

Augment During Training (ADT) pada DenseNet-121: Klasifikasi Tingkat Keparahan Retinopati Diabetik pada Citra Fundus

Willy Cahyadi¹, Daniel Martomanggolo Wonohadidjojo^{1,*}

¹ Program Studi Informatika, Universitas Ciputra Surabaya, Indonesia

* Correspondence: daniel.m.w@ciputra.ac.id

Copyright: © 2025 by the authors

Received: 11 Oktober 2025 | Revised: 29 Oktober 2025 | Accepted: 24 November 2025 | Published: 7 Desember 2025

Abstrak

Retinopati diabetik (RD) merupakan komplikasi serius diabetes mellitus sekaligus penyebab utama kebutaan usia produktif. Diagnosis manual berbasis citra fundus bersifat subjektif dan kurang efisien untuk skrining massal, sedangkan diagnosis otomatis masih terkendala ketidakseimbangan kelas yang membatasi kinerja dan generalisasi model. Penelitian ini bertujuan mengembangkan model klasifikasi tingkat keparahan RD menggunakan arsitektur DenseNet-121 dengan mengusulkan solusi inovatif berupa strategi augmentasi dinamis *Augment During Training* (ADT). Strategi ini dievaluasi dan dibandingkan dengan tiga pendekatan lain: *Baseline*, *Oversampling* Augmentasi, dan SMOTE. Dataset yang digunakan adalah APTOS 2019 Blindness Detection (3.662 citra fundus dengan lima tingkat keparahan RD). Hasil penelitian menunjukkan bahwa pendekatan model dengan ADT mencapai akurasi 83,65% (95% CI: 81,53–85,77%), *F1-score* 83,12%, *precision* 83,80%, dan *recall* 83,65%, mengungguli tiga pendekatan lain dengan peningkatan 3,27%. Analisis per kelas menunjukkan bahwa ADT efektif dalam mengatasi ketidakseimbangan data, dengan performa tinggi pada kelas *No DR* (0,98) dan *Moderate* (0,77), namun masih perlu peningkatan pada kelas minor seperti *Severe* dan *Proliferative*. Penelitian ini menyediakan landasan komputasi yang kuat untuk pengembangan alat bantu deteksi dini RD di masa depan.

Kata kunci: augmentasi dinamis; *deep learning*; densenet-121; ketidakseimbangan kelas; retinopati diabetik

Abstract

Diabetic Retinopathy (DR) is a severe complication of diabetes mellitus and a leading cause of blindness among the working-age population. Manual diagnosis using fundus images is often subjective and inefficient for mass screening, while automated diagnosis faces challenges due to class imbalance, limiting model performance and generalization. This study aims to develop a DR severity classification model based on the DenseNet-121 architecture by introducing a dynamic augmentation strategy called *Augment During Training* (ADT). This strategy was evaluated and compared against three other approaches: *Baseline*, *Oversampling* Augmentation, and SMOTE. We used the APTOS 2019 Blindness Detection dataset, comprising 3,662 fundus images across five severity levels. The results showed that the model with ADT achieved an accuracy of 83.65% (95% CI: 81.53–85.77%), an *F1-score* of 83.12%, *precision* of 83.80%, and *recall* of 83.65%, outperforming the other three approaches by a margin of 3.27%. Class-wise analysis demonstrated that ADT effectively addressed data imbalance, yielding high performance in the *No DR* (0.98) and *Moderate* (0.77) classes, though improvements are still needed for minority classes such as *Severe* and *Proliferative*. This study provides a solid computational foundation for developing future early detection tools for Diabetic Retinopathy.

Keywords: adaptive augmentation; *deep learning*; densenet-121; class imbalance; diabetic retinopathy



PENDAHULUAN

Diabetes mellitus merupakan masalah kesehatan utama di Indonesia dan global, dengan prevalensi pada populasi usia 15 tahun ke atas mencapai 11,7% menurut Survei Kesehatan Indonesia (SKI) 2023 (Santika, 2024), dan diproyeksikan meningkat sebesar 75,1% dari 18,69 juta kasus pada tahun 2020 menjadi 40,7 juta pada tahun 2045 (Wahidin et al., 2024). Komplikasi serius dari diabetes adalah retinopati diabetik (RD), yang merupakan komplikasi mikrovaskular yang menjadi penyebab utama kebutaan, memengaruhi 1 dari 3 penderita diabetes di dunia (Benhamza et al., 2024). Studi epidemiologi terbaru di Bandung Raya melaporkan bahwa kurang lebih 1 dari 4 pasien diabetes menderita RD, dengan prevalensi 19,46% dan 7,68% di antaranya tergolong *Vision Threatening Diabetic Retinopathy* (VTDR) yang memerlukan penanganan segera (Shaniaputri et al., 2022). Dengan proyeksi peningkatan kasus diabetes, diperkirakan kasus RD akan meningkat dari 5.739.732 pada tahun 2020 menjadi 12.497.915 pada tahun 2045 (Wahidin et al., 2024).

