

Analisis Peningkatan Kinerja Model InceptionResNetV2 dalam Identifikasi Wajah pada *Dataset In-the-Wild* melalui Optimasi *Hyperparameter*

Ferry Febiansah¹, Adhi Kusnadi², Aswan Supriyadi Sunge³

^{1,2,3} Magister Informatika, Universitas Nusa Putra.

Jl. Raya Cibolang No.21 Cisaat Sukabumi 43152, Indonesia

ferry.febiansah_cs24@nusaputra.ac.id

Abstrak

Identifikasi wajah yang didukung oleh *deep learning* memegang peranan krusial dalam berbagai aplikasi, termasuk sistem keamanan, autentikasi, serta verifikasi identitas. Akan tetapi, performa model pada *dataset in-the-wild* masih terkendala oleh variasi kondisi pencahayaan, perbedaan pose, ekspresi wajah yang beragam, dan kerumitan latar belakang. Penelitian ini bertujuan menganalisis peningkatan performa model InceptionResNetV2 melalui optimasi *hyperparameter* pada tugas identifikasi wajah, menggunakan *dataset primer* yang terdiri dari 5.180 citra wajah dari 153 individu dalam kondisi *in-the-wild*. Model ini dibangun menggunakan pendekatan *transfer learning* dengan memanfaatkan bobot awal dari *ImageNet* dan dilakukan *fine-tuning* pada 50 lapisan terakhir. Proses optimasi *hyperparameter* dilakukan dengan metode *grid search*, menggunakan kombinasi *learning rate* 0.001 dan 0.0005, serta *batch size* 8 dan 16. Evaluasi performa model dilakukan dengan mengukur metrik *accuracy*, *precision*, *recall*, dan *f1-score*, yang didukung oleh analisis *confusion matrix* dan visualisasi kasus prediksi. Hasil penelitian memperlihatkan bahwa model yang telah dioptimasi berhasil meningkatkan akurasi dari 0.8830 menjadi 0.9148, dengan peningkatan yang konsisten pada *precision* (dari 0.9063 menjadi 0.9249), *recall* (dari 0.9016 menjadi 0.9302), dan *f1-score* (dari 0.8839 menjadi 0.9139). Analisis *confusion matrix* menunjukkan peningkatan jumlah prediksi yang benar di berbagai kelas, sementara visualisasi *improvement case* membuktikan kemampuan model dalam memperbaiki kesalahan prediksi pada kondisi yang kompleks. Uji statistik *McNemar* menghasilkan nilai *p* sebesar 0.000639 (< 0.05), yang mengindikasikan bahwa peningkatan performa ini signifikan secara statistik. Dengan demikian, optimasi *hyperparameter* terbukti efektif dalam meningkatkan kinerja model pada *dataset in-the-wild*.

Kata kunci: identifikasi wajah, *deep learning*, InceptionResNetV2, *hyperparameter tuning*, *transfer learning*

Abstract

Deep learning-powered facial recognition is essential across various applications, including security systems, authentication processes, and identity verification. However, model performance on in-the-wild datasets faces challenges due to variations in lighting conditions, pose differences, diverse facial expressions, and background complexity. This study aims to analyze the performance improvement of the InceptionResNetV2 model through hyperparameter optimization for the face identification task, using a primary dataset comprising 5,180 face images of 153 individuals captured in in-the-wild conditions. The model employs transfer learning, utilizing initial weights from ImageNet, with fine-tuning applied to the final 50 layers. Hyperparameter optimization uses the grid search method, combining learning rates of 0.001 and 0.0005, and batch sizes of 8 and 16. Model performance is evaluated using accuracy, precision, recall, and F1-score metrics, supported by confusion matrix analysis and prediction case visualizations. Results show that the optimized model improved accuracy from 0.8830 to 0.9148, with consistent gains in precision (0.9063 to 0.9249), recall (0.9016 to 0.9302), and F1-score (0.8839 to 0.9139). Confusion matrix analysis indicates an increase in correct predictions across different classes, while improvement case visualizations demonstrate the model's ability to correct prediction errors under complex conditions. McNemar's statistical test yields a p-value of 0.000639 (< 0.05), indicating that the performance improvement is statistically significant. Therefore, hyperparameter optimization is effective in enhancing model performance on in-the-wild datasets.

Keywords: face recognition, *deep learning*, InceptionResNetV2, *hyperparameter optimization*, *transfer learning*

I. PENDAHULUAN

Perkembangan pesat teknologi *deep learning* dalam beberapa tahun terakhir telah memberikan dampak yang besar pada bidang pengolahan citra, terutama dalam sistem identifikasi wajah yang banyak digunakan dalam aplikasi keamanan dan validasi identitas [1]. Progres pada arsitektur *convolutional neural network* (CNN) dan model modern berbasis *deep learning* memungkinkan perolehan fitur wajah yang lebih presisi dan tangguh terhadap perbedaan data [2]. Akan tetapi, performa model pada *dataset in-the-wild* masih menemui kendala yang disebabkan oleh variasi pencahayaan, sudut pengambilan gambar, mimik wajah, serta kompleksitas latar belakang [3].

