

PERBANDINGAN PERFORMA RESNET-18, EFFICIENTNET-B0, DAN ENSEMBLE LEARNING DENGAN VISUALISASI GRAD-CAM PADA DETEKSI KATARAK

Wahyu Putra Pratama, Dwi Arman Prasetya, Andri Fauzan Adziima

Sains Data, Fakultas Ilmu Komputer, UPN "Veteran" Jawa Timur, Indonesia

Jl. Raya Rungkut Madya, Gunung Anyar, Surabaya

wppratama76@gmail.com

ABSTRAK

Katarak merupakan salah satu penyebab utama kebutaan di dunia yang memerlukan deteksi dini untuk mencegah penurunan penglihatan permanen. Namun, diagnosis manual oleh tenaga medis rentan terhadap bias subjektif dan membutuhkan waktu yang tidak singkat, sehingga diperlukan sistem deteksi otomatis berbasis kecerdasan buatan yang akurat sekaligus interpretatif. Penelitian ini bertujuan mengembangkan model klasifikasi katarak menggunakan pendekatan stacking ensemble yang menggabungkan ResNet-18 dan EfficientNet-B0 sebagai base learner dengan Logistic Regression sebagai meta-learner, untuk mengklasifikasikan citra slit-lamp mata ke dalam tiga kelas, yaitu normal, immature, dan mature. Untuk meningkatkan transparansi model, penelitian ini juga mengintegrasikan metode Gradient-weighted Class Activation Mapping (Grad-CAM) sebagai teknik Explainable Artificial Intelligence (XAI) guna memvisualisasikan area citra yang paling berkontribusi terhadap keputusan prediksi model. Dataset yang digunakan terdiri atas 2.421 citra yang diperoleh dari platform Roboflow dan Kaggle, dengan tahap prapemrosesan meliputi resize, augmentasi, normalisasi, dan pembagian data. Hasil eksperimen menunjukkan bahwa model ensemble stacking mencapai akurasi tertinggi sebesar 98,37%, melampaui performa model tunggal ResNet-18 (92,15%) dan EfficientNet-B0 (94,01%). Visualisasi Grad-CAM mengonfirmasi bahwa model secara konsisten memusatkan perhatian pada area sentral lensa yang mengalami kekeruhan, sesuai dengan karakteristik klinis katarak. Penelitian ini membuktikan bahwa pendekatan stacking ensemble yang dikombinasikan dengan XAI mampu menghasilkan sistem klasifikasi katarak yang akurat dan transparan, sehingga berkontribusi pada pengembangan sistem deteksi katarak berbasis kecerdasan buatan.

Kata kunci : Katarak, Deep Learning, ResNet-18, EfficientNet-B0, Stacking Ensemble, Grad-CAM, Explainable AI

1. PENDAHULUAN

Katarak menjadi salah satu penyebab kebutaan terbesar di dunia yang sebetulnya dapat dihindari apabila dideteksi dan ditangani sejak awal. Data dari World Health Organization (WHO) menunjukkan bahwa lebih dari 65 juta orang di seluruh dunia mengalami kebutaan yang disebabkan oleh katarak, dengan tingkat prevalensinya yang cenderung meningkat seiring bertambahnya usia seseorang[1].

Di Indonesia, data Kementerian Kesehatan RI menunjukkan bahwa katarak menyumbang sekitar 80% dari seluruh kasus kebutaan nasional[2], [3].

Kondisi ini diperparah oleh keterbatasan jumlah dokter spesialis mata dan rendahnya tingkat kesadaran masyarakat terhadap pemeriksaan dini, sehingga banyak kasus baru diketahui pada stadium lanjut. Secara medis, katarak adalah kekeruhan pada lensa mata yang menyebabkan penglihatan menjadi buram atau kabur [4].

Faktor risiko utamanya meliputi penuaan, paparan sinar ultraviolet berlebih, penggunaan obat kortikosteroid jangka panjang, serta komplikasi dari penyakit sistemik seperti diabetes melitus, yang menyebabkan perubahan metabolik pada lensa mata [5].

Peningkatan angka penderita diabetes di dunia turut berkontribusi terhadap meningkatnya prevalensi katarak, karena hiperglikemia kronis mempercepat kerusakan protein lensa dan memicu kekeruhan [6]

Meskipun diagnosis katarak dapat dilakukan melalui pemeriksaan klinis oleh dokter spesialis mata, proses tersebut masih menghadapi berbagai kendala, terutama keterbatasan jumlah tenaga medis dan fasilitas kesehatan di beberapa wilayah. Selain itu, diagnosis manual berpotensi menimbulkan variasi penilaian antar pemeriksa dan dapat menyebabkan keterlambatan deteksi dini. Kondisi ini menunjukkan perlunya sistem pendukung diagnosis yang mampu membantu proses identifikasi katarak secara otomatis, cepat, dan konsisten. Perkembangan Artificial Intelligence (AI), khususnya Deep Learning, telah membuka peluang dalam pengembangan sistem deteksi penyakit berbasis citra medis. Salah satu pendekatan yang banyak digunakan adalah Convolutional Neural Network (CNN), yang terbukti mampu mengekstraksi fitur visual secara otomatis dan menghasilkan performa klasifikasi yang tinggi pada berbagai kasus penyakit mata.[7][8].

Berbagai arsitektur CNN telah dikembangkan, di antaranya ResNet-18 dan EfficientNet-B0. ResNet-18 memiliki kemampuan mengatasi permasalahan vanishing gradient melalui mekanisme residual learning, sedangkan EfficientNet-B0 menawarkan efisiensi komputasi melalui metode compound scaling yang menyeimbangkan kedalaman, lebar, dan resolusi jaringan. Namun demikian, penggunaan model tunggal sering kali memiliki keterbatasan dalam menangkap kompleksitas fitur pada citra medis.