RD diklasifikasikan dalam lima tingkat keparahan: *No DR*, *Mild*, *Moderate*, *Severe*, dan *Proliferative DR*, berdasarkan jenis dan distribusi lesi mikrovaskular seperti mikroaneurisma, perdarahan intraretinal, dan neovaskularisasi (Yang et al., 2022). Perbedaan halus antar-kategori, seperti antara *Mild* (hanya mikroaneurisma) dan *Moderate* (mencakup perdarahan *dot-blot* dan eksudat keras), menyulitkan model dalam ekstraksi fitur dan klasifikasi akurat (Vishwanath et al., 2024). Diagnosis RD konvensional bergantung pada interpretasi manual citra fundus oleh dokter mata, yang tidak efisien untuk skrining massal, terutama di daerah terpencil yang berisiko tidak terdeteksi hingga tahap lanjut (Prajnaparamita, 2025). Selain itu, tantangan utama diagnosis otomatis adalah *class imbalance*, di mana kelas minor seperti *Severe*, dan *Proliferative* kurang terwakili dalam dataset, membuat model kesulitan dalam membedakan fitur halus antar kategori, yang menjadi hambatan teknis utama dalam akurasi klasifikasi otomatis (Atwany et al., 2022).

Perkembangan *Convolutional Neural Networks* (CNN), khususnya DenseNet-121, telah diterapkan dalam deteksi RD dengan akurasi mencapai 75–83% pada citra fundus (Khan et al., 2023). Namun, pendekatan ini sering kali gagal mengatasi masalah ketidakseimbangan kelas dan generalisasi pada data uji, seperti yang terlihat dalam penelitian Gangwar & Ravi. (2021) yang menggunakan arsitektur Inception-ResNet-v2 dengan akurasi 82.18%. Augmentasi statis dan SMOTE, meskipun dapat meningkatkan performa, memiliki keterbatasan mendasar karena augmentasi dilakukan hanya sekali sebelum pelatihan, yang tidak adaptif terhadap perubahan distribusi kelas selama proses pembelajaran. SMOTE, khususnya, menghasilkan data sintetis sekali saja, yang mengurangi diversitas fitur yang dibutuhkan untuk meningkatkan generalisasi (Matharaarachchi et al., 2024). Oleh karena itu, diperlukan pengembangan teknik augmentasi yang lebih dinamis dan adaptif agar mampu mengatasi ketidakseimbangan kelas secara efektif sepanjang proses pembelajaran.

Penelitian klasifikasi retinopati diabetik dengan CNN (misalnya Inception-ResNet-v2 dan DenseNet-121) masih banyak bergantung pada augmentasi statis atau oversampling konvensional seperti SMOTE. Pendekatan ini terbatas karena tidak adaptif terhadap perubahan distribusi kelas selama pelatihan. Akibatnya, performa pada kelas minor (*Severe* dan *Proliferative DR*) menurun, generalisasi model terbatas, dan klasifikasi tahap keparahan tinggi kurang optimal karena tumpang tindih morfologis antar-lesi (Luo et al., 2022; Ullah et al., 2022).

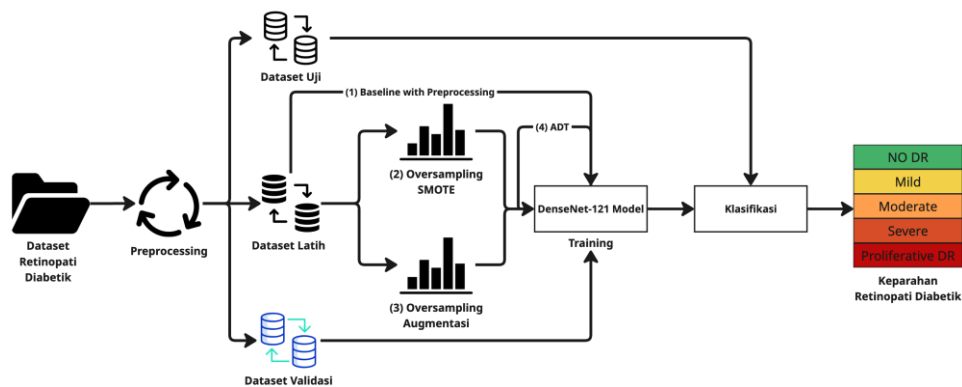
Sebagai solusi, penelitian ini membangun kerangka kerja diagnosis otomatis yang mengintegrasikan arsitektur DenseNet-121 dengan mekanisme inovatif *Augment During Training* (ADT). Solusi ini dirancang sebagai *pipeline* klasifikasi *end-to-end* yang mampu memproses citra fundus mentah, melakukan augmentasi variasi data secara *real-time* selama pelatihan, dan menghasilkan prediksi tingkat keparahan RD secara presisi tanpa menambah beban penyimpanan dataset. Keunggulan utama rancangan ini terletak pada kemampuan

adaptifnya dalam menyeimbangkan kelas secara dinamis untuk mengatasi keterbatasan augmentasi statis. Evaluasi kinerja model dilakukan melalui perbandingan empat skenario pendekatan, yaitu *Baseline* dengan *preprocessing*, *Oversampling* SMOTE, *Oversampling* Augmentasi Citra, dan ADT. Melalui *ablation study*, kami mengidentifikasi kontribusi spesifik dari setiap komponen solusi terhadap performa model secara keseluruhan.

Tujuan penelitian ini adalah meningkatkan akurasi klasifikasi RD dan efisiensi komputasi dengan membuktikan efektivitas ADT dalam mengatasi ketidakseimbangan data, serta memberikan fondasi metodologis bagi pengembangan model deteksi otomatis yang berpotensi mendukung alur skrining klinis di Indonesia. Penelitian ini memberikan pemahaman mendalam tentang strategi augmentasi adaptif yang mampu meningkatkan generalisasi model *deep learning* pada citra fundus retina dengan ketidakseimbangan kelas. Hasil yang diperoleh diharapkan dapat berkontribusi pada peningkatan aksesibilitas deteksi dini RD dan upaya pencegahan risiko kebutaan.

