

Rekognisi Tulisan Kaligrafi Dengan Menggunakan Metode *Convolutional Neural Network* Arsitektur *MobileNetV2*

Sami Irhamnillah^{1*}, Aldy Rialdy Atmadja², Ichsan Taufik³

^{1*,2,3}*Informatics Department, Faculty of Science and Technology, UIN Sunan Gunung Djati Bandung*

*Email: samiirhamnillah03@gmail.com

Abstract

This research aims to develop an automatic classification model to recognize the type of Arabic calligraphy writing using MobileNetV2 Convolutional Neural Network (CNN) architecture. Arabic calligraphy has a visual uniqueness and complexity of letterforms that become a challenge in the classification process, especially for ordinary people. The four main calligraphy types used in this research are Tsulust, Naskhi, Diwani, and Kufi. The research follows the CRISP-DM stages which include business understanding, data understanding, data preparation, modeling, evaluation, and deployment. The dataset used is the HICMA dataset consisting of 1.594 calligraphy images. The data is processed through cropping, normalization, and resizing to 224x224 pixels. The model was trained with epoch variations (10, 20, 30, and 40) to obtain the best configuration. The results show that the model at the 20th epoch has the most optimal performance with a testing accuracy of 97.52%. Evaluation of classification metrics showed high F1-Score values in the majority classes. The previously low-performing Kufi class was improved through data augmentation techniques to obtain an F1-Score value of 1.00. The model is then integrated into a Flask-based web application that allows users to upload images and receive classification results directly. The results of this research show that MobileNetV2 is effective for Arabic calligraphy type classification and can be practically implemented for educational purposes as well as digital preservation of Islamic culture.

Keywords: Arabic calligraphy, cnn, data augmentation, image classification, mobilenetv2

Abstrak

Penelitian ini bertujuan mengembangkan model klasifikasi otomatis untuk mengenali jenis tulisan kaligrafi Arab menggunakan arsitektur *Convolutional Neural Network* (CNN) *MobileNetV2*. Kaligrafi Arab memiliki keunikan visual dan kompleksitas bentuk huruf yang menjadi tantangan tersendiri dalam proses klasifikasi, terutama bagi masyarakat awam. Empat jenis kaligrafi utama yang digunakan dalam penelitian ini adalah Tsulust, Naskhi, Diwani, dan Kufi. Penelitian mengikuti tahapan *CRISP-DM* yang meliputi pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, dan deployment. Dataset yang digunakan merupakan dataset HICMA yang terdiri dari 1.594 citra kaligrafi. Data diproses melalui tahap *cropping*, normalisasi, dan *resizing* ke ukuran 224x224 piksel. Model dilatih dengan variasi epoch (10, 20, 30, dan 40) untuk memperoleh konfigurasi terbaik. Hasil menunjukkan bahwa model pada epoch ke-20 memiliki performa paling optimal dengan akurasi pengujian sebesar 97,52%. Evaluasi metrik klasifikasi menunjukkan nilai *F1-Score* tinggi pada kelas mayoritas. Kelas Kufi yang sebelumnya memiliki performa rendah ditingkatkan melalui teknik augmentasi data hingga memperoleh nilai *F1-Score* sebesar 1,00. Model ini kemudian diintegrasikan ke dalam aplikasi web berbasis *Flask* yang memungkinkan pengguna mengunggah gambar dan menerima hasil klasifikasi secara langsung. Hasil penelitian ini menunjukkan bahwa *MobileNetV2* efektif untuk klasifikasi jenis tulisan kaligrafi Arab dan dapat diimplementasikan secara praktis untuk keperluan edukasi serta pelestarian budaya Islam secara digital.

Kata kunci: Kaligrafi arab, cnn, augmentasi data, klasifikasi citra, *mobilenetv2*

1. Pendahuluan

Kaligrafi merupakan seni menulis indah yang memiliki nilai estetika tinggi dan mencerminkan kekayaan budaya Islam. Setiap gaya kaligrafi seperti Naskhi, Riq'ah, Thuluth, dan lainnya memiliki ciri khas dalam bentuk, struktur, dan goresan huruf yang membedakannya satu sama lain. Namun, proses pengenalan jenis tulisan kaligrafi secara manual seringkali sulit dilakukan, terutama bagi orang awam yang tidak memiliki pengetahuan khusus, karena kompleksitas struktur huruf dan banyaknya variasi bentuk yang digunakan [1]. Kesulitan ini menjadi tantangan tersendiri dalam pelestarian, pembelajaran, dan pengenalan seni kaligrafi Arab [2].

Beberapa penelitian telah dilakukan untuk mengatasi tantangan ini dengan memanfaatkan teknologi digital dan kecerdasan buatan. Misalnya, penelitian "Klasifikasi Jenis Tulisan Kaligrafi Menggunakan Metode CNN" menunjukkan bahwa CNN efektif dalam membedakan jenis tulisan kaligrafi berdasarkan fitur visual [2]. Aplikasi *Khatti* juga dikembangkan sebagai media interaktif untuk melatih penulisan kaligrafi Arab [3]. Pendekatan *autoencoder* juga pernah digunakan untuk pengenalan otomatis jenis kaligrafi [1]. Selain itu, penelitian *Dhad—A Children's Handwritten Arabic Characters Dataset for Automated Recognition* tentang dataset "Dhad" membuktikan efektivitas CNN dalam mengenali karakter Arab [4]. Terakhir, penelitian "*A Deep Learning Based Prediction of Arabic Manuscripts Handwriting Style*" menunjukkan bahwa MobileNet mampu mencapai akurasi tinggi yaitu 0.9623 dalam klasifikasi gaya tulisan Arab [5]. Ini membuktikan potensi CNN, khususnya MobileNetV2, dalam tugas-tugas visual seperti pengenalan kaligrafi.