Pendekatan ensemble learning menawarkan alternatif untuk meningkatkan performa klasifikasi dengan mengombinasikan prediksi dari beberapa model yang berbeda. Selain itu, dalam bidang medis, tingkat akurasi yang tinggi saja belum cukup untuk mendukung pengambilan keputusan klinis. Model yang digunakan juga perlu memiliki tingkat interpretabilitas yang baik sehingga alasan di balik suatu prediksi dapat dipahami oleh tenaga medis. Salah satu metode yang banyak digunakan untuk meningkatkan transparansi model adalah Gradient-weighted Class Activation Mapping (Grad-CAM), yang mampu memvisualisasikan area citra yang menjadi dasar pengambilan keputusan model.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk membandingkan performa model ResNet-18, EfficientNet-B0, dan pendekatan Ensemble Learning berbasis stacking dalam klasifikasi citra slit-lamp mata untuk deteksi katarak. Penelitian ini juga mengintegrasikan metode Grad-CAM untuk memberikan interpretasi visual terhadap hasil klasifikasi yang dihasilkan oleh masing-masing model. Sebagai upaya penyelesaian masalah, penelitian ini mengembangkan tiga skenario klasifikasi, yaitu menggunakan ResNet-18 sebagai model tunggal, EfficientNet-B0 sebagai model tunggal, dan Ensemble Learning berbasis stacking yang menggabungkan kedua model tersebut. Kinerja setiap model dievaluasi menggunakan metrik akurasi, precision, recall, dan F1-score. Selanjutnya, hasil klasifikasi dianalisis menggunakan Grad-CAM untuk mengidentifikasi area citra yang menjadi fokus perhatian model. Dengan pendekatan tersebut, penelitian ini diharapkan dapat memberikan informasi mengenai model yang paling efektif sekaligus menghasilkan sistem deteksi katarak yang lebih transparan dan dapat dipercaya dalam mendukung proses diagnosis klinis.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Penelitian mengenai deteksi katarak berbasis Deep Learning telah berkembang pesat dalam beberapa tahun terakhir. Salah satu penelitian yang dilakukan oleh *Automatic Cataract Diagnosis by Image-Based Interpretability* memanfaatkan arsitektur ResNet-18 dan ResNet-50 untuk mengklasifikasikan empat tingkat keparahan katarak, yaitu normal, mild, moderate, dan severe berdasarkan citra fundus[9].

Penelitian tersebut menggunakan kanal hijau (*green channel*) sebagai input utama dan menerapkan metode Gradient-weighted Class Activation Mapping (Grad-CAM) untuk menghasilkan visualisasi area yang menjadi perhatian model selama proses klasifikasi. Hasil penelitian menunjukkan bahwa ResNet-50 mampu mencapai akurasi hingga 95% pada klasifikasi biner dan 77% pada klasifikasi multikelas, sehingga menunjukkan potensi yang baik dalam mendeteksi serta melakukan penilaian tingkat keparahan katarak secara otomatis.

Penelitian lain berjudul *Application of AlexNet, EfficientNetV2B0, and VGG19 with Explainable AI for Cataract and Glaucoma Image Classification* melakukan klasifikasi citra mata ke dalam tiga kategori, yaitu normal, katarak, dan glaukoma menggunakan tiga arsitektur Convolutional Neural Network (CNN), yaitu AlexNet, VGG19, dan EfficientNetV2B0[10].

Penelitian tersebut mengevaluasi berbagai skenario pembagian data (*holdout split*) dan algoritma optimasi untuk memperoleh konfigurasi terbaik. Selain itu, metode Grad-CAM digunakan untuk meningkatkan interpretabilitas model. Hasil penelitian menunjukkan bahwa EfficientNetV2B0 dengan pembagian data 90:10 dan optimizer RMSProp memberikan performa terbaik dengan akurasi sebesar 89,77%, serta nilai precision, recall, dan F1-score yang seimbang. Visualisasi Grad-CAM juga menunjukkan bahwa model mampu memfokuskan perhatian pada area opasitas lensa yang menjadi indikator utama katarak.

Selanjutnya, penelitian berjudul *Cataract Early Detection Using ResNet150 for Eye Fundus Images* mengembangkan sistem deteksi dini katarak menggunakan arsitektur ResNet150 pada citra fundus mata.[11]

Sebelum proses pelatihan model, dilakukan tahap prapemrosesan berupa reduksi noise, peningkatan kontras, dan normalisasi citra untuk meningkatkan kualitas data. Penelitian tersebut juga membandingkan beberapa model berbasis transfer learning, seperti ResNet150, VGG19, dan DenseNet121. Berdasarkan hasil evaluasi menggunakan metrik akurasi, precision, recall, F1-score, dan ROC-AUC, ResNet150 menunjukkan performa terbaik dalam membedakan citra mata normal dan katarak, sehingga dinilai efektif untuk mendukung deteksi dini penyakit tersebut.

Berdasarkan penelitian-penelitian terdahulu, dapat disimpulkan bahwa arsitektur berbasis ResNet maupun EfficientNet memiliki kemampuan yang baik dalam klasifikasi citra katarak. Selain itu, penggunaan metode Explainable Artificial Intelligence (XAI) seperti Grad-CAM terbukti mampu meningkatkan interpretabilitas model dengan menunjukkan area citra yang menjadi dasar pengambilan keputusan. Namun, sebagian besar penelitian masih berfokus pada evaluasi model tunggal (*single architecture*) dan belum banyak yang melakukan perbandingan langsung antara ResNet-18, EfficientNet-B0, serta pendekatan Ensemble Learning berbasis stacking pada dataset yang sama. Oleh karena itu, penelitian ini dilakukan untuk membandingkan performa ketiga pendekatan tersebut sekaligus mengintegrasikan Grad-CAM guna memperoleh model deteksi katarak yang tidak hanya akurat, tetapi juga lebih transparan dan dapat dipercaya dalam mendukung proses diagnosis medis.

2.2. Katarak

Katarak merupakan kondisi ketika lensa mata mengalami kekeruhan progresif akibat perubahan struktur protein, sehingga mengganggu masuknya cahaya ke retina dan menurunkan ketajaman visual. Katarak masih menjadi penyebab utama kebutaan global, dengan kontribusi lebih dari 45% kasus kebutaan menurut laporan Lancet Global Health [12]

WHO juga menegaskan bahwa katarak merupakan penyakit mata degeneratif yang paling banyak ditemukan pada populasi usia lanjut, namun dapat pula muncul lebih cepat akibat faktor metabolik seperti diabetes, trauma, penggunaan steroid jangka panjang, atau kelainan kongenital.

Secara klinis, katarak diklasifikasikan ke dalam beberapa stadium perkembangan, yaitu: *incipient*, *immature*, *mature*, dan *hypermature* masing-masing menunjukkan tingkat kekeruhan lensa yang semakin menebal dan berdampak langsung pada penurunan fungsi visual.

Penelitian epidemiologis Khairallah et al.[4] menyatakan bahwa perkembangan katarak sering kali berlangsung perlahan, sehingga deteksi dini menjadi sangat penting untuk mencegah penurunan penglihatan permanen.

2.3. Arsitektur ResNet

ResNet (*Residual Network*) merupakan salah satu arsitektur jaringan saraf konvolusional (*Convolutional Neural Network / CNN*) yang diperkenalkan oleh He et al. pada tahun 2015. Arsitektur ini dikembangkan untuk mengatasi permasalahan *degradation problem* yang muncul ketika kedalaman jaringan (jumlah lapisan) ditingkatkan. Pada jaringan konvensional, penambahan lapisan yang lebih dalam seringkali tidak meningkatkan akurasi, bahkan dapat menurunkan performa akibat terjadinya masalah *vanishing gradient* dan *overfitting* [13].

Keunggulan utama ResNet terletak pada penggunaan konsep *residual learning* melalui mekanisme *skip connection* atau *shortcut connection*. Mekanisme ini memungkinkan aliran informasi melewati satu atau lebih lapisan tanpa mengalami transformasi non-linear, sehingga gradien dapat ditransmisikan secara lebih stabil selama proses *backpropagation*. Dengan cara ini, ResNet mampu melatih jaringan yang sangat dalam, bahkan hingga ratusan lapisan, tanpa mengalami degradasi performa yang signifikan [14].