Model dengan arsitektur modern seperti InceptionResNetV2 tetap menjadi salah satu metode yang efektif dalam pengenalan wajah karena kemampuannya mengintegrasikan pembelajaran *residual* dan ekstraksi fitur multi-skala [4]. Meski begitu, kinerja model *deep learning* sangat bergantung pada konfigurasi *hyperparameter* yang diterapkan selama proses pelatihan, termasuk *learning rate* dan ukuran *batch* yang memengaruhi kestabilan dan konvergensi model [5]. Pemilihan *hyperparameter* yang kurang tepat dapat mengakibatkan penurunan performa yang disebabkan oleh *overfitting* atau ketidakmampuan model mencapai konvergensi yang optimal [6].

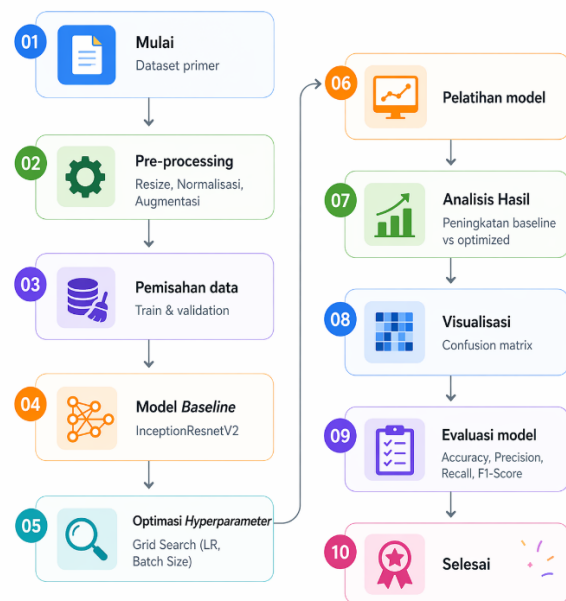
Studi terbaru menunjukkan bahwa optimasi *hyperparameter* menjadi aspek krusial dalam meningkatkan performa model *deep learning*, khususnya pada *dataset* dengan kompleksitas tinggi seperti *in-the-wild* [7]. Namun, sebagian besar studi hanya menitikberatkan pada peningkatan nilai metrik evaluasi tanpa melakukan analisis yang lebih mendalam terhadap perubahan performa model pada level kelas maupun kasus prediksi individual [8]. Oleh karena itu, diperlukan pendekatan analisis yang lebih menyeluruh untuk memahami dampak optimasi *hyperparameter* secara empiris.

Bertolak dari permasalahan tersebut, riset ini berfokus pada analisis peningkatan kinerja model InceptionResNetV2 dalam tugas identifikasi wajah pada *dataset in-the-wild* melalui optimasi *hyperparameter*. Evaluasi dilakukan tidak hanya secara kuantitatif menggunakan metrik *accuracy*, *precision*, *recall*, dan *f1-score*, tetapi juga melalui analisis *confusion matrix* serta visualisasi kasus prediksi untuk memberikan pemahaman yang lebih mendalam tentang peningkatan performa model.

II. METODE PENELITIAN

Penelitian ini berfokus pada analisis peningkatan performa model *deep learning* dalam mengenali wajah pada *dataset in-the-wild* melalui optimasi *hyperparameter*. Rancangan metodologi penelitian dilakukan secara sistematis, meliputi beberapa tahap yaitu pengumpulan *dataset*, pra-pemrosesan data, perancangan model, optimasi *hyperparameter*, proses pelatihan model, serta evaluasi dan analisis terhadap hasil yang diperoleh.

Diagram alur penelitian yang diimplementasikan divisualisasikan dalam Gambar 1, yang merepresentasikan rangkaian tahapan proses secara komprehensif, mulai dari fase pengumpulan data hingga tahap analisis hasil akhir.



Gambar 1. Tahapan penelitian

A. Dataset

Data yang dipakai dalam riset ini adalah *dataset primer*, yang dikumpulkan sendiri oleh penulis. Himpunan data ini terdiri dari 5.180 gambar wajah yang mewakili 153 orang. Semua gambar didapatkan dalam kondisi *in-the-wild*, termasuk variasi pencahayaan, perspektif, ekspresi muka, dan latar belakang yang berbeda-beda [9]. Karakteristik ini mencerminkan keadaan lingkungan yang sebenarnya, sehingga mengharuskan model mempunyai kemampuan generalisasi yang baik dalam melakukan pengenalan wajah [10].