Meskipun berbagai pendekatan telah dilakukan, tantangan masih muncul dalam hal ketersediaan dataset kaligrafi yang representatif dan kebutuhan akan model yang ringan namun akurat. Oleh karena itu, penelitian ini bertujuan untuk mengembangkan model klasifikasi jenis tulisan kaligrafi Arab dengan memanfaatkan

arsitektur MobileNetV2. Fokus utama penelitian ini adalah menguji efektivitas MobileNetV2 dalam mengenali empat jenis kaligrafi: Naskhi, Thuluth, Diwani, dan Kufi [6]. Penelitian ini diharapkan dapat memberikan kontribusi dalam bidang pelestarian budaya, pengembangan teknologi edukatif, serta mendukung pengenalan kaligrafi Arab secara otomatis dan efisien.

2. Metoda Penelitian

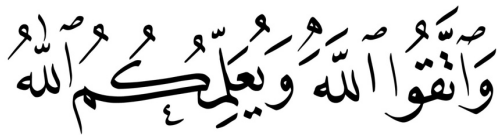
Penelitian ini menggunakan metode CRISP-DM (*Cross-Industry Standard Process for Data Mining*), yang merupakan pendekatan sistematis yang dimaksudkan untuk menyelesaikan perancangan dan pelatihan model serta proses data mining yang efektif yang terdiri dari enam tahap [7], [8]. Dalam penelitian ini, seluruh proses CRISP-DM digunakan hingga tahap *deployment*.

2.1. Business Understanding

Penelitian ini berfokus pada pembuatan model klasifikasi jenis tulisan kaligrafi Arab yang diharapkan dapat membantu mengidentifikasi gaya kaligrafi secara efektif dan akurat. Kesulitan dalam membedakan gaya kaligrafi Arab, terutama bagi masyarakat awam, menjadi tantangan dalam pelestarian dan pengembangan seni Islam, serta berpotensi menghambat proses digitalisasi dan pembelajaran kaligrafi. Dari hal tersebut, tujuan dilakukannya penelitian ini adalah menemukan solusi untuk pengenalan otomatis jenis tulisan kaligrafi Arab agar proses identifikasi dapat dilakukan secara cepat dan tepat, sehingga memperluas aksesibilitas dan pemahaman masyarakat terhadap seni kaligrafi. Dengan mengusulkan model MobileNetV2, yang memanfaatkan keunggulan arsitektur *deep learning* dalam mengekstraksi fitur visual secara otomatis, penelitian ini diharapkan dapat meningkatkan akurasi identifikasi pada klasifikasi gambar kaligrafi Arab.

2.2. Data Understanding

Penelitian ini menggunakan dataset HICMA (*Handwriting Identification for Calligraphy and Manuscripts in Arabic*) yang dipublikasikan sebagai dataset pertama yang tersedia secara terbuka untuk pengenalan tulisan tangan dan kaligrafi Arab pada manuskrip serta karya seni nyata. Dataset ini terdiri dari 5.031 gambar yang terbagi ke dalam lima kelas utama, yaitu gaya Kufic (Gambar 4), Thuluth (Gambar 2), Naskh (Gambar 1), Diwani (Gambar 3), dan Muhaqaaq. Setiap kelas mencerminkan karakteristik visual yang khas, sehingga sangat penting untuk ekstraksi fitur dan klasifikasi jenis tulisan kaligrafi Arab. Dataset ini dikumpulkan dari berbagai sumber, antara lain situs *Free Islamic Calligraphy*, manuskrip Ibn Bawab Qur'an dari Chester Beatty Library, serta koleksi pribadi manuskrip dan tulisan religius abad ke-17 [9].



Gambar 1. Naskhi



Gambar 2. Tsulust



Gambar 3. Diwani

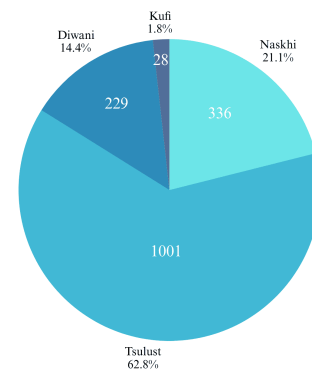


Gambar 4. Kufi

2.2.1 Deskripsi Dataset

Dataset awal yang ditemukan terdiri dari 5.031 gambar terbagi beberapa kelas tulisan kaligrafi Arab, yaitu naskhi, kufi, tsulust, diwani, dan Muhaqaaq. Namun, dalam

penelitian ini, hanya empat kelas yang dipilih dan digunakan untuk membangun model, yaitu Naskhi 336 gambar, Tsulust 1001 gambar, Diwani 229 gambar, dan Kufi 28 gambar, total gambar yang akan digunakan berjumlah 1594 gambar. Distribusi proporsi jumlah gambar dari masing-masing kelas dapat dilihat pada Gambar 5 dalam bentuk diagram pie.

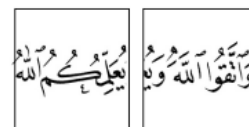


Gambar 5. Jumlah Dataset

Alasan pemilihan keempat kelas ini adalah karena keempatnya merupakan jenis kaligrafi yang paling banyak ditemukan di Indonesia menurut seorang kaligrafer berasal dari Indramayu dan cukup representatif untuk membangun model awal pengenalan jenis tulisan kaligrafi.

2.3. Data Preparation

Pada tahap Data Preparation, dataset diolah agar siap digunakan dalam model klasifikasi [10]. Karena dataset awal berupa gambar kalimat Arab yang berbentuk panjang, dilakukan proses *cropping* sehingga setiap satu gambar kalimat dijadikan dua bagian. Contoh hasil *cropping* dapat dilihat pada Gambar 6.



Gambar 6. Cropping Image

Hal ini bertujuan untuk meningkatkan jumlah sampel serta memastikan setiap input

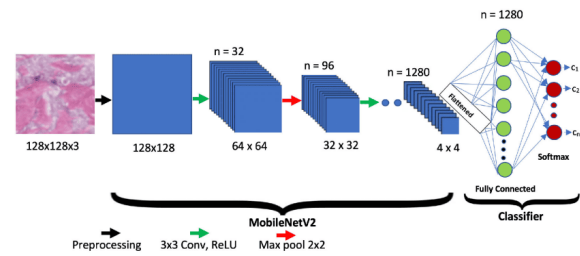
yang masuk ke model memiliki proporsi yang relevan dan tidak terlalu panjang.