METODE

Penelitian ini menerapkan desain eksperimental komparatif untuk mengevaluasi kinerja arsitektur DenseNet-121 dalam klasifikasi tingkat keparahan RD. Dataset yang digunakan bersumber dari APTOS 2019 Blindness Detection, terdiri dari 3.662 citra fundus teranotasi dalam lima kelas: *No DR* (1,805), *Mild* (370), *Moderate* (999), *Severe* (193), dan *Proliferative DR* (295). Guna menjamin validitas evaluasi pada data yang tidak seimbang, dataset dibagi menjadi data latih (70%), validasi (20%), dan uji (10%) menggunakan teknik *stratified sampling* untuk mempertahankan proporsi kelas yang konsisten di setiap subset.

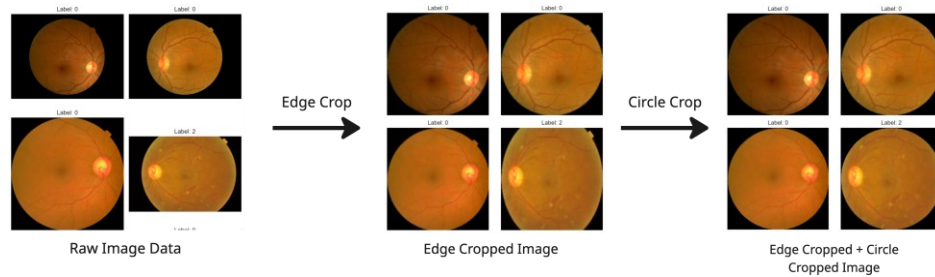


Gambar 1. Rancangan model klasifikasi retinopati diabetik

Tahapan preprocessing dilakukan seragam untuk menstandarisasi input model pada gambar 2. Proses ini mencakup pemotongan area retina (*cropping*) berbasis deteksi tepi *Canny* dan masker lingkaran untuk mengeliminasi *noise* latar belakang, perubahan ukuran (*resizing*) menjadi 224×224 piksel, serta normalisasi nilai piksel guna mempercepat konvergensi model.

Metode *Oversampling* Augmentasi, SMOTE, dan ADT diterapkan pada dataset latih; *Oversampling* Augmentasi dan SMOTE menggunakan teknik augmentasi statis, sedangkan ADT menggunakan augmentasi *on-the-fly*. Meskipun *Oversampling* Augmentasi dan SMOTE efektif dalam meningkatkan jumlah data pada kelas minoritas, keduanya memiliki keterbatasan serupa: augmentasi yang dilakukan bersifat statis dan tidak adaptif terhadap perubahan distribusi kelas selama pelatihan (Matharaarachchi et al., 2024). Sebaliknya, ADT memberikan variasi data secara dinamis selama pelatihan, yang dapat mengurangi *overfitting* dan meningkatkan keberagaman model tanpa menambah ukuran dataset, menjadikannya lebih efisien secara komputasi (Hao et al., 2021). Tabel 1 merangkum parameter augmentasi yang digunakan untuk *Oversampling* Augmentasi dan ADT, yang bertujuan untuk

memvariasikan orientasi, skala, dan distorsi citra, serta mengisi area di luar batas transformasi.



Gambar 2. Tahapan *preprocessing* data citra fundus

Tabel 1. Parameter transformasi augmentasi untuk *oversampling* augmentasi dan adt

Parameter	Nilai	Deskripsi
rotation_range	180	Rotasi acak dalam rentang 0 hingga 180 derajat
horizontal_flip	True	Pencerminan horizontal secara acak (50% probabilitas)
vertical_flip	True	Pencerminan vertikal secara acak (50% probabilitas)
zoom_range	0,15	Zoom in/out acak dengan faktor 0.85x hingga 1.15x
shear_range	0,1	Transformasi <i>shear</i> dengan intensitas 0-10%
fill_mode	<i>constant</i>	Mode pengisian untuk piksel di luar <i>boundary</i>
cval	0	Nilai konstan untuk fill_mode

Model dibangun menggunakan *backbone* DenseNet-121 dengan bobot *pre-trained* ImageNet (DenseNet-BC-121-32-no-top.h5). Modifikasi arsitektur dilakukan dengan menambahkan lapisan *Global Average Pooling*, *Dropout* (0,5), dan *Batch Normalization* sebelum klasifikasi *softmax*. Pelatihan menggunakan optimizer *Adam* (*learning rate* 0,00005) dan *loss function categorical cross-entropy* dengan *batch size* 32 selama 50 *epoch*. Untuk memastikan reproduktibilitas, *random seed* diinisialisasi pada nilai 2025. Evaluasi kinerja dilakukan menggunakan metrik Akurasi, *Precision*, *Recall*, dan *F1-Score*. Validitas statistik diperkuat dengan perhitungan *Confidence Interval* (CI) 95% berbasis *bootstrapping* (1000 sampel) serta Uji McNemar untuk menilai signifikansi perbedaan performa antar-model pada data uji.

HASIL DAN PEMBAHASAN

Hasil

Analisis hasil difokuskan pada evaluasi kemampuan generalisasi model DenseNet-121 dalam mengklasifikasikan 367 citra data uji independen (*unseen data*) di bawah empat skenario pelatihan yang berbeda. Bagian ini memaparkan perbandingan efektivitas antara pendekatan *Baseline*, strategi *Oversampling* Augmentasi, dan metode usulan *Augment During Training* (ADT) dalam menangani ketidakseimbangan kelas. Kinerja model dianalisis berdasarkan capaian metrik evaluasi kuantitatif serta hasil uji signifikansi statistik untuk memvalidasi keandalan peningkatan performa yang diamati.