Secara umum, blok dasar ResNet disebut sebagai *residual block*, yang terdiri dari beberapa lapisan konvolusi, *batch normalization*, fungsi aktivasi (*ReLU*), serta jalur pintas (*shortcut path*). Jalur pintas ini menambahkan masukan awal (*input*) dengan keluaran transformasi (*output*) dari lapisan konvolusi, sehingga model belajar fungsi residu (*residual function*) alih-alih fungsi transformasi secara langsung. Formulasi matematis dari residual learning dapat dituliskan sebagai berikut:

2.4. Arsitektur EfficientNet

EfficientNet merupakan salah satu arsitektur *Convolutional Neural Network (CNN)* yang diperkenalkan oleh Tan dan Le (2019). Arsitektur ini dikembangkan dengan tujuan untuk mengatasi permasalahan efisiensi dalam pembelajaran mendalam, terutama pada model dengan jumlah parameter yang besar. EfficientNet menggunakan pendekatan *compound scaling method*, yaitu metode yang secara bersamaan melakukan skala pada tiga dimensi utama jaringan: kedalaman (*depth*), lebar (*width*), dan resolusi citra (*resolution*) secara terkoordinasi. Pendekatan ini terbukti memberikan hasil yang lebih baik dibandingkan dengan metode konvensional yang hanya berfokus pada peningkatan satu dimensi tertentu saja [15].

Pendekatan ini terbukti memberikan hasil yang lebih baik dibandingkan dengan metode konvensional yang hanya berfokus pada peningkatan satu dimensi tertentu saja. MBConv merupakan pengembangan dari *depthwise separable convolution* yang sebelumnya digunakan pada MobileNetV2. Blok MBConv terdiri dari tiga tahap utama, yaitu *expansion layer*, *depthwise convolution*, serta *projection layer*. Di samping itu, MBConv juga dilengkapi dengan blok *squeeze-and-excitation (SE)* yang berperan dalam memberikan bobot perhatian pada setiap saluran (*channel attention*), sehingga jaringan dapat lebih memusatkan perhatiannya pada fitur-fitur yang paling relevan dari citra masukan [16].

2.5. Stacking Ensemble

Metode ensemble merupakan pendekatan dalam pembelajaran mesin yang bertujuan untuk meningkatkan performa model dengan menggabungkan beberapa model prediktif. Salah satu teknik ensemble yang cukup populer adalah *stacking* (*stacked generalization*), yaitu metode yang menggabungkan output dari beberapa model dasar (*base learners*) menggunakan model lain yang disebut *meta-learner*. Berbeda dengan metode ensemble sederhana seperti bagging atau boosting, stacking tidak hanya menggabungkan prediksi, tetapi juga mempelajari cara terbaik untuk mengkombinasikan output tersebut sehingga menghasilkan prediksi yang lebih akurat. Dalam pendekatan stacking, beberapa model dasar dilatih secara independen menggunakan dataset yang sama. Pada penelitian ini, model ResNet dan EfficientNet digunakan sebagai *base learners* karena keduanya memiliki karakteristik ekstraksi fitur yang berbeda. ResNet unggul dalam menangani jaringan yang dalam melalui residual connection, sementara EfficientNet mengoptimalkan skala model secara efisien melalui pendekatan *compound scaling*. Perbedaan karakteristik ini memungkinkan stacking untuk memanfaatkan keunggulan masing-masing model dalam menangkap pola pada citra slit-lamp. Output dari masing-masing *base learner*, berupa probabilitas kelas, kemudian digunakan sebagai input bagi *meta-learner*. Meta-learner ini dilatih untuk

mempelajari hubungan antara prediksi model dasar dan label sebenarnya, sehingga mampu memperbaiki kesalahan yang mungkin dilakukan oleh masing-masing model individu. Dengan demikian, stacking tidak hanya meningkatkan akurasi, tetapi juga meningkatkan stabilitas dan generalisasi model terhadap data baru [17].

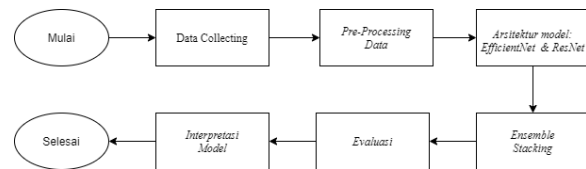
2.6. Grad-CAM

Dalam penerapan model deep learning pada bidang medis, interpretabilitas menjadi aspek yang sangat penting untuk meningkatkan kepercayaan terhadap hasil prediksi. Model berbasis Convolutional Neural Network umumnya bersifat *black box*, sehingga sulit untuk memahami bagaimana model mengambil keputusan. Hal ini menjadi tantangan dalam aplikasi klinis, karena tenaga medis membutuhkan bukti visual yang dapat mendukung hasil diagnosis yang dihasilkan oleh sistem. Untuk mengatasi permasalahan tersebut, penelitian ini menggunakan metode Grad-CAM (*Gradient-weighted Class Activation Mapping*) sebagai teknik interpretasi model. Grad-CAM bekerja dengan memanfaatkan gradien dari kelas target terhadap feature maps pada lapisan konvolusi terakhir untuk menghasilkan peta aktivasi (*heatmap*) [18]. Heatmap ini menunjukkan area pada citra yang memiliki kontribusi paling besar terhadap keputusan model [19].

Dalam implementasinya, Grad-CAM diterapkan pada model ResNet, EfficientNet, serta model hasil ensemble stacking. Hasil visualisasi berupa heatmap kemudian dioverlay pada citra asli slit-lamp untuk mengidentifikasi bagian mata yang menjadi fokus perhatian model. Pada kasus katarak, area yang diharapkan memiliki intensitas tinggi adalah bagian lensa yang mengalami kekeruhan, sehingga dapat dibandingkan dengan interpretasi klinis oleh dokter. Penggunaan Grad-CAM tidak hanya membantu dalam memahami keputusan model, tetapi juga berperan dalam mengevaluasi apakah model telah belajar fitur yang relevan secara medis. Jika model memberikan perhatian pada area yang tidak relevan, maka hal tersebut dapat menjadi indikasi adanya bias atau kesalahan dalam proses pelatihan. Oleh karena itu, interpretasi visual menjadi langkah penting dalam validasi model sebelum diterapkan dalam lingkungan klinis.

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan eksperimental untuk membandingkan performa model deep learning berbasis arsitektur ResNet dan EfficientNet, serta metode ensemble stacking dalam klasifikasi citra slit-lamp mata. Selain itu, penelitian ini juga menerapkan metode Grad-CAM untuk meningkatkan interpretabilitas hasil prediksi model. Tahap penelitian ditampilkan pada Gambar 1.