Setelah proses cropping, setiap gambar diubah ke ukuran seragam 224x224 piksel, sedangkan nilai piksel dinormalisasi untuk mempersiapkan data piksel agar sesuai dengan kebutuhan input model MobileNetV2 [11]. Normalisasi ini dilakukan dengan mengubah rentang nilai piksel yang awalnya berada antara 0 hingga 255 menjadi rentang standar. Umumnya, setiap nilai piksel dibagi dengan 255 sehingga berada pada kisaran $[0, 1]$ [12]. Namun, MobileNetV2 khususnya ketika digunakan dengan bobot pra-latih (*pretrained*) melalui pustaka Keras memerlukan input dengan rentang nilai antara $[-1, 1]$ untuk menghasilkan performa terbaik [13]. Oleh karena itu, diperlukan langkah tambahan untuk mengubah nilai dari rentang $[0, 1]$ menjadi $[-1, 1]$. Proses ini dilakukan dengan mengalikan nilai piksel dengan 2 lalu mengurangnya dengan 1. Transformasi ini bertujuan untuk menyelaraskan distribusi data masukan dengan kondisi saat model awal dilatih, sehingga dapat mempercepat proses pelatihan, meningkatkan stabilitas konvergensi, serta meningkatkan akurasi model [14].

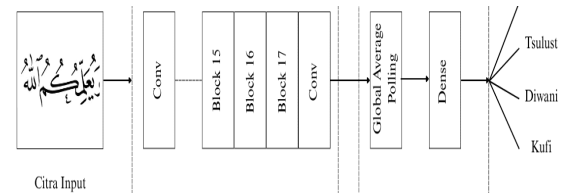
Dataset kemudian dibagi menjadi dua bagian: training dan test dengan pembagian 80% untuk data latih, 10% data validasi dan 10% untuk data uji [15].

2.4. Modeling

Pada tahap pemodelan, digunakan arsitektur *deep learning* MobileNetV2 dengan bobot awal (*pre-trained weight*) dari ImageNet untuk membangun model klasifikasi. Model dilatih menggunakan data latih dengan optimizer adam dan nilai learning rate default sebesar 0.001, *loss function* yang digunakan adalah *sparse_categorical_crossentropy*, sedangkan *metrics* menggunakan *accuracy* dan ukuran batch yang diterapkan selama pelatihan adalah 32. Selama proses pelatihan, dilakukan evaluasi performa model secara berkala.



Gambar 7. Arsitektur MobileNetV2 [16]



Gambar 8. Arsitektur MobileNetV2 yang digunakan

Gambar 7 menunjukkan arsitektur dasar MobileNetV2 yang digunakan sebagai fondasi dalam penelitian ini. Pada gambar 8 merupakan arsitektur MobileNetV2 yang digunakan pada penelitian ini. Model dimuat tanpa lapisan klasifikasi teratas, kemudian ditambahkan lapisan *Global Average Pooling* dan *Dense* untuk menghasilkan output klasifikasi empat kelas kaligrafi: Naskhi, Tsulust, Diwani, dan Kufi. Proses klasifikasi diawali dengan ekstraksi fitur menggunakan lapisan *konvolusi* awal dan tiga blok utama MobileNetV2 (blok 15–17), lalu dilanjutkan ke tahap klasifikasi [17]. Fungsi aktivasi softmax digunakan pada lapisan akhir untuk mengubah output menjadi probabilitas dari setiap kelas [18].

2.5. Evaluation

Pada tahap evaluasi model, dilakukan pengukuran kinerja model MobileNetV2 dalam mengklasifikasikan jenis tulisan kaligrafi Arab menggunakan empat metrik utama yaitu *accuracy*, *precision*, *recall*, dan *F1-score*. Keempat metrik ini dipilih untuk memberikan gambaran komprehensif terkait performa model, baik secara keseluruhan maupun pada setiap jenis kaligrafi secara spesifik [19]. Setiap metrik memberikan pandangan yang unik terhadap kinerja model, khususnya dalam menilai ketepatan dan kelengkapan klasifikasi untuk setiap kelas kaligrafi [20]. Penggunaan

keempat metrik ini memastikan evaluasi yang lebih mendalam terhadap performa model, sehingga model tidak hanya memiliki akurasi yang baik secara keseluruhan, tetapi juga andal dalam hal ketepatan, sensitivitas, dan keseimbangan klasifikasi [21], pada masing-masing jenis kaligrafi, yaitu Naskhi, Tsuluts, Diwani, dan Kufi.

2.6. Deployment

Dalam penelitian ini, model MobileNetV2 yang telah dilatih diimplementasikan ke dalam aplikasi berbasis web menggunakan *Flask*. Aplikasi ini dirancang untuk memungkinkan pengguna mengunggah gambar tulisan kaligrafi Arab, yang kemudian akan diklasifikasikan secara otomatis oleh model. Hasil prediksi jenis tulisan kaligrafi ditampilkan beserta nilai *confidence score*. Implementasi ini bertujuan untuk membuktikan bahwa model yang dikembangkan dapat digunakan secara praktis dan mudah diakses oleh pengguna umum, serta menjadi alat bantu edukatif dalam mengenal jenis-jenis kaligrafi Arab. Sistem ini dihosting secara lokal dan kode sumber tersedia sebagai dasar untuk pengembangan lanjutan.

3. Hasil Penelitian

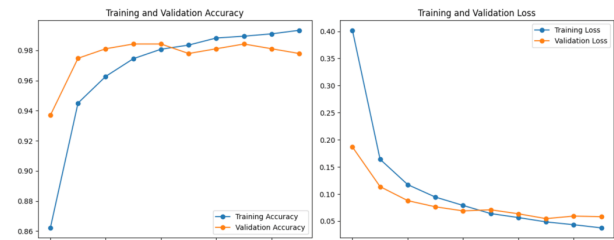
Pelatihan dan pengujian pada penelitian ini dilakukan dengan mencoba beberapa epoch untuk mengetahui epoch berapa yang menghasilkan akurasi model yang bagus. epoch yang dicoba antara lain 10, 20, 30 dan 40 beserta parameter yang sebelumnya telah ditentukan. Hasil pengujian telah dicatat menggunakan tabel untuk pengamatan hasil.