Tabel 2 menyajikan perbandingan kinerja antara pendekatan-pendekatan yang diuji. ADT mencapai akurasi tertinggi sebesar 83,65% dengan CI 95% antara 81,53% - 85,77%, meskipun memiliki *Efficiency Ratio* (ER) 1,97, yang lebih rendah dibandingkan *Baseline* yang mencatatkan ER 2,56. SMOTE dan *Oversampling* Augmentasi masing-masing mencatatkan akurasi 80,38% dan 80,11%, dengan ER yang lebih rendah, yaitu 1,18 dan 1,09. Waktu pelatihan untuk ADT lebih cepat (4223 detik) dibandingkan dengan SMOTE (6714 detik) dan *Oversampling* Augmentasi (7174 detik).

Tabel 2. Perbandingan performa model pada data uji, waktu pelatihan dan efisiensi model

Pendekatan	Akurasi (95% CI)	Precision (95% CI)	Recall (95% CI)	F1-Score (95% CI)	Waktu Latih (detik)	Efficiency Ratio ($F1/time \times 10^{-4}$)
ADT	83,65% [81,53– 85,77]	83,80% [81,51– 85,75]	83,65% [81,53– 85,77]	83,12% [80,94– 85,29]	4223	1,97
<i>Oversampling</i> SMOTE	80,38% [77,84– 82,86]	79,10% [76,14– 81,62]	80,38% [77,84– 82,86]	79,42% [76,65– 81,80]	6714	1,18
<i>Baseline</i>	80,38% [78,66– 82,83]	78,87% [77,21– 81,79]	80,38% [78,66– 82,83]	79,46% [77,45– 82,01]	3103	2,56
<i>Oversampling</i> Augmentasi	80,11% [77,36– 82,02]	77,81% [75,55– 80,30]	80,11% [77,36– 82,02]	78,66% [76,05– 80,59]	7174	1,09

Analisis performa per kelas pada Tabel 3 mengindikasikan keunggulan model ADT, yang secara signifikan mengungguli model lain pada kelas *No DR* dan *Moderate* dengan skor masing-masing 0,9888 dan 0,7793. Namun, pada kelas *Severe* dan *Proliferative DR*, ADT memiliki *F1-Score* yang lebih rendah, masing-masing 0,2857 dan 0,5217. Dibandingkan dengan model lainnya, ADT memiliki kinerja yang lebih baik di semua kelas, dengan peningkatan signifikan pada kelas *No DR* dan *Moderate*. Namun, ADT menunjukkan kesulitan dalam mengklasifikasikan kelas minoritas, terlihat dari kinerjanya yang lebih rendah pada kelas *Severe* dan *Proliferative DR*.

Tabel 3. *F1-Score* per kelas untuk setiap model pada data uji

Kelas	F1-Baseline	F1-Augmentasi	F1-SMOTE	F1-ADT	Support
0 (<i>No DR</i>)	0,9780	0,9726	0,9725	0,9888	181
1 (<i>Mild</i>)	0,6666	0,6388	0,5882	0,7317	37
2 (<i>Moderate</i>)	0,7142	0,7373	0,7407	0,7793	100
3 (<i>Severe</i>)	0,1428	0,0740	0,2142	0,2857	19
4 (<i>Proliferative</i> DR)	0,5263	0,4615	0,5172	0,5217	30

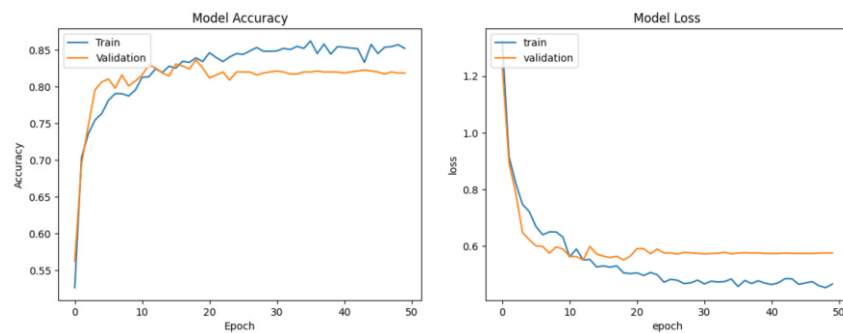
Berdasarkan data kuantitatif pada tabel 4, terlihat adanya tren positif kenaikan akurasi sebesar 3% pada model ADT, akan tetapi hasil uji McNemar untuk perbandingan dengan SMOTE, Augmentasi, dan *Baseline* menunjukkan p-value lebih besar dari 0,05 (0,0896 dan 0,0609). Hal ini mengindikasikan bahwa perbedaan kinerja antara ADT dan model lainnya tidak signifikan secara statistik. Uji McNemar dengan $p > 0,05$ menunjukkan bahwa perbedaan kinerja antara model mungkin dipengaruhi oleh faktor seperti ukuran sampel kecil dan ketidakseimbangan kelas, yang mengurangi kemampuan deteksi perbedaan signifikan.