Pengambilan gambar dilakukan dengan kamera digital Nikon D7000 beresolusi 16,2 megapiksel yang dikombinasikan dengan lensa telefoto 80–200 mm f/2.8. Penggunaan lensa telefoto dimaksudkan untuk menghasilkan fokus yang tajam pada wajah subjek

dan menghasilkan efek buram pada latar belakang, sehingga fitur wajah menjadi lebih dominan. Kamera dipasang pada tripod untuk menjaga kestabilan posisi serta konsistensi sudut pengambilan gambar.

B. Pre-processing data

Peningkatan kualitas dan standardisasi data sebelum digunakan dalam proses pembelajaran dicapai melalui serangkaian tahapan prapemrosesan. Selain itu, diterapkan teknik augmentasi data pada set data pelatihan, meliputi rotasi dan *horizontal flipping* untuk memperkaya variasi data dan mengurangi risiko *overfitting* [11].

C. Pengembangan model

Dalam riset ini, arsitektur InceptionResNetV2 diimplementasikan dengan strategi *transfer learning* [4]. Bobot yang telah dilatih sebelumnya pada *ImageNet* dimanfaatkan, diikuti dengan *fine-tuning* pada 50 lapisan terakhir guna mengoptimalkan model untuk *dataset* yang digunakan. Pemanfaatan *transfer learning* ditunjukkan dapat mendongkrak kinerja model, khususnya pada *dataset* berukuran kecil, sekaligus mempercepat laju konvergensi.

D. Optimasi hyperparameter

Optimasi *hyperparameter* dilakukan dengan metode *grid search*, menggabungkan *learning rate* 0.001 dan 0.0005 serta *batch size* 8 dan 16 [12], [13]. *Hyperparameter* ini sangat memengaruhi stabilitas pelatihan dan kinerja model. Setiap kombinasi dipakai untuk melatih model dan dievaluasi berdasarkan kinerja pada data validasi. Untuk meningkatkan stabilitas pelatihan, dipakai *callback* seperti *EarlyStopping*, *ReduceLROnPlateau*, dan *ModelCheckpoint* [14].

E. Pelatihan model

Pelatihan dilakukan memakai data pelatihan dengan *parameter* yang ditentukan dalam tahap optimasi. Model dilatih sampai konvergensi tercapai, menghasilkan model *baseline* dan model hasil optimasi untuk perbandingan kinerja.

F. Evaluasi model

Kinerja model dievaluasi memakai metrik *accuracy*, *precision*, *recall*, dan *f1-score* untuk mengukur kinerja klasifikasi secara komprehensif [15]. Penggunaan multi-metrik diperlukan untuk memberikan gambaran kinerja yang lebih menyeluruh pada masalah multi-kelas. Secara matematis, metrik evaluasi dirumuskan sebagai berikut:

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

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

Dalam evaluasi, TP, TN, FP, dan FN masing-masing mewakili *True Positive*, *True Negative*, *False Positive*, dan *False Negative*.

Confusion matrix juga dimanfaatkan untuk pemeriksaan kinerja model yang lebih mendalam pada tiap-tiap kelas. Perbandingan visual dari hasil prediksi pun dijalankan untuk mengetahui perbedaan kinerja antara model dasar dan model yang telah dioptimalkan.

Untuk memvalidasi signifikansi perbedaan kinerja dari kedua model tersebut, diterapkan uji statistik *McNemar* [16], yang dirumuskan sebagai berikut:

$$\chi^2 = \frac{(b-c)^2}{b+c} \quad (5)$$

Dalam hal ini, (b) melambangkan frekuensi kesalahan klasifikasi oleh model *baseline* yang berhasil dikoreksi oleh model yang dioptimalkan, sementara (c) melambangkan frekuensi klasifikasi yang semula benar oleh model *baseline* namun kemudian salah setelah dioptimalkan.

G. Analisis hasil

Analisis dilakukan berdasarkan metrik evaluasi dan visualisasi hasil prediksi.

III. HASIL DAN PEMBAHASAN

A. Deskripsi dataset

Dataset yang dipakai dalam penelitian ini merupakan *dataset primer* berkarakter *in-the-wild* yang dikumpulkan secara mandiri oleh peneliti. Totalnya terdapat 5.180 citra wajah dari 153 individu, dengan variasi tinggi pada pencahayaan, pose, ekspresi, serta kondisi lingkungan sekitar. Keberagaman tersebut menghasilkan distribusi data yang tidak merata, sehingga menambah tingkat kompleksitas dalam proses identifikasi wajah.