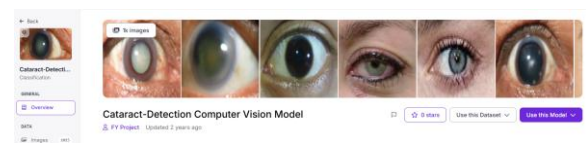


Gambar 1. Alur Penelitian

Gambar 1 menyajikan diagram alur penelitian yang menggambarkan tahapan-tahapan utama yang dilakukan dalam penelitian ini. Uraian lebih lanjut mengenai setiap tahapan dijelaskan secara rinci pada subbagian berikut.

3.1. Sumber Data

Pengumpulan data pada penelitian ini dilakukan dengan memanfaatkan dua platform open-source yang umum digunakan dalam bidang pengolahan citra dan machine learning, yaitu Roboflow dan Kaggle. Kedua platform tersebut dipilih karena menyediakan akses yang luas terhadap dataset citra serta mendukung proses manajemen data seperti anotasi, augmentasi, dan pembagian dataset secara sistematis. Dataset yang digunakan diperoleh dari repositori publik yang relevan dengan studi oftalmologi, khususnya citra mata yang diambil menggunakan perangkat slit-lamp, sehingga sesuai dengan kebutuhan penelitian dalam analisis katarak berbasis citra. Dataset ini bersifat terbuka dan dapat diakses publik, sebagaimana ditunjukkan pada Gambar 2.



Gambar 2. Sumber dataset

3.2. Pra-pemrosesan Data

Tahap berikutnya adalah data *preprocessing*, yaitu proses persiapan data agar siap digunakan dalam pelatihan model. Tahap awal yang dilakukan adalah data cleaning, yakni proses penyaringan dengan menghapus citra yang tidak memenuhi standar kualitas, seperti citra yang buram, terlalu gelap, maupun rusak. Hal ini bertujuan untuk memastikan bahwa hanya data berkualitas baik yang digunakan dalam proses pelatihan, sehingga model dapat belajar secara lebih optimal dan efektif.

Langkah selanjutnya adalah *resize* citra, yaitu mengubah ukuran seluruh gambar retina menjadi seragam, misalnya 224×224 piksel. Tujuannya adalah agar citra dapat diproses oleh CNN tanpa menimbulkan perbedaan dimensi. Normalisasi piksel juga dilakukan untuk menyesuaikan nilai intensitas warna, biasanya dalam rentang 0–1, sehingga perhitungan lebih stabil selama pelatihan model.

Setelah itu, dilakukan augmentasi data untuk memperkaya variasi citra. Teknik augmentasi dapat berupa rotasi, *flipping* horizontal/vertikal, perubahan *brightness*, dan *zooming*. Proses ini bertujuan untuk mengurangi risiko overfitting dengan cara membuat

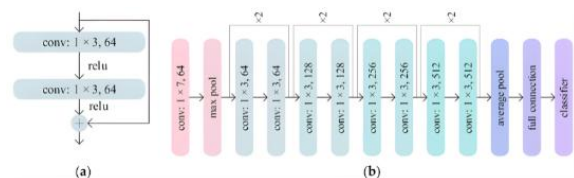
model terbiasa melihat data dalam kondisi yang lebih bervariasi.

Data kemudian dibagi menjadi tiga subset, yaitu *train*, *validation*, dan *test*. Data latih (*train*) digunakan untuk membangun model, data validasi (*validation*) digunakan untuk memantau performa model selama pelatihan, dan data uji (*test*) digunakan untuk mengukur performa akhir model. Proporsi yang umum digunakan adalah 70% untuk *train*, 20% untuk *test*, dan 10% untuk *validation*.

Selain itu, dilakukan juga *class balancing* untuk mengatasi ketidakseimbangan jumlah sampel pada tiap kelas. Misalnya, jika jumlah citra normal lebih banyak dibandingkan citra katarak tingkat lanjut, maka dilakukan oversampling atau augmentasi tambahan pada kelas minoritas agar model tidak bias terhadap kelas mayoritas.

3.3. ResNet-18

Arsitektur ResNet dibangun dari susunan *residual block* yang saling terhubung dalam beberapa tahap (*stage*) sehingga memungkinkan jaringan menjadi sangat dalam tanpa mengalami degradasi performa. Setiap blok residual terdiri atas lapisan konvolusi, normalisasi, fungsi aktivasi, serta mekanisme *skip connection* yang menjadi ciri khas utama ResNet. Gambar 3 berikut menyajikan ilustrasi arsitektur ResNet beserta komponen penyusunnya yang meliputi *stem layer*, *residual block* pada tiap tahap, serta lapisan akhir untuk proses klasifikasi.



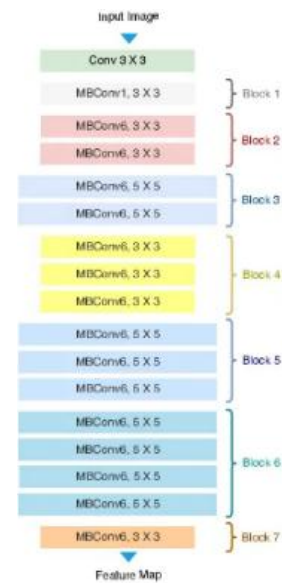
Gambar 3. Model arsitektur ResNet-18

3.4. EfficientNet-B0

Keunggulan utama dari EfficientNetB0 terletak pada metode penskalaan yang terkoordinasi dan dikenal sebagai compound scaling, yang secara proporsional meningkatkan tiga aspek sekaligus, yaitu kedalaman jaringan (*depth*), lebar jaringan (*width*), serta resolusi gambar masukan. Pendekatan ini menghasilkan keseimbangan yang baik antara akurasi tinggi dan efisiensi komputasi, sehingga menjadikan EfficientNetB0 sebagai pilihan yang sangat tepat untuk aplikasi klasifikasi yang menuntut performa tinggi namun tetap hemat dalam penggunaan sumber daya komputasi..

Gambar 4 berikut menyajikan ilustrasi arsitektur EfficientNet. Model ini diawali dengan lapisan konvolusi awal berukuran kernel 3×3 untuk mengekstraksi fitur dasar dari citra masukan, kemudian diikuti oleh rangkaian blok MBConv dengan ukuran kernel bervariasi (3×3 dan 5×5). Setiap blok disusun secara hierarkis dari Block 1 hingga Block 7, dengan jumlah lapisan dan

kompleksitas yang meningkat secara bertahap. Tahap akhir menghasilkan feature map yang kaya akan representasi fitur, yang kemudian digunakan untuk proses klasifikasi lebih lanjut.