Hasil pada gambar 3 memvisualisasikan dinamika pembelajaran model ADT, yang merepresentasikan progres akurasi dan nilai *loss* sepanjang durasi pelatihan 50 *epoch*. Pada *epoch* ke-15, akurasi model mencapai kisaran 0,83 hingga 0,85 dan menunjukkan konvergensi. Setelah itu, akurasi validasi stabil di sekitar 0,81, sementara akurasi *training* mulai stabil di sekitar 0,84 mulai *epoch* ke-20, yang menunjukkan bahwa model tidak mengalami *overfitting*. Pada grafik *loss*, nilai *validation loss* mulai stabil pada *epoch* ke-25,

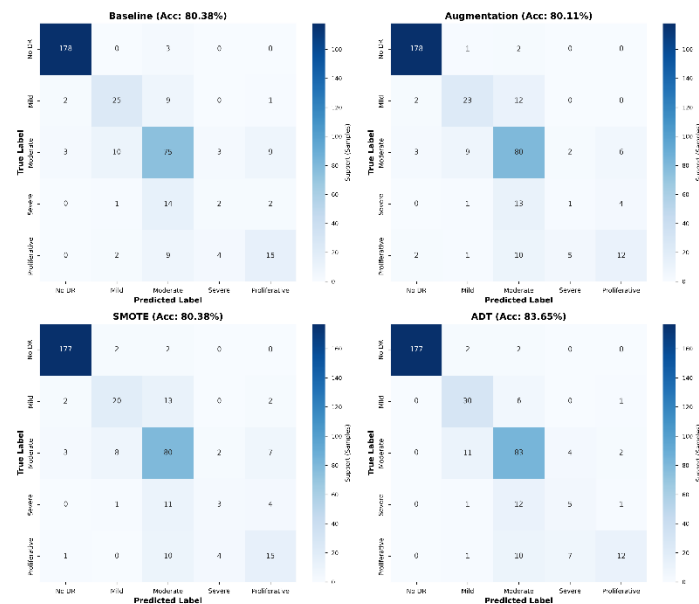
menunjukkan bahwa model telah mencapai titik stabil dengan nilai kurang dari 0,2, yang sebanding antara *training loss* dan *validation loss*.

Tabel 4. Hasil uji signifikansi mcnemar antar-model pada data uji

Perbandingan	χ^2	<i>p-value</i>
ADT vs SMOTE	2,881	0,0896
ADT vs Augmentasi	3,512	0,0609
ADT vs Baseline	2,881	0,0896



Gambar 3. Hasil akurasi dan loss model dengan pendekatan adt

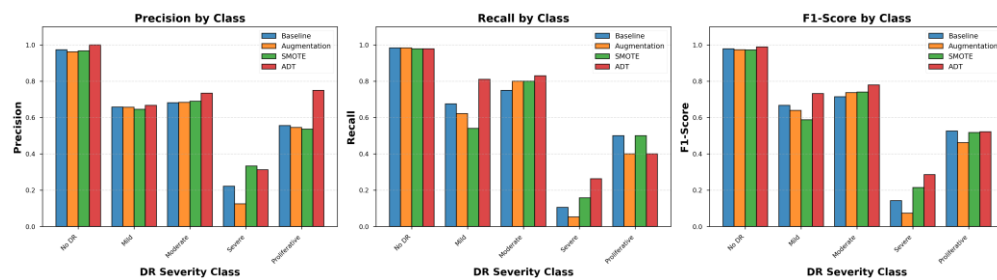


Gambar 4. Confusion Matrix hasil klasifikasi model pada data uji

Selanjutnya, pada gambar 4 menyajikan analisis komparatif performa antar-model melalui diagram *barchart*, yang merinci distribusi *Precision*, *Recall*, dan *F1-Score* secara spesifik untuk setiap kelas pada berbagai pendekatan model. Model ADT menunjukkan performa unggul, terutama pada kelas *No DR* dan *Moderate*. Model ADT juga menunjukkan kinerja terbaik pada *Proliferative* DR di sisi *Precision*, serta konsisten memberikan hasil yang baik di semua kelas lainnya.

Sementara itu, gambar 5 menyajikan analisis detail performa klasifikasi melalui *Confusion Matrix* pada dataset pengujian. Pada kelas *No DR*, semua model hampir sepenuhnya mengklasifikasikan sampel dengan benar, dengan 177-178 dari 181 sampel terklasifikasi dengan tepat. Namun, pada kelas *Severe*, semua model menunjukkan *Recall* yang rendah, dengan banyak sampel *Severe* yang salah diklasifikasikan sebagai *Moderate*

atau *Proliferative*. Model ADT memberikan akurasi terbaik, terutama pada kelas *No DR*, *Mild* dan *Moderate*, meskipun kesulitan dalam mendeteksi kelas *Severe* dan *Proliferative* DR.



Gambar 5. Performa model per kelas pada data uji (*precision, recall, f1-score*)

Pembahasan

Temuan utama dari penelitian ini menunjukkan bahwa strategi augmentasi dinamis ADT secara konsisten mengungguli pendekatan augmentasi statis dan *oversampling* konvensional dalam mengklasifikasikan retinopati diabetik. ADT mencapai akurasi 83,65%, mengungguli *baseline*, *oversampling augmentasi*, dan SMOTE dengan selisih 3–3,27% (Tabel 2). Meskipun peningkatan ini tidak signifikan secara statistik ($p > 0,05$; Tabel 4), peningkatan akurasi 3–4% tetap memiliki relevansi klinis yang penting dalam skrining massal. Hal ini dapat meningkatkan jumlah pasien yang terdiagnosis dengan benar dan memungkinkan intervensi dini untuk mencegah progresivitas penyakit dan kebutaan (Aggarwal et al., 2021). Keterbatasan signifikansi statistik disebabkan oleh ukuran sampel uji yang kecil ($n=367$) dan ketidakseimbangan kelas yang ekstrem, yang mencerminkan tantangan dunia nyata dalam setting klinis (Song et al., 2021). Semua eksperimen menggunakan arsitektur model yang sama (DenseNet-121), sehingga perbedaan hasil utama berasal dari strategi penanganan ketidakseimbangan kelas, bukan kapasitas model.