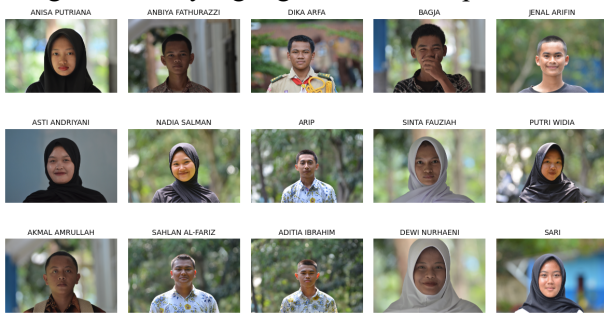
Dataset selanjutnya dibagi dengan skema *hold-out*: 80% untuk pelatihan dan 20% untuk validasi. Pembagian *dataset* dilakukan secara stratifikasi pada level citra untuk setiap kelas, menjaga proporsi distribusi data antar kelas pada set pelatihan dan validasi dalam skenario *closed-set face identification*. Dari pembagian tersebut, 4.144 citra dipakai sebagai data pelatihan, sementara 1.036 citra digunakan sebagai data validasi.

Kerumitan *dataset* menimbulkan tantangan khusus bagi model, khususnya dalam membedakan individu yang memiliki karakteristik visual serupa. Pada kondisi tertentu, perubahan pencahayaan atau

sudut pengambilan gambar dapat memengaruhi representasi fitur wajah yang dipelajari model, sehingga berpotensi meningkatkan tingkat kesalahan klasifikasi.

Untuk menggambarkan karakteristik data, contoh citra yang dipilih secara acak ditampilkan pada Gambar 2. Sampel tersebut menonjolkan variasi signifikan pada pencahayaan, pose, dan latar belakang. Variasi ini menjadi faktor utama yang memengaruhi performa model dalam proses identifikasi wajah.

Selain itu, pada visualisasi tersebut setiap kelas hanya diwakili oleh satu citra. Pendekatan ini menghindari pengulangan visual dan menjaga agar representasi distribusi sampel tetap seimbang terhadap keseluruhan *dataset*, sehingga memberikan ilustrasi yang representatif mengenai tingkat keragaman data yang digunakan dalam penelitian ini.



Gambar 2. Contoh *dataset*

B. Pengembangan model & hyperparameter

Dalam penelitian ini, pengembangan model melibatkan perbandingan kinerja antara model *baseline* dan model yang telah dioptimalkan hyperparameternya dengan arsitektur InceptionResNetV2. Model *baseline*, yang dilatih menggunakan pengaturan awal tanpa optimasi, berfungsi sebagai tolok ukur untuk menilai peningkatan kinerja.

Proses optimasi *hyperparameter* diimplementasikan melalui metode *grid search*, yang menguji berbagai kombinasi nilai *learning rate* dan *batch size*. Setiap kombinasi ini diterapkan dalam pelatihan, dan evaluasi dilakukan berdasarkan kinerja pada set data validasi.

Hasil eksperimen yang diperoleh dari proses *grid search* terangkum dalam Tabel 1.

Tabel 1. Hasil *baseline* & *grid search hyperparameter*

No	<i>Learning rate</i>	<i>Batch size</i>	<i>Val Accuracy</i>	<i>Val loss</i>
1	<i>Baseline</i>	<i>default</i>	0.8923	0.4505
2	0.0005	16	0.9195	0.4942

3	0.001	16	0.9177	0.4298
4	0.001	8	0.8896	0.5778
5	0.0005	8	0.8824	0.5835

Berdasarkan data pada Tabel 1, model *baseline* mencatatkan akurasi validasi 0.8923 dan nilai *loss* 0.4505. Optimasi *hyperparameter* menghasilkan sejumlah konfigurasi yang melampaui performa *baseline*.

Akurasi tertinggi dicapai dengan *learning rate* 0.0005 dan *batch size* 16, dengan akurasi validasi mencapai 0.9195. Di sisi lain, kombinasi *learning rate* 0.001 dan ukuran *batch* 16 menghasilkan nilai *loss* terendah, yaitu 0.4298, menandakan proses pelatihan yang lebih konsisten.

Temuan ini menegaskan bahwa optimasi *hyperparameter* memiliki pengaruh besar terhadap kinerja model. Kenaikan akurasi sekitar 2.7% dibandingkan dengan *baseline* menunjukkan bahwa pengaturan parameter yang tepat dapat meningkatkan kemampuan model dalam melakukan generalisasi pada *dataset* yang belum pernah dilihat sebelumnya (*in-the-wild*).

Lebih lanjut, ukuran *batch* 16 secara konsisten memberikan hasil yang lebih unggul dibandingkan ukuran *batch* 8. Hal ini mengindikasikan bahwa kestabilan *gradient* selama pelatihan berpengaruh positif terhadap mutu pembelajaran model.

C. Hasil pelatihan

Hasil pelatihan memperlihatkan perbedaan signifikan antara model dasar dan model yang *hyperparameter*-nya telah dioptimalkan. Model dasar mencapai akurasi validasi 0.8923 dengan *loss* 0.4505, menandakan kemampuan mempelajari pola dasar, tetapi belum optimal untuk kompleksitas *dataset* alami.