Gambar 4. Model arsitektur EfficientNet-B0

3.5. Stacking Ensemble

Metode stacking ensemble pada penelitian ini bekerja melalui dua tingkatan (*level*) pemrosesan yang saling berurutan. Pada Level 0, ResNet18 dan EfficientNet-B0 difungsikan sebagai *base learner* yang dilatih secara independen menggunakan data pelatihan. Untuk menghasilkan prediksi meta yang tidak bias, proses pelatihan pada level ini dilakukan dengan skema *k-fold cross-validation*, di mana pada setiap iterasi lipatan, masing-masing *base learner* dilatih pada sebagian data pelatihan dan menghasilkan prediksi probabilitas pada bagian data yang tidak digunakan (*out-of-fold*). Proses ini diulang sebanyak *k* iterasi hingga seluruh data pelatihan memiliki nilai prediksi *out-of-fold* dari masing-masing model. Hasil prediksi probabilitas dari ResNet18 dan EfficientNet-B0 kemudian digabungkan secara horizontal membentuk matriks fitur meta baru, di mana setiap baris merepresentasikan satu sampel dengan fitur berupa probabilitas prediksi dari kedua *base learner*.

Pada Level 1, matriks fitur meta tersebut digunakan sebagai masukan untuk melatih *meta learner* berupa Logistic Regression. Model ini mempelajari bobot optimal dari kombinasi probabilitas kedua *base learner* terhadap label *ground truth* yang sesungguhnya, sehingga secara adaptif menentukan seberapa besar kontribusi masing-masing *base learner* dalam keputusan akhir. Pada tahap inferensi terhadap data uji, kedua *base learner* yang telah sepenuhnya dilatih pada seluruh data pelatihan menghasilkan prediksi probabilitas secara paralel, lalu probabilitas tersebut digabungkan membentuk matriks fitur meta yang kemudian diumpankan ke Logistic Regression untuk menghasilkan keputusan klasifikasi

final. Dengan mekanisme ini, *meta learner* tidak sekadar merata-ratakan prediksi, melainkan mempelajari pola komplementer dari kedua *base learner* untuk menghasilkan keputusan yang lebih robust dibandingkan apabila masing-masing model digunakan secara terpisah.

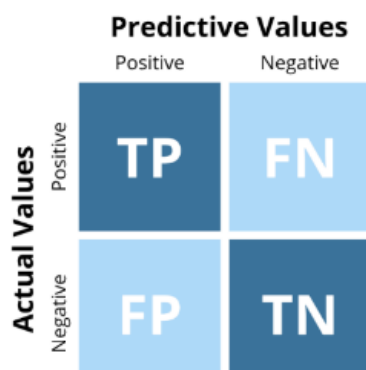
3.6. Evaluasi

Setelah model menghasilkan prediksi, dilakukan tahap evaluasi untuk menilai seberapa baik kinerja sistem. Evaluasi dilakukan menggunakan metrik umum klasifikasi, seperti akurasi untuk mengukur ketepatan prediksi, precision dan recall untuk menilai keseimbangan dalam mendeteksi kasus positif, F1-score untuk mengukur performa secara keseluruhan, serta AUC untuk menilai kualitas pemisahan antar kelas.

Selain metrik yang telah disebutkan sebelumnya, penelitian ini juga memanfaatkan evaluasi confusion matrix yang menyajikan gambaran mengenai jumlah prediksi yang benar maupun yang keliru untuk setiap kelas. Melalui confusion matrix, tidak hanya akurasi model secara keseluruhan yang dapat diketahui, tetapi juga seberapa efektif model dalam mengklasifikasikan kondisi mata, seperti mata normal dan katarak. Confusion matrix merupakan sebuah tabel evaluasi yang digunakan untuk menilai performa model klasifikasi, khususnya dalam pendekatan supervised learning. Tabel ini menggambarkan kemampuan model dalam membedakan berbagai kelas dengan mencatat jumlah prediksi yang tepat dan tidak tepat pada masing-masing kategori. Pada klasifikasi biner, confusion matrix memiliki format 2x2, sedangkan untuk klasifikasi multi-kelas, tabel ini berbentuk $n \times n$ yang menyesuaikan dengan jumlah kelas yang dianalisis.

Tabel 1. Confusion matrix

Actual Class	Predicted Class		Total
	Yes	No	
Yes	TP	FN	P
No	FP	TN	N
Total	P'	N'	P + N



Gambar 5. Confussion Matrix

Confusion matrix adalah representasi dalam bentuk matriks yang digunakan untuk mengevaluasi

kinerja model klasifikasi dengan cara membandingkan hasil prediksi model terhadap nilai yang sebenarnya[20].

Dalam proses evaluasi model klasifikasi, terdapat empat komponen utama yang digunakan untuk mengukur performa model, yaitu True Positive (TP), False Positive (FP), False Negative (FN), dan True Negative (TN). True Positive (TP) mengacu pada jumlah data yang diprediksi positif oleh model dan memang terbukti benar secara aktual. Sebaliknya, False Positive (FP) menggambarkan kondisi di mana data yang sesungguhnya bernilai negatif justru salah diklasifikasikan sebagai positif oleh model. False Negative (FN) merujuk pada situasi di mana data yang seharusnya masuk ke dalam kategori positif malah diprediksi sebagai negatif. Sementara itu, True Negative (TN) merupakan jumlah data yang diprediksi sebagai negatif dan memang benar-benar bernilai negatif secara aktual.

3.7. Interpretasi Hasil

Untuk memberikan transparansi dan keterpercayaan terhadap keputusan klasifikasi yang dihasilkan oleh model stacking ensemble, penelitian ini mengintegrasikan metode *Explainable Artificial Intelligence* (XAI) berupa Gradient-weighted Class Activation Mapping (Grad-CAM). Grad-CAM bekerja dengan memanfaatkan gradien dari skor kelas target yang mengalir kembali ke lapisan konvolusi terakhir masing-masing *base learner*, yaitu ResNet18 dan EfficientNet-B0, untuk menghasilkan peta aktivasi (*activation map*) yang menunjukkan wilayah pada citra masukan yang paling berkontribusi terhadap keputusan prediksi model.

Secara mekanisme, Grad-CAM menghitung rata-rata gradien secara global terhadap setiap kanal pada lapisan konvolusi terakhir untuk memperoleh bobot kepentingan tiap kanal. Bobot tersebut kemudian digunakan untuk menghasilkan kombinasi linier dari peta fitur (*feature map*) pada lapisan tersebut, yang selanjutnya dilewatkan melalui fungsi aktivasi ReLU untuk hanya mempertahankan nilai positif yang berkontribusi terhadap kelas yang diprediksi. Hasil akhirnya berupa peta kalor (*heatmap*) yang divisualisasikan secara bertumpang tindih (*overlay*) di atas citra asli, di mana wilayah berwarna merah atau hangat mengindikasikan area yang paling berpengaruh terhadap keputusan klasifikasi, sementara wilayah berwarna biru atau dingin mengindikasikan area yang kurang relevan.

Visualisasi *heatmap* yang dihasilkan kemudian diinterpretasikan untuk mengevaluasi apakah region yang menjadi fokus model secara visual konsisten dengan area yang secara domain relevan terhadap kelas yang diprediksi. Proses interpretasi ini tidak hanya berfungsi sebagai validasi kualitatif terhadap keandalan model, tetapi juga memberikan wawasan mengenai pola visual yang dipelajari oleh masing-masing arsitektur dalam membentuk keputusan akhir pada sistem stacking ensemble.