3.1. Hasil Pelatihan Tanpa Augmentasi

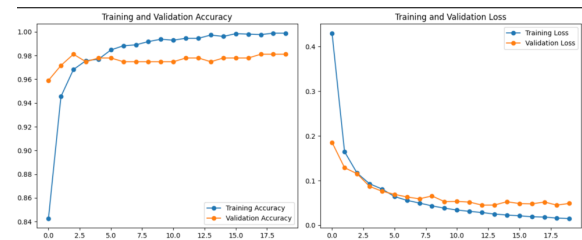
Model dilatih dengan empat skenario epoch, 10, 20, 30, dan 40. Hasil performa model pada masing-masing skenario disajikan pada Tabel 1 dan divisualisasikan pada Gambar 3 hingga Gambar 6.

Tabel 1. Hasil Analisis Epoch

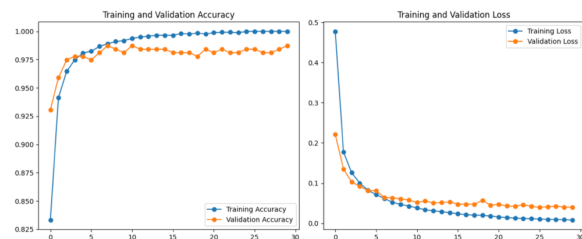
Epoch	Training accuracy	Validation Accuracy	Testing Accuracy
10	99.14%	0.9779	58.22%
20	99.93%	0.9811	97.52%
30	100%	0.9874	96.58%
40	100%	0.9842	97.52%



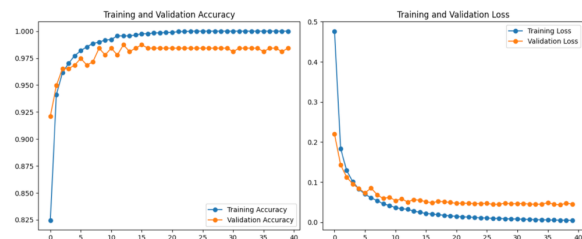
Gambar 9. Epoch 10



Gambar 10. Epoch 20



Gambar 11. Epoch 30

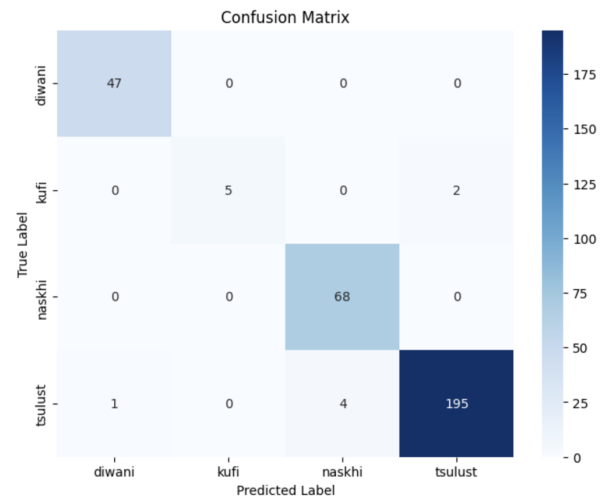


Gambar 12. Epoch 40

Berdasarkan hasil pada Tabel 1, dapat diamati bahwa peningkatan jumlah epoch secara umum menyebabkan peningkatan akurasi pelatihan (*training accuracy*), bahkan mencapai 100% pada epoch 30 dan 40. Namun demikian, tren pada akurasi validasi dan pengujian menunjukkan indikasi bahwa peningkatan jumlah epoch tidak selalu berbanding lurus dengan peningkatan performa model secara keseluruhan. Pada epoch 10, akurasi pelatihan memang tinggi (99.14%), tetapi akurasi validasi rendah (97.79%) dibandingkan dengan epoch 20 yaitu (98.11), yang mengindikasikan kemungkinan model belum cukup belajar atau terdapat kesalahan

dalam proses validasi. Pada epoch 20, model mulai menunjukkan hasil yang optimal dengan akurasi validasi sebesar 98.11% dan akurasi pengujian mencapai 97.52%. Saat epoch ditingkatkan menjadi 30, meskipun akurasi pelatihan dan validasi meningkat, akurasi pengujian justru menurun menjadi 96.58%, yang mengindikasikan awal terjadinya overfitting. Hal ini mengisyaratkan bahwa model terlalu menyesuaikan diri dengan data pelatihan sehingga kemampuan generalisasinya menurun. Pada epoch 40, akurasi validasi sedikit menurun namun akurasi pengujian kembali meningkat ke angka 97.52%, menunjukkan bahwa model masih mampu mempertahankan performa generalisasi yang baik. Berdasarkan hasil ini, dapat disimpulkan bahwa model cenderung mulai overfitting setelah epoch ke-30, dan epoch 20 atau 40 dapat dipertimbangkan sebagai titik optimal pelatihan dengan keseimbangan antara akurasi validasi dan pengujian.

Berdasarkan pertimbangan akurasi pelatihan, validasi, dan pengujian yang telah diuraikan sebelumnya, maka dipilihlah model pada epoch ke-20 sebagai model terbaik. Model ini menunjukkan keseimbangan yang optimal antara kemampuan belajar terhadap data pelatihan dan kemampuan generalisasi terhadap data pengujian, tanpa indikasi overfitting yang signifikan. Oleh karena itu, model pada epoch 20 digunakan untuk evaluasi lebih lanjut menggunakan *confusion matrix* guna mengetahui performa klasifikasi pada masing-masing kelas kaligrafi.



Gambar 13. *Confusion Matrix*

Berdasarkan *confusion matrix* pada Gambar 13, dapat diketahui bahwa model memiliki kemampuan klasifikasi yang sangat baik, khususnya pada kelas Tsulust, Naskhi, dan Diwani. Pada kelas Diwani, sebanyak 47 sampel berhasil diklasifikasikan dengan benar tanpa adanya kesalahan klasifikasi ke kelas lain. Hal ini menunjukkan bahwa fitur visual dari tulisan Diwani cukup khas dan dapat dikenali dengan mudah oleh model. Sementara itu, pada kelas Kufi, hanya 5 sampel yang berhasil diklasifikasikan dengan benar, sedangkan 2 lainnya justru salah diklasifikasikan sebagai Tsulust.