Setelah optimasi *hyperparameter*, semua konfigurasi menunjukkan peningkatan performa dibandingkan model dasar. Konfigurasi terbaik berdasarkan akurasi, yaitu *learning rate* 0.0005 dan *batch size* 16, menghasilkan akurasi validasi 0.9195. Sementara itu, kombinasi *learning rate* 0.001 dan *batch size* 16 menghasilkan *loss* terendah, yaitu 0.4298, yang mengindikasikan pembelajaran yang lebih stabil.

Perbedaan ini mengindikasikan bahwa optimasi *hyperparameter* memengaruhi dinamika pelatihan model secara signifikan. Model yang dioptimalkan menunjukkan konvergensi yang lebih baik, dengan peningkatan akurasi lebih tinggi dan stabilitas pembelajaran yang lebih baik dibandingkan model

dasar. Selain itu, *batch size* 16 secara konsisten memberikan hasil yang lebih optimal dibandingkan *batch size* 8, menunjukkan pentingnya kestabilan gradien selama pelatihan dalam peningkatan performa model.

D. Evaluasi model

Perbandingan kinerja antara model dasar dan model yang telah dioptimalkan dilakukan melalui evaluasi menggunakan metrik-metrik seperti *accuracy*, *precision*, *recall*, dan *f1-score*. Rincian hasil evaluasi tersebut dapat dilihat pada Tabel 2.

Tabel 2. Perbandingan performa model *baseline* dan *optimized*

Model	Accuracy	Precision	Recall	F1-score
Baseline	0.8830	0.9063	0.9016	0.8839
Optimized	0.9148	0.9249	0.9302	0.9139

Berdasarkan data yang disajikan pada Tabel 2, model yang telah dioptimasi memperlihatkan kemajuan kinerja di semua *parameter* evaluasi jika dibandingkan dengan model *baseline*. Tingkat akurasi meningkat dari 0.8830 menjadi 0.9148, mengindikasikan peningkatan pada kapasitas model dalam melakukan klasifikasi secara menyeluruh.

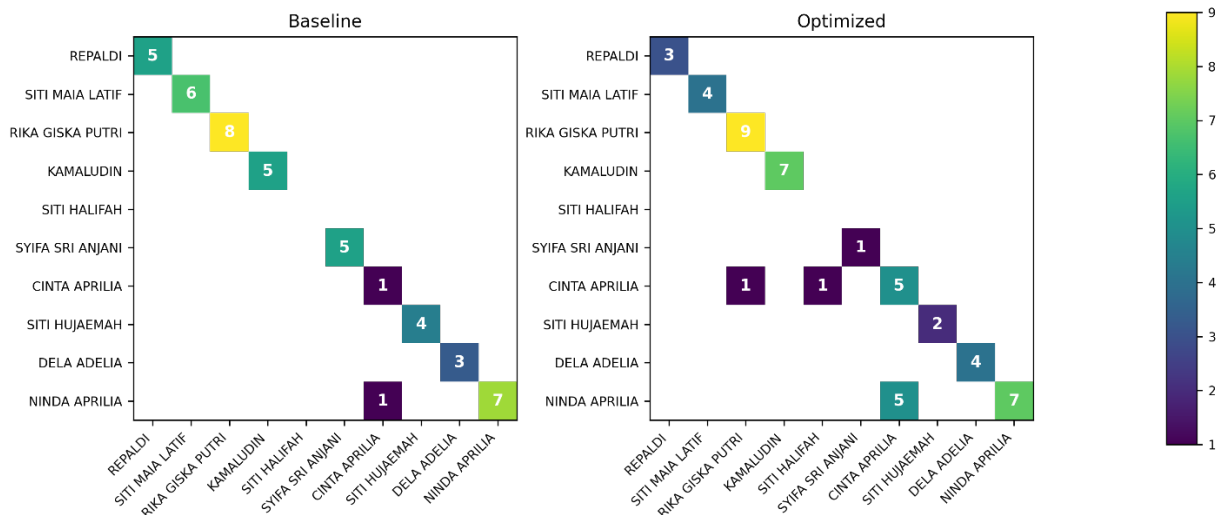
Kenaikan pada *precision* dan *recall* mengindikasikan bahwa model hasil optimasi lebih

ajek dalam mengenali kelas yang tepat, serta mampu meminimalisir kesalahan klasifikasi. Hal ini juga terefleksikan pada nilai *f1-score* yang meningkat menjadi 0.9139, yang memperlihatkan keseimbangan yang lebih baik antara *precision* dan *recall*.

E. Analisis hasil

Analisis hasil bertujuan untuk mengevaluasi secara mendalam pengaruh optimasi *hyperparameter* terhadap peningkatan performa model dalam tugas identifikasi wajah pada *dataset in-the-wild*. Pendekatan analisis tidak hanya mengandalkan metrik kuantitatif, tetapi juga diperkaya dengan visualisasi distribusi kesalahan, peningkatan per kelas, serta analisis kasus prediksi.