4. HASIL DAN PEMBAHASAN

4.1. Sumber Data

Dataset yang digunakan dalam penelitian ini terdiri atas citra mata hasil pemeriksaan slit-lamp, yang merupakan teknik pencitraan oftalmologi untuk mengamati struktur anterior mata, khususnya lensa. Setiap citra dalam dataset telah diklasifikasikan ke dalam tiga kelas utama, yaitu normal, immature, dan mature. Tabel 2 menyajikan distribusi dataset berdasarkan kelas.

Tabel 2. Distribusi dataset

Jumlah Total Dataset	Normal	Immature	Mature
	649	884	888
	2421		

Kelas normal merepresentasikan kondisi lensa mata yang jernih tanpa adanya kekeruhan, sedangkan kelas immature menunjukkan tahap awal perkembangan katarak dengan tingkat kekeruhan yang masih parsial. Adapun kelas mature menggambarkan kondisi katarak lanjut, di mana kekeruhan lensa telah menyeluruh dan signifikan memengaruhi kualitas penglihatan.

4.2. Pra-premrosesan Data

Tahap preprocessing data dilakukan untuk memastikan bahwa data yang digunakan dalam penelitian siap untuk dianalisis. Proses ini bertujuan untuk meningkatkan kualitas data serta meminimalkan kesalahan yang dapat memengaruhi hasil analisis pada tahap selanjutnya.

dilakukan prapemrosesan dengan teknik augmentasi data untuk meningkatkan keragaman dataset. Augmentasi dilakukan dengan menerapkan beberapa transformasi citra, antara lain perubahan ukuran (resize), rotasi dengan sudut tertentu, flipping secara horizontal dan vertikal, penyesuaian tingkat kecerahan (brightness), serta zooming. Transformasi ini bertujuan untuk mensimulasikan variasi kondisi pengambilan citra di dunia nyata, seperti perbedaan sudut pandang, pencahayaan, dan posisi objek, sehingga model dapat belajar lebih robust terhadap variasi tersebut.

Implementasi proses augmentasi data pada penelitian ini dilakukan menggunakan library *PyTorch*, khususnya modul *torchvision.transforms* yang menyediakan berbagai fungsi transformasi citra. Adapun rangkaian transformasi yang digunakan dalam penelitian ini ditunjukkan pada kode program berikut.

```
transform = transforms.Compose([
    transforms.Resize((224, 224)),
    transforms.RandomHorizontalFlip(),
    transforms.RandomRotation(20),
    transforms.ColorJitter(brightness=
0.3, contrast=0.3),
    transforms.RandomAffine(degrees=10
, translate=(0.1, 0.1)),
    transforms.ToTensor(),
    transforms.Normalize([0.485,
0.456, 0.406],
```

4.3. Model Arsitektur

Selama proses pelatihan berlangsung, model menggunakan data latih yang telah melalui proses augmentasi dan secara berkala dievaluasi menggunakan data validasi pada setiap epoch. Pelatihan dijalankan sebanyak 25 epoch, yang berarti keseluruhan data latih diproses sebanyak 25 kali guna menyesuaikan bobot model melalui algoritma backpropagation. Dalam setiap siklus pelatihan, dua indikator utama yang dipantau adalah nilai loss dan akurasi, baik pada data latih maupun data validasi. Nilai loss mencerminkan sejauh mana prediksi model menyimpang dari label yang sebenarnya, sedangkan akurasi menggambarkan proporsi prediksi yang berhasil dilakukan dengan benar. Perbandingan grafik dari kedua metrik tersebut pada data latih dan validasi menjadi hal yang sangat penting untuk mendeteksi adanya indikasi overfitting maupun underfitting pada model.

4.4. ResNet-18

Pada tahap awal eksperimen, dilakukan pelatihan model berbasis arsitektur ResNet sebagai salah satu *baseline model* dalam klasifikasi citra slit-lamp mata. Proses pelatihan ini bertujuan untuk mengevaluasi kemampuan model dalam mengekstraksi fitur visual yang relevan serta membedakan antara kelas yang ditentukan. Selama proses training, performa model dipantau menggunakan metrik seperti akurasi dan loss pada data latih dan validasi untuk memastikan bahwa model mampu belajar dengan baik tanpa mengalami overfitting.

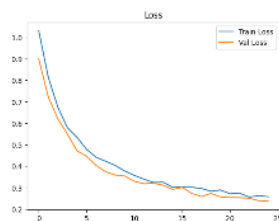
Tabel 3. Hasil pelatihan per-epoch

Epoch	Valid Loss	Valid Accuracy
1	0.3080	0.5847
2	0.7231	0.7624
3	0.6205	0.8306
4	0.5497	0.8223
5	0.4714	0.8843
...	...	...
22	0.2532	0.9070
23	0.2510	0.9112
24	0.2387	0.9215
25	0.2359	0.9215

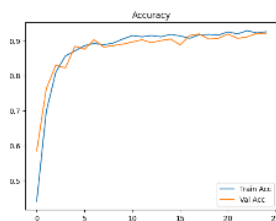
Berdasarkan hasil pelatihan model ResNet yang tersaji pada tabel, dapat diamati bahwa performa model mengalami peningkatan yang cukup signifikan seiring dengan bertambahnya jumlah epoch. Pada tahap awal pelatihan, tepatnya di epoch pertama, nilai validation accuracy masih tercatat sebesar 0,5847 dengan validation loss sebesar 0,3080, yang mengindikasikan bahwa model masih berada dalam tahap awal proses pembelajaran. Seiring berjalannya proses pelatihan, akurasi validasi terus meningkat secara konsisten hingga mencapai angka 0,9215 pada epoch ke-24 dan ke-25, sementara nilai validation loss mengalami penurunan menjadi 0,2359. Pola perkembangan ini menunjukkan bahwa model

semakin mampu melakukan generalisasi yang baik terhadap data validasi.

Meskipun terdapat fluktuasi kecil pada beberapa epoch, seperti peningkatan *validation loss* pada epoch ke-2, secara keseluruhan tren yang dihasilkan menunjukkan konvergensi yang stabil antara penurunan loss dan peningkatan akurasi. Hal ini mengindikasikan bahwa model tidak mengalami overfitting yang signifikan dan mampu mempelajari representasi fitur yang relevan dari citra slit-lamp. Dengan demikian, performa model ResNet dapat dikategorikan baik dalam melakukan klasifikasi, serta layak digunakan sebagai pembandingan terhadap model lain dalam penelitian ini.