Kondisi ini mencerminkan bahwa model masih mengalami kesulitan dalam mengenali tulisan Kufi, yang kemungkinan besar disebabkan oleh jumlah data latih yang sangat terbatas dibandingkan dengan kelas lainnya. Untuk kelas Naskhi, model menunjukkan performa yang sangat baik dengan mengklasifikasikan seluruh 68 sampel secara benar, tanpa adanya kesalahan prediksi.

Hal ini mengindikasikan bahwa model telah memahami pola karakteristik tulisan Naskhi dengan baik. Adapun pada kelas Tsulust, sebanyak 195 sampel berhasil diklasifikasikan dengan benar, meskipun terdapat beberapa kesalahan klasifikasi, yaitu lima sampel diprediksi sebagai Naskhi dan satu sampel sebagai Diwani. Ini menunjukkan bahwa meskipun tingkat akurasi klasifikasi Tsulust sangat tinggi, masih terdapat sedikit tumpang tindih dalam fitur visual antara kelas Tsulust dengan kelas lainnya. Secara keseluruhan, confusion matrix ini memperlihatkan bahwa model pada epoch ke-20 memiliki kemampuan generalisasi yang baik, khususnya dalam membedakan kelas mayoritas, sedangkan untuk kelas minoritas seperti Kufi, perlu dilakukan pendekatan tambahan seperti augmentasi data atau penyesuaian strategi pelatihan agar performa klasifikasinya dapat meningkat.

Tabel 2. Hasil Evaluasi Model Pada Epoch 20

	Precision	Recall	F1-Score
Diwani	0,98	1,00	0,99
Kufi	1,00	0,71	0,83
Naskhi	0,93	1,00	0,96
Tsulust	0,99	0,97	0,98
Accuracy			0,98
Macro Avg	0,98	0,92	0,94
Weighted Avg	0,98	0,98	0,97

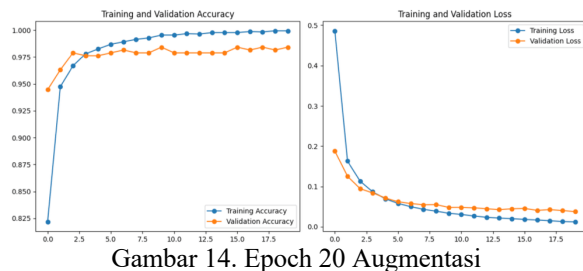
Pada Tabel 2 adalah hasil evaluasi metrik klasifikasi pada epoch ke-20, model menunjukkan performa yang sangat baik secara keseluruhan dengan akurasi mencapai 98%. Jika dilihat lebih lanjut berdasarkan masing-masing kelas, model memiliki nilai *precision*, *recall*, dan *F1-score* yang tinggi untuk hampir semua kelas. Pada kelas Diwani, nilai *precision* mencapai 0,98 dan *recall* 1,00, menghasilkan *F1-score* sebesar 0,99, yang mengindikasikan bahwa hampir seluruh prediksi untuk kelas Diwani benar dan tidak ada sampel yang

terlewat. Kelas Kufi, meskipun memiliki *precision* sempurna sebesar 1,00 (semua prediksi Kufi benar), namun *recall* hanya sebesar 0,71, sehingga *F1-score* menjadi 0,83. Hal ini menguatkan temuan pada *confusion matrix* sebelumnya bahwa model sering gagal mengenali sebagian sampel Kufi sebagai kelas yang benar, yang kemungkinan besar disebabkan oleh jumlah data yang tidak seimbang. Untuk kelas Naskhi, nilai *precision* sebesar 0,93 dan *recall* 1,00 menghasilkan *F1-score* yang tinggi sebesar 0,96, yang menandakan keseimbangan yang sangat baik antara kemampuan model dalam mengenali dan mengklasifikasikan tulisan Naskhi. Sementara itu, kelas Tsulust memiliki *precision* 0,99 dan *recall* 0,97, dengan *F1-score* sebesar 0,98, mencerminkan kinerja yang sangat stabil dan andal. Secara agregat, nilai *macro average* untuk *precision*, *recall*, dan *F1-score* masing-masing adalah 0,98, 0,92, dan 0,94, sedangkan *weighted average* masing-masing adalah 0,98, 0,98, dan 0,97. Hal ini menunjukkan bahwa meskipun ada ketidakseimbangan data antar kelas, model tetap mampu mempertahankan performa yang tinggi dan seimbang secara keseluruhan.

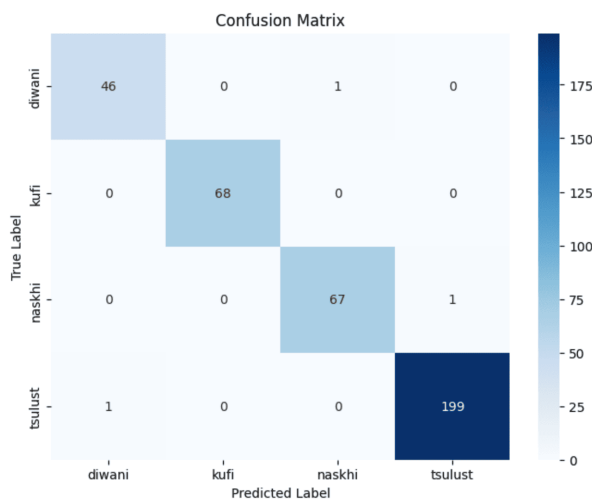
3.2. Hasil Pelatihan dengan Augmentasi

Mengingat performa klasifikasi pada kelas Kufi cenderung lebih rendah dibandingkan kelas lainnya, khususnya pada metrik *recall* yang hanya mencapai 0,71, maka dilakukan langkah lanjutan berupa augmentasi data secara khusus untuk kelas tersebut. Tujuan dari augmentasi ini adalah untuk meningkatkan jumlah variasi sampel tulisan Kufi dalam data pelatihan, Teknik augmentasi yang digunakan mencakup rotasi dalam rentang ± 10 derajat dari sumbu tengah, horizontal 5% dari dimensi gambar, zooming diterapkan secara acak antara 0.9x hingga 1.1x, yang berarti gambar dapat diperkecil 10% atau diperbesar 10%, dan perubahan kontras warna gambar. Setelah proses augmentasi diterapkan, jumlah data kelas Kufi meningkat secara signifikan, sehingga diharapkan distribusi data antar kelas menjadi lebih seimbang. Evaluasi ulang dilakukan untuk mengetahui apakah

penambahan variasi data ini dapat meningkatkan performa klasifikasi pada kelas Kufi tanpa menurunkan akurasi pada kelas lain. Hasil evaluasi setelah augmentasi akan dibandingkan dengan model sebelumnya untuk menilai efektivitas pendekatan ini.