Perbandingan *confusion matrix* antara model *baseline* dan model hasil optimasi, seperti yang ditunjukkan pada Gambar 3, memperlihatkan peningkatan yang konsisten pada nilai diagonal utama yang mewakili jumlah prediksi benar pada setiap kelas. Pada model *baseline*, masih ditemukan sejumlah kesalahan klasifikasi yang tersebar pada beberapa kelas, yang mengindikasikan adanya keterbatasan model dalam membedakan individu dengan karakteristik visual yang serupa. Sebaliknya, pada model hasil optimasi, distribusi kesalahan menjadi lebih terkonsentrasi dan jumlah prediksi benar meningkat, yang menunjukkan peningkatan kemampuan diskriminatif model.



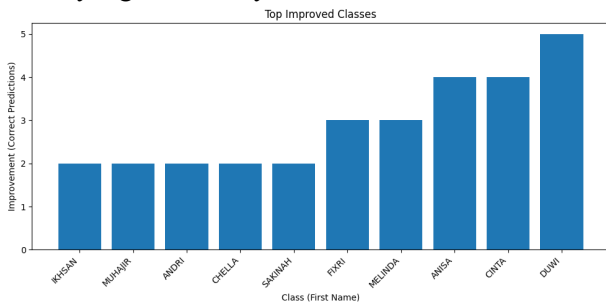
Gambar 3. Perbandingan *confusion matrix* model *baseline* dan model optimasi

Analisis lebih lanjut terhadap distribusi kesalahan mengungkapkan bahwa sebagian besar kesalahan pada model *baseline* terjadi pada citra dengan kondisi pencahayaan yang kurang optimal serta variasi pose wajah yang signifikan. Hal ini sesuai dengan karakteristik *dataset in-the-wild* yang memiliki tingkat kompleksitas tinggi. Setelah optimasi

hyperparameter dilakukan, model menunjukkan kemampuan yang lebih baik dalam menangani variasi tersebut, yang tercermin dari berkurangnya kesalahan klasifikasi pada kelas-kelas yang sebelumnya sulit dikenali.

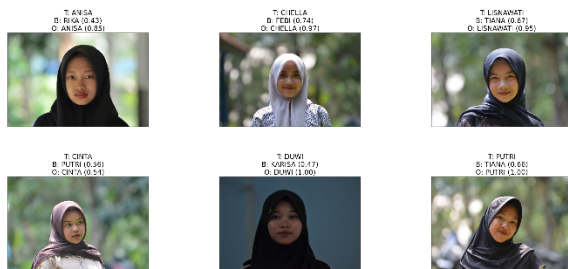
Visualisasi peningkatan per kelas, seperti yang ditunjukkan pada Gambar 4, memperlihatkan bahwa

beberapa kelas mengalami peningkatan jumlah prediksi benar secara signifikan, dengan peningkatan hingga lima prediksi benar pada kelas tertentu. Hal ini mengindikasikan bahwa optimasi *hyperparameter* tidak hanya meningkatkan performa secara global, tetapi juga memberikan dampak nyata pada tingkat kelas, khususnya pada kelas dengan karakteristik visual yang sebelumnya sulit dibedakan.



Gambar 4. Peningkatan jumlah prediksi benar

Selain itu, analisis kasus prediksi yang ditunjukkan pada Gambar 5 memberikan bukti empiris mengenai peningkatan performa model. Pada beberapa contoh citra, model *baseline* gagal mengklasifikasikan dengan benar, terutama pada kondisi pencahayaan rendah atau pose wajah yang tidak frontal. Akan tetapi, model hasil optimasi mampu mengoreksi kesalahan tersebut dan menghasilkan prediksi yang benar. Hal ini menunjukkan bahwa optimasi *hyperparameter* meningkatkan kemampuan model dalam mengekstraksi fitur wajah yang lebih stabil terhadap variasi visual.



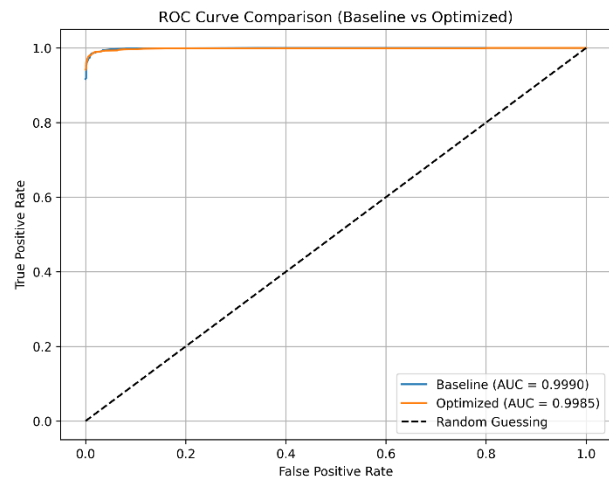
Gambar 5. Peningkatan performa model pada contoh citra