Gambar 6. Grafik accuracy dan loss pada train



Gambar 7. Grafik accuracy dan loss pada validation model

Berdasarkan hasil pelatihan model ResNet-18 yang ditunjukkan pada grafik, kinerja model mengalami peningkatan yang konsisten seiring bertambahnya epoch. Nilai *training loss* dan *validation loss* menunjukkan tren penurunan yang stabil, dari kisaran awal di atas 1,0 dan 0,9 hingga mencapai sekitar 0,26 dan 0,24 pada akhir pelatihan. Hal ini mengindikasikan bahwa model mampu mempelajari representasi data dengan baik tanpa mengalami fluktuasi yang signifikan. Selain itu, kurva *training accuracy* dan *validation accuracy* memperlihatkan peningkatan yang cukup cepat pada fase awal, kemudian cenderung melandai dan konvergen di kisaran nilai di atas 0,90. Kedekatan antara kurva pelatihan dan validasi baik pada metrik loss maupun akurasi menunjukkan bahwa model tidak mengalami overfitting secara berarti, sehingga memiliki kemampuan generalisasi yang baik terhadap data validasi. Secara keseluruhan, hasil ini menunjukkan bahwa arsitektur ResNet-18 efektif dalam menangkap pola pada data yang digunakan dalam penelitian ini.

4.5. EfficientNet-B0

Pada tahap selanjutnya, dilakukan pelatihan model berbasis arsitektur EfficientNet untuk dibandingkan dengan performa model sebelumnya. EfficientNet dipilih karena kemampuannya dalam mengoptimalkan kinerja jaringan melalui pendekatan *compound scaling*, yang menyeimbangkan kedalaman, lebar, dan resolusi model secara sistematis. Pendekatan ini memungkinkan model untuk mencapai akurasi yang tinggi dengan jumlah

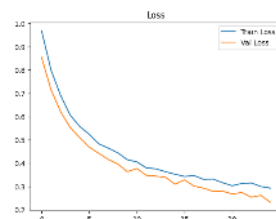
parameter yang relatif lebih efisien dibandingkan arsitektur CNN konvensional.

Selama proses pelatihan, model EfficientNet dilatih menggunakan data yang sama dengan skenario pada model sebelumnya, serta dievaluasi pada setiap epoch menggunakan metrik *validation loss* dan *validation accuracy*. Pemantauan terhadap kedua metrik ini bertujuan untuk mengamati pola pembelajaran model, termasuk kecepatan konvergensi serta stabilitas performa selama proses training. Hasil pelatihan kemudian dianalisis untuk melihat sejauh mana EfficientNet mampu meningkatkan kemampuan klasifikasi dibandingkan dengan model ResNet yang telah diuji sebelumnya.

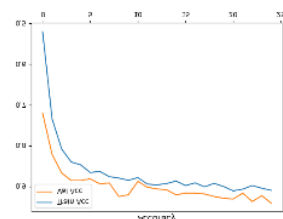
Tabel 4. Hasil pelatihan per-epoch

Epoch	Valid Loss	Valid Accuracy
1	0.8538	0.7190
2	0.7140	0.8202
3	0.6204	0.8657
4	0.5533	0.8843
5	0.5094	0.8843
...	...	...
22	0.2738	0.9153
23	0.2538	0.9360
24	0.2617	0.9215
25	0.2323	0.9401

Berdasarkan hasil pelatihan model EfficientNet yang ditunjukkan pada Tabel, terlihat bahwa model menunjukkan peningkatan performa yang cukup signifikan seiring bertambahnya epoch. Pada epoch pertama, nilai *validation accuracy* sudah mencapai 0,7190 dengan *validation loss* sebesar 0,8538, yang menunjukkan bahwa model memiliki kemampuan awal yang cukup baik dalam mengenali pola pada citra. Seiring dengan proses pelatihan, akurasi validasi meningkat secara konsisten hingga mencapai nilai tertinggi sebesar 0,9401 pada epoch ke-25, sementara *validation loss* menurun menjadi 0,2323.



Gambar 8. Grafik accuracy dan loss pada train



Gambar 9. Grafik accuracy dan loss pada validation

Pada Gambar 8 dan 9 Berdasarkan hasil pelatihan model EfficientNet hingga epoch ke-25, diperoleh performa yang menunjukkan konvergensi yang baik. Nilai *training loss* sebesar 0,2928 dan *validation loss* sebesar 0,2323 memperlihatkan tren penurunan yang stabil selama proses pelatihan, yang mengindikasikan bahwa model mampu meminimalkan kesalahan prediksi secara efektif. Selain itu, nilai *training accuracy* mencapai 0,9085, sedangkan *validation*

accuracy sebesar 0,9401, yang menunjukkan tingkat akurasi klasifikasi yang tinggi. Menariknya, nilai akurasi pada data validasi sedikit lebih tinggi dibandingkan data pelatihan, yang mengindikasikan bahwa model tidak mengalami overfitting dan memiliki kemampuan generalisasi yang baik. Secara keseluruhan, hasil ini menunjukkan bahwa arsitektur EfficientNet efektif dalam mengekstraksi fitur dan menangkap pola kompleks pada data, sehingga layak digunakan untuk tahap evaluasi lanjutan maupun implementasi pada kasus nyata.

4.6. Stacking Ensemble

Setelah diperoleh hasil prediksi dari model ResNet dan EfficientNet-B0, tahap selanjutnya dalam penelitian ini adalah menggabungkan kedua model tersebut menggunakan pendekatan ensemble stacking. Pada metode ini, output prediksi dari masing-masing model dasar (base learner) tidak langsung dijadikan hasil akhir, melainkan digunakan sebagai fitur masukan bagi model tingkat kedua (meta learner). Pendekatan ini bertujuan untuk memanfaatkan keunggulan masing-masing model sehingga dapat meningkatkan performa klasifikasi secara keseluruhan. Meta-learner yang digunakan dalam penelitian ini adalah Logistic Regression, yang bekerja dengan cara memodelkan hubungan antara fitur masukan dan probabilitas suatu kelas menggunakan fungsi sigmoid sebagai dasar perhitungannya. Output probabilitas dari ResNet dan EfficientNet dijadikan sebagai variabel input, kemudian Logistic Regression akan mempelajari bobot optimal untuk mengkombinasikan kedua prediksi tersebut. Secara mekanisme, model ini akan memberikan nilai probabilitas untuk setiap kelas berdasarkan kombinasi linear dari input, yang kemudian dipetakan ke dalam rentang 0 hingga 1 menggunakan fungsi aktivasi sigmoid. Kelas dengan probabilitas tertinggi akan dipilih sebagai hasil prediksi akhir.