Gambar 14. Epoch 20 Augmentasi



Gambar 15. Confusion Matrix Epoch 20 Augmentasi

Setelah dilakukan augmentasi model memperoleh *accuracy* 0.9990 hal ini mengalami sedikit penurunan sebelum dilakukan augmentasi pada dataset kufi yang mencapai 0.9993. Namun, meskipun terjadi penurunan kecil pada *training accuracy*, *validation accuracy* meningkat dari 0.9811 menjadi 0.9842, dan yang paling signifikan adalah peningkatan *testing accuracy* dari 97.52% menjadi 99.22%. Hal ini menunjukkan bahwa augmentasi pada dataset Kufi memberikan dampak positif terhadap performa model, karena membuat model lebih robust, mampu melakukan generalisasi lebih baik, serta mengurangi risiko overfitting dibandingkan sebelum dilakukan augmentasi. Pada dataset kelas Kufi, terlihat adanya peningkatan performa model secara

keseluruhan, khususnya pada kemampuan mengenali tulisan Kufi yang sebelumnya menjadi tantangan. Hal ini dapat dilihat pada *confusion matrix* terbaru pada Gambar 15, di mana jumlah sampel Kufi yang berhasil diklasifikasikan dengan benar meningkat secara signifikan. Model kini mampu mengenali 68 sampel Kufi dengan benar, tanpa terjadi kesalahan klasifikasi ke kelas lain. Ini menunjukkan bahwa augmentasi berhasil membantu model dalam memahami pola dan karakteristik khas dari tulisan Kufi.

Peningkatan ini juga tercermin pada metrik evaluasi pada tabel 3. F1-Score kelas Kufi mengalami lonjakan menjadi 1.00, yang berarti model tidak hanya akurat dalam memprediksi kelas Kufi (*precision*), tetapi juga sensitif terhadap keberadaannya (*recall*). Sementara itu, performa pada kelas lain seperti Tsulust, Naskhi, dan Diwani tetap stabil dan tinggi, dengan F1-Score di atas 0.98, yang menandakan bahwa augmentasi pada satu kelas tidak menurunkan akurasi klasifikasi kelas lain.

Secara keseluruhan, pendekatan augmentasi data terbukti efektif untuk mengatasi ketidakseimbangan dataset, khususnya dalam meningkatkan representasi kelas minoritas. Hasil ini memperkuat argumen bahwa strategi augmentasi merupakan langkah penting dalam meningkatkan performa model *deep learning*, terutama saat menghadapi keterbatasan jumlah data pada kelas tertentu.

Tabel 3. Hasil Evaluasi Model Augmentasi Pada Epoch 20

	Precision	Recall	F1-Score
Diwani	0,98	0,98	0,98
Kufi	1,00	1,00	1,00
Naskhi	0,99	0,99	0,99
Tsulust	0,99	0,99	0,99
Accuracy			0,99
Macro Avg	0,99	0,99	0,99
Weighted Avg	0,99	0,99	0,99

3.3. Deployment

Model terbaik yang dipilih yaitu model dengan hasil pelatihan augmentasi, yang kemudian dikonversi menjadi file dengan ekstensi h5 agar dapat digunakan pada *platform* website dengan menggunakan bahasa python *framework flask*. Pada website yang dibangun, pengguna bisa memasukkan gambar yang akan diprediksi. Gambar 16 merupakan tampilan antarmuka website yang dibangun.



Gambar 16. Tampilan antarmuka website

Setelah pengguna sudah input sebuah gambar kemudian, model tersebut akan memberikan prediksi berdasarkan input gambar yang diberikan. Gambar 17 menampilkan website dengan model yang dibangun menggunakan arsitektur MobileNetV2 berhasil melakukan prediksi gambar jenis kaligrafi tsulust dengan benar dan tingkat kepercayaan sebesar 97.74%.



Gambar 17. Tampilan antarmuka setelah melakukan prediksi

3.4. Uji Test Gambar

Data uji eksternal untuk menguji kinerja model dalam mengklasifikasikan jenis tulisan kaligrafi Arab. Data ini tidak termasuk dalam dataset pelatihan dan validasi. Data ini terdiri dari sepuluh gambar kaligrafi dengan empat kelas tulisan: Naskhi, Tsulust, Diwani, dan Kufi. Hasil prediksi model dibandingkan dengan label kelas sebenarnya, seperti yang ditunjukkan pada Tabel 4 berikut.

