkt.v10i1.6181>.
- [20] A. K. Umam, E. Wijayanti, and A. A. Chamid, "Penerapan NLP Pada Chatbot Telegram Untuk Informasi Seputar Handphone," *Bull. Comput. Sci. Res.*, vol. 5, no. 4, pp. 329–339, 2025, doi: 10.47065/bulletincsr.v5i4.540.
- [21] S. Elysia and Herianto, "Chatbot Berbasis Retrieval Augmented Generation (RAG) untuk Peningkatan Layanan Informasi Sekolah," *J. TIFDA Technol. Inf. Data Anal.*, vol. 1, no. 2, pp. 52–58, Dec. 2024, doi: 10.70491/tifda.v1i2.52.
- [22] M. L. Husain, Y. Wibisono, and A. Anisyah, "Development of an Academic Services Chatbot Based on Retrieval-Augmented Generation (RAG)," *Brill. Res. Artif. Intell.*, vol. 5, no. 2, pp. 727–735, Aug. 2025, doi: 10.47709/brilliance.v5i2.6719.
- [23] L. S. Hartono, E. I. Setiawan, and V. Singh, "Retrieval Augmented Generation-Based Chatbot for Prospective and Current University Students," *Int. J. Eng. Sci. Inf. Technol.*, vol. 5, no. 3, pp. 268–277, 2025, doi: 10.52088/ijesty.v5i3.951.
- [24] I. P. H. Putro, J. Antoni, M. K. Adhitya, and ..., "Retrieval-Augmented Generation (RAG) Chatbot for Handling Customer Complaints in the Energy Sector," *J. Infomedia Edots*, vol. 10, no. 2, pp. 105–111, 2025, doi: <http://dx.doi.org/10.30811/jim.v10i2.7169>.
- [25] T. Q. Ramadhani, N. Q. Nada, and N. D. S., "Penerapan Metode Retrieval-Augmented Generation (RAG) Pada Chatbot E-Commerce Berbasis Gemini Ai," *J. Ilm. Ilk. - Ilmu Komput. Inform.*, vol. 8, no. 2, pp. 301–313, 2025, doi: 10.47324/ilkominfo.v8i2.384.
- [26] S. Es, J. James, L. Espinosa-Anke, S. Schockaert, and E. Gradients, "RAGAS: Automated Evaluation of Retrieval Augmented Generation," in *System Demonstrations*, Association for Computational Linguistics, 2024, pp. 150–158. doi: 10.18653/v1/2024.eacl-demo.16.
- [27] Ragas Documentation, "Context utilization." Nov. 03, 2025. [Online]. Available: https://docs.ragas.io/en/v0.1.21/concepts/metrics/context_utilization.html