Keunggulan ADT terletak pada kemampuannya untuk menciptakan lingkungan pelatihan *stochastic* yang lebih mendekati distribusi data dunia nyata. Dengan menghasilkan variasi data unik setiap *epoch*, ADT berfungsi sebagai regularisasi implisit yang mengurangi risiko *overfitting* dibandingkan augmentasi statis yang terbatas (Kumar et al., 2025). Selain itu, penggunaan DenseNet-121 dengan arsitektur konektivitas padatnya mengoptimalkan propagasi gradien dan penggunaan kembali fitur, memungkinkan model untuk menangkap variasi augmentasi dinamis secara lebih efektif (Mas'ud & Junta, 2024). Efisiensi komputasi ADT tercermin dalam rasio efisiensi 1,97 pada Tabel 4, yang menunjukkan keseimbangan optimal antara peningkatan akurasi dan beban komputasi. Hal ini membuktikan bahwa ADT tidak hanya meningkatkan performa tetapi juga praktis untuk implementasi klinis.

Analisis *F1-Score* per kelas (Tabel 3) menunjukkan bahwa ADT memiliki kinerja tinggi pada kelas *No DR* (98,8%) dan *Moderate* (77,9%), namun kinerjanya menurun tajam pada kelas *Severe* dan *Proliferative* DR. Sebagian besar kasus *Severe* salah diklasifikasikan sebagai *Moderate* atau *Proliferative*. Penurunan ini tidak hanya dipengaruhi oleh ukuran sampel yang kecil, tetapi juga oleh tumpang tindih morfologis antara *Severe*, dengan perdarahan dan eksudat luas, dan tahap awal *Proliferative*, di mana neovaskularisasi sulit terdeteksi (Zhao et al., 2023). *Confusion matrix* pada gambar 5 menunjukkan konsistensi kesalahan klasifikasi yang mendukung hipotesis ini. Seperti yang dilaporkan oleh Yi et al. (2024), angiografi lesi pada tahap ini sering kali tumpang tindih, membuat model kesulitan membedakan IRMA yang *advanced* dari neovaskularisasi awal. Penelitian sebelumnya (Hu et al., 2024; Khan et al., 2023) juga menunjukkan bahwa deteksi tahap *Severe* dan *Proliferative* adalah tantangan umum dalam retinopati diabetik. Oleh karena itu, misklasifikasi *Severe*

sebagai *Proliferative* kemungkinan besar disebabkan oleh tumpang tindih morfologis dan angiografik, bukan semata-mata oleh ketidakseimbangan kelas.

Penerapan strategi ADT pada arsitektur DenseNet-121 terbukti menghasilkan akurasi 83,65%, melampaui rentang kinerja referensi sebelumnya seperti studi Khan et al. (2023) dan Gangwar & Ravi (2021) yang berkisar pada 75–83%. Meskipun akurasi ADT lebih tinggi, *F1-Score* untuk kelas *Severe* dan *Proliferative* tetap menunjukkan penurunan. Namun, perbaikan utama yang ditawarkan oleh ADT terletak pada stabilitas pelatihan yang lebih baik tanpa *overfitting*, serta kemampuan untuk mempertahankan performa pada kelas minor tanpa memperburuk *confusion*. Sebagai tambahan, studi Matharaarachchi et al. (2024) menunjukkan bahwa augmentasi statis seperti SMOTE tidak cukup efektif dalam menangani perbedaan *subclass* yang *imbalanced*, yang menegaskan bahwa ADT memberikan kemajuan yang signifikan dalam regularisasi dan generalisasi, meskipun bukan solusi final, dalam deteksi retinopati diabetik yang kompleks.

Stabilitas pelatihan ADT dapat terlihat pada gambar 3, dengan konvergensi yang stabil hingga *epoch* ke-50 tanpa *overfitting*, diperkuat oleh arsitektur DenseNet-121 yang memanfaatkan *deep supervision* dan konektivitas padat untuk memitigasi masalah *vanishing gradient* dan meningkatkan efisiensi propagasi gradien (Hasan et al., 2021). Namun, *confusion matrix* pada gambar 5 menunjukkan kelemahan semua model dalam mendeteksi *Severe* dan *Proliferative* DR, masalah yang juga tercatat dalam literatur global dan nasional. Alokasi pengawasan klinis pada tahap *Severe* sangat penting untuk mencegah progresi ke tahap *Proliferative* yang dapat menyebabkan kebutaan (Puteri et al., 2022). Kami berhipotesis bahwa augmentasi berbasis citra global seperti ADT masih kurang efektif dalam menangkap perbedaan subtil antar-lesi. Pendekatan *multi-task learning* yang mencakup klasifikasi keparahan dan segmentasi lesi kunci seperti mikroaneurisma dan neovaskularisasi dapat diuji sebagai langkah penelitian selanjutnya. Temuan ini menegaskan bahwa *underdiagnosis* pada kelas *Severe* dan *Proliferative* DR merupakan isu kritis yang perlu menjadi fokus penelitian lanjutan guna menjamin keamanan implementasi klinis dan memastikan model berperan efektif sebagai alat triase berbasis AI, bukan sebagai alat diagnosis mandiri.