Untuk memperkuat hasil analisis, dilakukan uji statistik menggunakan *McNemar*. Hasil pengujian menunjukkan nilai *p-value* sebesar 0.000639, yang lebih kecil dari 0.05. Hal ini mengindikasikan bahwa perbedaan performa antara model *baseline* dan model hasil optimasi bersifat signifikan secara statistik dan bukan disebabkan oleh faktor kebetulan. Secara keseluruhan, hasil analisis menunjukkan bahwa optimasi *hyperparameter* memberikan dampak yang signifikan terhadap peningkatan performa model, baik dari segi akurasi, distribusi prediksi per kelas, maupun kemampuan model dalam menangani variasi

kondisi pada *dataset in-the-wild*. Meskipun demikian, masih terdapat beberapa kesalahan klasifikasi pada kelas dengan kemiripan visual tinggi, yang menunjukkan bahwa kompleksitas *dataset* tetap menjadi tantangan utama dalam pengembangan sistem identifikasi wajah.

Selain melalui metrik klasifikasi dan inspeksi visual, kinerja model turut dianalisis memakai kurva *Receiver Operating Characteristic (ROC)*, seperti yang tergambar pada Gambar 6. Kurva ROC berperan dalam menilai keefektifan model dalam membedakan antar kelas, dengan mempertimbangkan keseimbangan antara *True Positive Rate (TPR)* dan *False Positive Rate (FPR)*.

Berdasarkan analisis kurva *ROC*, kedua model menunjukkan kinerja yang sangat baik dengan nilai *AUC* yang mendekati 1. Model *baseline* memperoleh *AUC* sebesar 0.9990, sedangkan model hasil optimasi mencapai 0.9985. Hal ini mengindikasikan bahwa kedua model memiliki kemampuan diskriminatif yang sangat tinggi dalam membedakan antar kelas, meskipun perbedaan performanya relatif tidak signifikan secara praktis.



Gambar 6. Komparasi kurva *ROC* model *baseline* vs model optimasi

Akan tetapi, model hasil optimasi tetap menunjukkan performa yang lebih baik pada metrik utama seperti *accuracy*, *precision*, *recall*, dan *f1-score*. Hal ini mengisyaratkan bahwa optimasi *hyperparameter* memiliki dampak yang lebih besar pada peningkatan performa klasifikasi secara langsung, dibandingkan dengan kemampuan diskriminatif global yang diukur oleh *ROC*.

Dengan demikian, analisis *ROC* memperkuat kesimpulan bahwa kedua model memiliki kemampuan pemisahan kelas yang sangat baik, namun peningkatan performa pada model hasil optimasi lebih menonjol pada aspek klasifikasi praktis yang relevan dengan tugas identifikasi wajah.

IV. KESIMPULAN

Penelitian ini membuktikan bahwa optimasi *hyperparameter* dapat secara signifikan meningkatkan kinerja model InceptionResNetV2 pada tugas identifikasi wajah dengan menggunakan *dataset in-the-wild* yang bersifat kompleks. Berdasarkan rangkaian percobaan, kombinasi *learning rate* sebesar 0.0005 dan *batch size* 16 menghasilkan performa tertinggi, dengan akurasi validasi naik dari 0.8923 menjadi 0.9195. Pada tahap evaluasi akhir, model yang telah dioptimasi juga meningkatkan metrik lainnya: *accuracy* meningkat dari 0.8830 menjadi 0.9148, *precision* dari 0.9063 menjadi 0.9249, *recall* dari 0.9016 menjadi 0.9302, serta *f1-score* dari 0.8839 menjadi 0.9139.

Analisis *confusion matrix*, visualisasi peningkatan prediksi, serta uji statistik McNemar ($p\text{-value} = 0.000639 < 0.05$) menegaskan bahwa perbaikan tersebut bersifat signifikan secara statistik. Temuan ini menegaskan peran krusial optimasi *hyperparameter* dalam memperluas kemampuan generalisasi model *deep learning* untuk identifikasi wajah di lingkungan *in-the-wild*. Untuk penelitian selanjutnya, disarankan penekanan pada pengembangan arsitektur yang lebih adaptif serta eksplorasi teknik augmentasi data yang lebih kompleks.

UCAPAN TERIMA KASIH

Penulis menyampaikan rasa terima kasih yang sebesar-besarnya kepada semua pihak, baik individu maupun institusi, yang telah berperan serta dalam menuntaskan riset ini. Penghargaan khusus ditujukan kepada Universitas Nusa Putra atas penyediaan sarana, prasarana, dan suasana akademik yang mendukung kelancaran pelaksanaan penelitian ini.