Tabel 4. Hasil Data Uji Eksternal

No	File	Kelas	Prediksi	Kesesuaian
1	06.JPG	Diwani	Diwani	Ya
2	07.JPG	Diwani	Diwani	Ya
3	08.JPG	Kufi	Kufi	Ya
4	09.JPG	Kufi	Kufi	Ya
5	01.JPG	Naskhi	Naskhi	Ya
6	02.JPG	Naskhi	Tsulust	Tidak
7	03.JPG	Naskhi	Nakhi	Ya
8	04.JPG	Tsulust	Tsulust	Ya
9	05.JPG	Tsulust	Tsulust	Ya
10	10.JPG	Tsulust	Tsulust	Ya

Tabel 4 menunjukkan bahwa model berhasil mengklasifikasikan 9 dari 10 gambar dengan benar, tetapi satu gambar dengan kelas Naskhi yang diprediksi sebagai Tsulust mengalami kesalahan klasifikasi. Kesalahan ini terjadi karna citra uji Naskhi memiliki bentuk tulisan melingkar yang menyerupai karakteristik Tsulust, sedangkan secara umum tulisan Naskhi cenderung ditulis mendatar dan lebih sederhana. Kondisi ini menyebabkan model mengalami ambiguitas dalam membedakan kedua kelas yang memiliki kemiripan struktur visual. Hasilnya menunjukkan bahwa meskipun model dapat menemukan jenis tulisan kaligrafi dengan baik, masih ada kemungkinan kesalahan prediksi, terutama pada kelas dengan karakteristik bentuk huruf yang serupa. Secara keseluruhan, temuan ini menunjukkan bahwa arsitektur CNN

MobileNetV2 memiliki kemampuan untuk generalisasi data di luar himpunan pelatihan, tetapi masih ada keterbatasan yang perlu diperhatikan dalam penelitian lebih lanjut.

3.5. Pembahasan

Kelas kufi dalam dataset ini memiliki jumlah data yang paling sedikit dibandingkan kelas lainnya. yang berpotensi menyebabkan ketidakseimbangan data dan menurunkan performa model pada kelas tersebut. Untuk mengatasi permasalahan ini, dilakukan teknik augmentasi citra seperti rotasi, perubahan kecerahan dan kontras, zooming, horizontal 5% dari dimensi gambar, kemudian resize menjadi 224x224 yang bertujuan untuk menambahkan keragaman visual pada data kufi. Setelah augmentasi diterapkan, hasil evaluasi menunjukkan adanya peningkatan signifikan pada performa model terhadap kelas Kufi. Nilai *precision* mencapai 1.00, *recall* 0.98, dan *f1-score* meningkat menjadi 0.99, yang menunjukkan bahwa model mampu mengenali pola tulisan Kufi secara konsisten dan akurat.

Peningkatan performa ini tercermin pula dalam *confusion matrix*, di mana mayoritas prediksi terhadap citra Kufi berhasil diklasifikasikan secara tepat ke dalam kelas yang sesuai. Hal ini mendukung temuan sebelumnya bahwa augmentasi data dapat membantu mengurangi bias pada kelas minoritas dan meningkatkan kemampuan generalisasi model [22]. Selain itu, performa pada kelas lain seperti Diwani, Naskhi, dan Tsulust juga tetap stabil dengan *f1-score* di atas 0.95, yang menunjukkan bahwa augmentasi yang dilakukan tidak berdampak negatif pada pembelajaran model secara keseluruhan.

Hasil pengujian deployment menunjukkan bahwa model memiliki kinerja yang cukup baik dalam mengklasifikasikan jenis tulisan kaligrafi pada data uji di luar dataset pelatihan. Pada pengujian untuk jenis kaligrafi tsulust berhasil diprediksi dengan benar, dan memiliki tingkat akurasi sebesar 97.74%, pada hasil ini menunjukkan bahwa model tidak hanya berhasil belajar dari data pelatihan, tetapi juga

cukup andal saat digunakan oleh pengguna untuk memprediksi jenis kaligrafi.

4. Kesimpulan

Penelitian ini berhasil membangun dan mengimplementasikan model klasifikasi jenis tulisan kaligrafi Arab menggunakan arsitektur MobileNetV2 yang dioptimalkan melalui pendekatan CRISP-DM. Model yang dikembangkan menunjukkan performa terbaik pada epoch ke-20 dengan akurasi pengujian sebesar 97.52% dan keseimbangan metrik klasifikasi antar kelas yang baik. Melalui evaluasi per kelas, diketahui bahwa model dapat mengenali tulisan Tsulust, Naskhi, dan Diwani dengan tingkat akurasi tinggi, namun memerlukan augmentasi tambahan pada kelas Kufi. Setelah augmentasi diterapkan, performa model memiliki *accuracy* 0.9990 hal ini mengalami sedikit penurunan sebelum dilakukan augmentasi pada dataset kufi yang mencapai 0.9993. Namun, meskipun terjadi penurunan kecil pada *training accuracy*, *validation accuracy* meningkat dari 0.9811 menjadi 0.9842, dan yang paling signifikan adalah peningkatan *testing accuracy* dari 97.52% menjadi 99.22%, dengan F1-Score kelas Kufi mencapai 1.00 tanpa menurunkan performa pada kelas lain. Selain pengembangan model, penelitian ini juga berhasil mengintegrasikan model klasifikasi ke dalam aplikasi berbasis web, yang memungkinkan pengguna melakukan pengenalan jenis kaligrafi secara otomatis. Hal ini membuktikan bahwa model tidak hanya efektif secara teknis, tetapi juga aplikatif dalam konteks nyata sebagai alat bantu edukasi dan pelestarian budaya Islam.

5. Saran

Penelitian selanjutnya dapat dikembangkan dengan melakukan eksplorasi terhadap teknik augmentasi lanjutan seperti *generative augmentation* berbasis GAN, serta perbandingan performa dengan arsitektur model lain seperti EfficientNet atau ResNet. Selain itu, aplikasi yang telah dibangun dapat dikembangkan lebih lanjut dengan antarmuka yang lebih interaktif, fitur penyimpanan riwayat klasifikasi, serta integrasi basis data

tulisan kaligrafi. Penelitian lanjutan juga dapat melibatkan pengujian terhadap pengguna akhir untuk mengevaluasi usability dan akurasi prediksi dalam skenario dunia nyata. Upaya ini diharapkan dapat memperluas pemanfaatan teknologi dalam pelestarian dan pembelajaran seni kaligrafi Arab secara digital.