Meskipun pendekatan ADT menunjukkan hasil yang menjanjikan, penelitian ini memiliki beberapa keterbatasan. Pertama, dataset APTOS 2019 berasal dari populasi *India*, sehingga validitas model pada populasi Indonesia perlu diuji lebih lanjut. Kedua, jumlah sampel uji yang terbatas, terutama pada kelas minoritas (*Severe*: 19, *Proliferative* DR: 30), menyebabkan hasil uji McNemar yang tidak signifikan, mengindikasikan bahwa perbedaan performa antar-metode belum cukup kuat. Ketiga, kesulitan model dalam mendeteksi kelas *Severe* dan *Proliferative* DR (*F1-Score* 0,2857 dan 0,5217) mencerminkan tantangan dalam membedakan lesi dengan fitur visual yang tumpang tindih. Penelitian lanjutan sebaiknya mengintegrasikan teknik *deep learning* terkini, seperti *Vision Transformer* atau *hybrid CNN-Transformer*, yang lebih efektif dalam menangkap dependensi jangka panjang pada citra retina, serta menerapkan *cost-sensitive learning* atau *focal loss* untuk meningkatkan perhatian pada kelas minoritas. Validasi eksternal pada dataset Indonesia juga diperlukan untuk mengevaluasi generalisasi domain dan kelayakan klinis, sementara eksplorasi metode *ensemble* dapat meningkatkan kekuatan diskriminatif pada kelas-kelas yang lebih sulit.

SIMPULAN

Penelitian ini mengonfirmasi bahwa ADT secara signifikan meningkatkan akurasi model klasifikasi retinopati diabetik, mencapai 83,65%, serta efisiensi komputasi dengan *Efficiency Ratio* (ER) 1,97, mengungguli pendekatan lain seperti *oversampling* augmentasi dan SMOTE. ADT terbukti memperbaiki generalisasi model tanpa perlu memperbesar dataset atau menyimpan data sintesis tambahan. Meskipun model ini menunjukkan performa terbaik pada kelas *No DR* dan *Moderate*, tantangan tetap teridentifikasi pada klasifikasi kelas

Severe dan *Proliferative* DR yang terkait dengan ketidakseimbangan data. Penelitian ini menunjukkan potensi ADT untuk mendukung skenario skrining massal, terutama di daerah dengan keterbatasan infrastruktur medis, dengan menyediakan landasan komputasi yang kuat bagi pengembangan alat bantu deteksi dini retinopati diabetik berbasis AI yang efisien di masa depan. Ke depannya, integrasi model berbasis ADT ini ke dalam alur skrining klinis di Indonesia berpotensi memperluas aksesibilitas deteksi dini, mengurangi risiko kebutaan, dan meningkatkan hasil kesehatan secara keseluruhan.

REFERENSI

- Aggarwal, R., Sounderajah, V., Martin, G., Ting, D. S. W., Karthikesalingam, A., King, D., Ashrafian, H., & Darzi, A. (2021). Diagnostic accuracy of deep learning in medical imaging: A systematic review and meta-analysis. *Npj Digital Medicine*, 4(1), 65. <https://doi.org/10.1038/s41746-021-00438-z>
- Atwany, M. Z., Sahyoun, A. H., & Yaqub, M. (2022). Deep Learning Techniques for Diabetic Retinopathy Classification: A Survey. *IEEE Access*, 10, 28642–28655. <https://doi.org/10.1109/ACCESS.2022.3157632>
- Benhamza, M., Dahlui, M., & Said, M. A. (2024). Determining direct, indirect healthcare and social costs for diabetic retinopathy management: A systematic review. *BMC Ophthalmology*, 24(1), 424. <https://doi.org/10.1186/s12886-024-03665-6>
- Gangwar, A. K., & Ravi, V. (2021). Diabetic Retinopathy Detection Using Transfer Learning and Deep Learning. In V. Bhateja, S.-L. Peng, S. C. Satapathy, & Y.-D. Zhang (Eds.), *Evolution in Computational Intelligence*. (1176), 679–689. Springer Singapore. https://doi.org/10.1007/978-981-15-5788-0_64
- Hao, R., Namdar, K., Liu, L., Haider, M. A., & Khalvati, F. (2021). A Comprehensive Study of Data Augmentation Strategies for Prostate Cancer Detection in Diffusion-Weighted MRI Using Convolutional Neural Networks. *Journal of Digital Imaging*, 34(4), 862–876. <https://doi.org/10.1007/s10278-021-00478-7>
- Hasan, N., Bao, Y., Shawon, A., & Huang, Y. (2021). DenseNet Convolutional Neural Networks Application for Predicting COVID-19 Using CT Image. *SN Computer Science*, 2(5), 389. <https://doi.org/10.1007/s42979-021-00782-7>
- Hu, A., Schmidt, M. H. H., & Heinig, N. (2024). Microglia in retinal angiogenesis and diabetic retinopathy. *Angiogenesis*, 27(3), 311–331. <https://doi.org/10.1007/s10456-024-09911-1>
- Khan, M. B., Ahmad, M., Yaakob, S. B., Shahrir, R., Rashid, M. A., & Higa, H. (2023). Automated Diagnosis of Diabetic Retinopathy Using Deep Learning: On the Search of Segmented Retinal Blood Vessel Images for Better Performance. *Bioengineering*, 10(4), 413. <https://doi.org/10.3390/bioengineering10040413>
- Kumar, S., Asiamah, P., Jolaoso, O., & Esiowu, U. (2025). Enhancing Image Classification with Augmentation: Data Augmentation Techniques for Improved Image Classification (No. arXiv:2502.18691). arXiv. <https://doi.org/10.48550/arXiv.2502.18691>
- Luo, X., Zhang, B., Wang, W., Xu, Y., Lai, Z., Jin, X., & Zhang, D. (2022). A deep convolutional neural network for diabetic retinopathy detection via mining local and long-range dependence. *CAAI Transactions on Intelligence Technology*, 7(4), 701–713. <https://doi.org/10.1049/cit2.12155>
- Mas'ud, R. A., & Junta, Z. (2024). Optimasi Convolutional Neural Networks untuk Deteksi Kanker Payudara menggunakan Arsitektur DenseNet. *Edumatic: Jurnal Pendidikan Informatika*, 8(1), 310–318. <https://doi.org/10.29408/edumatic.v8i1.25883>
- Matharaarachchi, S., Domaratzki, M., & Muthukumarana, S. (2024). Enhancing SMOTE for imbalanced data with abnormal minority instances. *Machine Learning with Applications*, 18, 100597. <https://doi.org/10.1016/j.mlwa.2024.100597>