Penulis juga memberikan apresiasi atas sokongan dan asistensi dari kolega yang telah berpartisipasi dalam tahapan pengumpulan dan analisis data, serta validasi model. Selain itu, penulis mengucapkan terima kasih kepada semua pihak yang telah menyumbangkan ide, petunjuk, dan dukungan, baik secara langsung maupun tidak langsung, selama proses penulisan penelitian ini.

Harapan penulis adalah agar temuan dari riset ini dapat memberikan sumbangsih yang berguna bagi kemajuan ilmu pengetahuan, khususnya di bidang pengolahan citra dan *deep learning*.

REFERENSI

- [1] S. Khan, M. Naseer, M. Hayat, S. W. Zamir, F. S. Khan, and M. Shah, "Transformers in Vision: A Survey," *ACM Comput. Surv.*, vol. 54, no. 10s, pp. 1–41, Jan. 2022, doi: 10.1145/3505244.
- [2] V. Tomar, N. Kumar, and A. R. Srivastava, "Single sample face recognition using deep learning: a survey," *Artif. Intell. Rev.*, vol. 56, no. S1, pp. 1063–1111, Oct. 2023, doi: 10.1007/s10462-023-10551-y.
- [3] M. Wang and W. Deng, "Deep face recognition: A survey," *Neurocomputing*, vol. 429, pp. 215–244, Mar. 2021, doi: 10.1016/j.neucom.2020.10.081.
- [4] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning," 2016. [Online]. Available: <https://arxiv.org/abs/1602.07261>
- [5] T. Yu and H. Zhu, "Hyper-Parameter Optimization: A Review of Algorithms and Applications," 2020. doi: 10.48550/arxiv.2003.05689.
- [6] C. Zhang, S. Bengio, M. Hardt, B. Recht, and O. Vinyals, "Understanding deep learning requires rethinking generalization," 2017. [Online]. Available: <https://arxiv.org/abs/1611.03530>
- [7] M. Feurer and F. Hutter, "Hyperparameter Optimization," 2019, pp. 3–33. doi: 10.1007/978-3-030-05318-5_1.
- [8] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," *BMC Genomics*, vol. 21, no. 1, p. 6, Dec. 2020, doi: 10.1186/s12864-019-6413-7.
- [9] G. Sari, S. B. Okcu, D. Demir, İ. Göçer, and C. Çiğla, "Face recognition in the wild: augmentation-based robustness against real-world degradations," in *Artificial Intelligence for Security and Defence Applications III*, Y. Yitzhaky, R. Prabhu, and H. J. Kuijff, Eds., SPIE, Oct. 2025, p. 20. doi: 10.1117/12.3069782.
- [10] Y. Huang, Z. Wu, J. Chen, and H. Xiang, "Privacy-Preserving Face Recognition Method Based on Randomization and Local Feature Learning," *J. Imaging*, vol. 10, no. 3, p. 59, Feb. 2024, doi: 10.3390/jimaging10030059.
- [11] Diniati Ruaika and Shofwatul Uyun, "Optimisation of Residual Network Using Data Augmentation and Ensemble Deep Learning for Butterfly Image Classification," *IJID (International J. Informatics Dev.)*, vol. 12, no. 2, pp. 350–361, Jan. 2024, doi: 10.14421/ijid.2023.4038.
- [12] Y. Jaber, P. Dharmasena, A. Nassif, and N.

- Nassif, "Hyperparameter Optimization of Neural Networks Using Grid Search for Predicting HVAC Heating Coil Performance," *Buildings*, vol. 15, no. 15, p. 2753, Aug. 2025, doi: 10.3390/buildings15152753.
- [13] N. H. Tiep *et al.*, "A New Hyperparameter Tuning Framework for Regression Tasks in Deep Neural Network: Combined-Sampling Algorithm to Search the Optimized Hyperparameters," *Mathematics*, vol. 12, no. 24, p. 3892, Dec. 2024, doi: 10.3390/math12243892.
- [14] L. Xu, X. Chen, and X. Yang, "Tourism image classification based on convolutional neural network SqueezeNet—Taking Slender West Lake as an example," *PLoS One*, vol. 19, no. 1, p. e0295439, Jan. 2024, doi: 10.1371/journal.pone.0295439.
- [15] O. Rainio, J. Teuho, and R. Klén, "Evaluation metrics and statistical tests for machine learning," *Sci. Rep.*, vol. 14, no. 1, p. 6086, Mar. 2024, doi: 10.1038/s41598-024-56706-x.
- [16] M. Martinović, K. Dokic, and D. Pudić, "Comparative Analysis of Machine Learning Models for Predicting Innovation Outcomes: An Applied AI Approach," *Appl. Sci.*, vol. 15, no. 7, 2025, doi: 10.3390/app15073636.