6. Daftar Pustaka

- [1] R. Al-Hmouz, “Deep learning autoencoder approach: Automatic recognition of artistic Arabic calligraphy types”.
- [2] O. V. Putra, A. Musthafa, M. Nur, and M. Rido, “Classification of Calligraphy Writing Types Using Convolutional Neural Network Method (CNN),” *Procedia Eng. Life Sci.*, vol. 2, Nov. 2021, doi: 10.21070/pels.v2i0.1136.
- [3] M. Aldayel and K. Saud University, “Khatti: An Interactive Mobile Application for Training of Arabic Calligraphy,” presented at the Intelligent Human Systems Integration (IHSI 2024) Integrating People and Intelligent Systems, 2024. doi: 10.54941/ahfe1004536.
- [4] S. AlMuhaideb *et al.*, “Dhad—A Children’s Handwritten Arabic Characters Dataset for Automated Recognition,” *Appl. Sci.*, vol. 14, no. 6, p. 2332, Mar. 2024, doi: 10.3390/app14062332.
- [5] M. Khayyat and L. Elrefaei, “A Deep Learning Based Prediction of Arabic Manuscripts Handwriting Style,” *Int. Arab J. Inf. Technol.*, vol. 17, no. 5, pp. 702–712, Sept. 2020, doi: 10.34028/iajit/17/5/3.
- [6] A. A. Ahmad, “KAJIAN SENI KALIGRAFI ISLAM PADA MASJID AGUNG DARUSSALAM WATANSOPPENG”.
- [7] C. Schröer, F. Kruse, and J. M. Gómez, “A Systematic Literature Review on Applying CRISP-DM Process Model,” *Procedia Comput. Sci.*, vol. 181, pp. 526–534, 2021, doi: 10.1016/j.procs.2021.01.199.
- [8] R. B. Setiawan and N. Lukman, “Attendance System Face Recognition using Convolutional Neural Network (CNN),” *CoreID J.*, vol. 1, no. 3, Mar. 2024, doi: 10.60005/coreid.v1i3.16.
- [9] A. Ismail, Z. Kamel, and R. Mahmoud, “HICMA: The Handwriting Identification for Calligraphy and Manuscripts in Arabic Dataset,” in *Proceedings of ArabicNLP 2023*, Singapore (Hybrid): Association for Computational Linguistics, 2023, pp. 24–32. doi: 10.18653/v1/2023.arabicnlp-1.3.
- [10] F. Abdusyukur, “PENERAPAN ALGORITMA SUPPORT VECTOR MACHINE (SVM) UNTUK KLASIFIKASI PENCEMARAN NAMA BAIK DI MEDIA SOSIAL TWITTER,” *Komputa J. Ilm. Komput. Dan Inform.*, vol. 12, no. 1, pp. 73–82, May 2023, doi: 10.34010/komputa.v12i1.9418.
- [11] S. Ramdani and A. Rahmatulloh, “Implementasi Mobilenet untuk Klasifikasi Gambar dan Deteksi Emosi Menggunakan KERAS,” *J. Sist. Dan Teknol. Inf. JustIN*, vol. 12, no. 2, p. 259, Apr. 2024, doi: 10.26418/justin.v12i2.73389.
- [12] S. F. D. Wardhana and A. Nugroho, “Perbandingan Arsitektur MobileNetV2 dan MobileNetV3 Dalam Klasifikasi Jenis Jeruk,” *J. Ilmu Komput. Dan Bisnis*, vol. 16, no. 1, pp. 25–34, May 2025, doi: 10.47927/jikb.v16i1.916.
- [13] M. N. Baihaqi, A. Sugiharto, and H. Tantyoko, “Classification of Real and Fake Images Using Error Level Analysis Technique and MobileNetV2 Architecture,” *J. Masy. Inform.*, vol. 16, no. 1, pp. 54–68, May 2025, doi: 10.14710/jmasif.16.1.73283.
- [14] J. R. Hidayat, “Implementasi Klasifikasi Citra Berbasis Tensorflow Untuk Mendeteksi Penyakit Tanaman Pada Aplikasi Agroskan,” vol. 15, no. 1.
- [15] S. Muhartini, A. Sunyoto, and A. H. Muhammad, “Implementasi Metode Deep Learning CNN Dalam Klasifikasi Tajong (Sarung) Samarinda,” *J. SENOPATI*

- Sustain. Ergon. Optim. Appl. Ind. Eng.*, vol. 6, no. 1, pp. 28–41, Oct. 2024, doi: 10.31284/j.senopati.2024.v6i1.6573.
- [16] M. Akay *et al.*, “Deep Learning Classification of Systemic Sclerosis Skin Using the MobileNetV2 Model,” *IEEE Open J. Eng. Med. Biol.*, vol. 2, pp. 104–110, 2021, doi: 10.1109/OJEMB.2021.3066097.
- [17] A. R. Dani and I. Handayani, “Klasifikasi Motif Batik Yogyakarta Menggunakan Metode GLCM dan CNN,” *J. Teknol. Terpadu*, vol. 10, no. 2, pp. 142–156, Dec. 2024, doi: 10.54914/jtt.v10i2.1451.
- [18] Y. Y. Pane and J. J. Sihombing, “Klasifikasi Jenis Burung menggunakan Metode Transfer Learning,” *J. Teknol. Terpadu*, vol. 9, no. 2, pp. 89–94, Dec. 2023, doi: 10.54914/jtt.v9i2.744.
- [19] M. G. Somoal and A. R. Dzikrillah, “Komparasi MobileNETV2 dengan Kustomisasi Transfer Learning dan Hyperparameter untuk Identifikasi Tumor Otak,” *J. Teknol. Inf. Dan Ilmu Komput.*, vol. 12, no. 1, pp. 229–240, Feb. 2025, doi: 10.25126/jtiik.20251219582.
- [20] A. Afriani, H. Sujaini, and N. Candraningrum, “Analisis Perbandingan Metode Pengklasifikasi Gambar Jenis Tulisan Kaligrafi Arab,” *J. Edukasi Dan Penelit. Inform. JEPIN*, vol. 10, no. 1, p. 68, Apr. 2024, doi: 10.26418/jp.v10i1.72863.
- [21] M. Fadli and R. A. Saputra, “KLASIFIKASI DAN EVALUASI PERFORMA MODEL RANDOM FOREST UNTUK PREDIKSI STROKE,” vol. 12, no. 02, 2023.
- [22] I. Badruddin, T. R. Nugrahanto, O. D. Pangesti, and T. Agustin, “PENDEKATAN PENGURANGAN OVERFITTING PADA MOBILENET UNTUK KLASIFIKASI CITRA SAMPAH,” 2024.