- Prajnaparamita, I. (2025). Diabetic Retinopathy: A Growing Concern for Diabetes Patients in Indonesia. *Journal Scientific of Mandalika (JSM)*, 6(4), 768–778.
- Puteri, V., Lassie, N., & Huda, M. N. (2022). Gambaran Karakteristik Pasien Retinopati Diabetik yang Dilakukan Pembedahan Vitrektomi Di RSKM Padang Eye Center Tahun 2019-2020. *Scientific Journal*, 1(3), 177–191. <https://doi.org/10.56260/sciena.v1i3.40>
- Santika, E. F. (2024). Indonesia's Diabetes Prevalence Rises to 11.7% in 2023. Databoks. Retrieved November 17, 2025, from [Indonesia's Diabetes Prevalence Rises to 11.7% in 2023](#)
- Shaniaputri, T., Iskandar, E., & Fajriansyah, A. (2022). Prevalensi Retinopati Diabetik di Puskesmas di Bandung Raya Periode Januari 2019-Desember 2020. *eJournal Kedokteran Indonesia*, 10(1), 39–45. <https://doi.org/10.23886/ejki.10.119.39-45>
- Song, B., Li, S., Sunny, S., Gurushanth, K., Mendonca, P., Mukhia, N., Patrick, S., Gurudath, S., Raghavan, S., Tsusennaro, I., Leivon, S. T., Kolor, T., Shetty, V., Bushan, V., Ramesh, R., Peterson, T., Pillai, V., Wilder-Smith, P., Sigamani, A., ... Liang, R. (2021). Classification of imbalanced oral cancer image data from high-risk population. *Journal of Biomedical Optics*, 26(10). <https://doi.org/10.1117/1.JBO.26.10.105001>
- Ullah, N., Mohmand, M. I., Ullah, K., Gismalla, M. S. M., Ali, L., Khan, S. U., & Ullah, N. (2022). Diabetic retinopathy detection using genetic algorithm-based CNN features and error correction output code SVM framework. *Wireless Communications and Mobile Computing*, 2022, Article ID 7095528. <https://doi.org/10.1155/2022/7095528>
- Vishwanath, C. S., Divya, K. L., Vamsika, A., Murali, B. S., & Dinesh, B. (2024). Diabetic Retinopathy Classification by Enhanced Image Filtration Techniques Using Deep Learning. *2024 OPJU International Technology Conference (OTCON) on Smart Computing for Innovation and Advancement in Industry 4.0*, 1–7. India: IEEE. <https://doi.org/10.1109/OTCON60325.2024.10687473>
- Wahidin, M., Achadi, A., Besral, B., Kosen, S., Nadjib, M., Nurwahyuni, A., Ronoatmodjo, S., Rahajeng, E., Pane, M., & Kusuma, D. (2024). Projection of diabetes morbidity and mortality till 2045 in Indonesia based on risk factors and NCD prevention and control programs. *Scientific Reports*, 14(1), 5424. <https://doi.org/10.1038/s41598-024-54563-2>
- Yang, Z., Tan, T.-E., Shao, Y., Wong, T. Y., & Li, X. (2022). Classification of diabetic retinopathy: Past, present and future. *Frontiers in Endocrinology*, 13, 1079217. <https://doi.org/10.3389/fendo.2022.1079217>
- Yi, D., Baltov, P., Hua, Y., Philip, S., & Sharma, P. K. (2024). Compound Scaling Encoder-Decoder (CoSED) Network for Diabetic Retinopathy Related Bio-Marker Detection. *IEEE Journal of Biomedical and Health Informatics*, 28(4), 1959–1970. <https://doi.org/10.1109/JBHI.2023.3313785>
- Zhao, P. Y., Bommakanti, N., Yu, G., Aaberg, M. T., Patel, T. P., & Paulus, Y. M. (2023). Deep learning for automated detection of neovascular leakage on ultra-widefield fluorescein angiography in diabetic retinopathy. *Scientific Reports*, 13(1), 9165. <https://doi.org/10.1038/s41598-023-36327-6>