Penggunaan Logistic Regression sebagai meta-learner dipilih karena sifatnya yang sederhana, cepat, serta efektif dalam menggabungkan prediksi dari beberapa model tanpa menambah kompleksitas yang signifikan. Selain itu, model ini cenderung lebih stabil dan tidak mudah mengalami overfitting dibandingkan model yang lebih kompleks, terutama ketika jumlah fitur input relatif kecil seperti pada kasus stacking ini. Pada tahap ensemble stacking, mekanisme meta-learner Logistic Regression diimplementasikan dengan memanfaatkan output probabilitas dari model ResNet dan EfficientNet-B0 sebagai fitur input. Secara teknis dalam kode, prediksi dari masing-masing model diambil menggunakan hasil forward pass yang kemudian dikonversi menjadi probabilitas (softmax). Output tersebut selanjutnya digabungkan (concatenate) menjadi satu matriks fitur baru (X_{meta}), di mana setiap baris merepresentasikan satu sampel dengan kombinasi probabilitas dari kedua model. Label asli tetap digunakan sebagai target (y_{meta}). Setelah itu, model Logistic Regression

dilatih menggunakan fungsi $\text{fit}(X_{meta}, y_{meta})$ untuk mempelajari bobot optimal dari masing-masing prediksi model dasar. Pada tahap inferensi, digunakan fungsi $\text{predict}()$ untuk menghasilkan kelas akhir berdasarkan kombinasi linear dari probabilitas yang telah dipelajari. Setelah proses pelatihan dan pengujian dilakukan, langkah selanjutnya adalah membandingkan hasil akurasi dari setiap model untuk melihat efektivitas pendekatan yang digunakan.

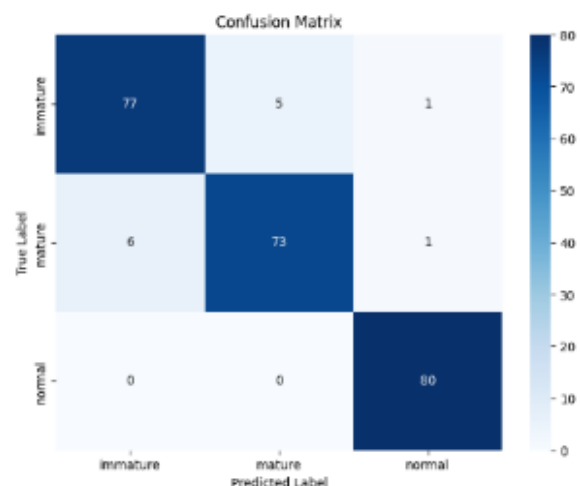
Tabel 5. Hasil Perbandingan Performa Setiap Model

Model	Akurasi
ResNet18	0.9215
EfficientNetB0	0.9401
Ensemble Stacking	0.9837

Dari tabel di atas, terlihat bahwa metode Ensemble Stacking memberikan performa terbaik dengan akurasi 0.9837 (98.37%), jauh lebih tinggi dibandingkan model tunggal seperti ResNet18 (92.15%) dan EfficientNetB0 (94.01%). Hal ini menunjukkan bahwa kombinasi beberapa model mampu menangkap pola data dengan lebih baik dibandingkan model individual.

4.7. Evaluasi

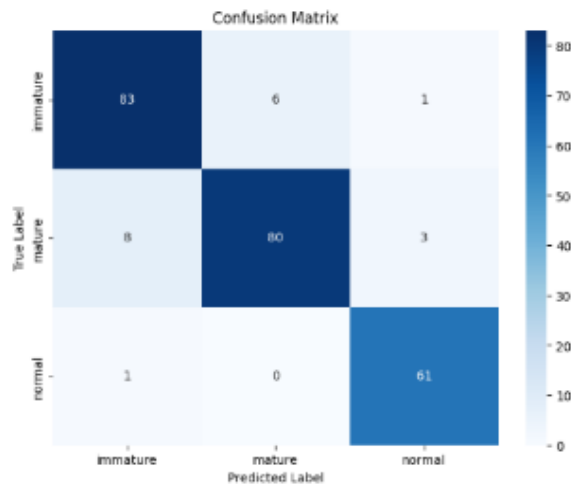
Evaluasi model dilakukan menggunakan data uji yang belum pernah dilihat oleh model sebelumnya. Evaluasi dengan menampilkan *confusion matrix* untuk melihat distribusi hasil klasifikasi secara rinci (*true positive*, *false positive*, *true negative*, *false negative*). Confusion matrix ini digunakan untuk menilai kemampuan model dalam mengklasifikasikan setiap kelas secara tepat serta mengidentifikasi jenis kesalahan klasifikasi yang terjadi.



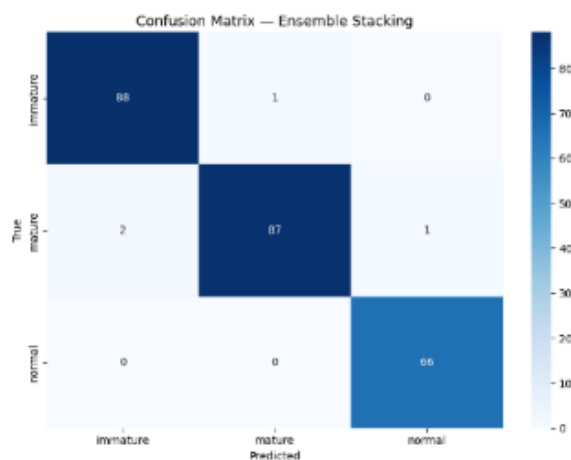
Gambar 10. *Confusion Matrix* hasil testing model EfficientNet

Model EfficientNet-B0 menunjukkan performa klasifikasi yang baik pada ketiga kelas. Kelas *immature* berhasil memprediksi 77 dari 83 sampel dengan benar, sementara 5 sampel salah terklasifikasi sebagai *mature* dan 1 sampel sebagai *normal*. Kelas

mature memiliki 73 prediksi benar dari 80 sampel, dengan 6 sampel salah terklasifikasi sebagai *immature* dan 1 sampel sebagai *normal*. Kelas *normal* mencapai hasil sempurna dengan seluruh 80 sampel terprediksi benar tanpa kesalahan. Secara keseluruhan, model ini bisa memprediksi 230 benar dari 243 sampel, dengan kebingungan utama terjadi antara kelas *immature* dan *mature* yang memiliki kemiripan fitur visual.



Gambar 11. Confusion Matrix hasil testing model ResNet



Gambar 12. Confusion Matrix hasil testing model Ensemble Stacking

Model ResNet18 menunjukkan performa yang sebanding dengan EfficientNet-B0. Kelas *immature* memprediksi 83 dari 90 sampel dengan benar, dengan 6 sampel salah terklasifikasi sebagai *mature* dan 1 sampel sebagai *normal*. Kelas *mature* memiliki 80 prediksi benar dari 91 sampel, dengan 8 sampel salah terklasifikasi sebagai *immature* dan 3 sampel sebagai *normal*, menjadikannya kelas dengan tingkat kesalahan tertinggi. Kelas *normal* hampir sempurna dengan 61 dari 62 sampel terprediksi benar, dan hanya 1 sampel salah terklasifikasi sebagai *immature*. Secara keseluruhan, model dapat memprediksi 224 benar dari 243 sampel, dengan pola kebingungan yang serupa antara kelas *immature* dan *mature*.

Berdasarkan hasil evaluasi model *ensemble stacking* yang divisualisasikan melalui *confusion matrix* pada Gambar 12, model menunjukkan performa klasifikasi yang sangat baik pada ketiga kelas yang diuji, yaitu *immature*, *mature*, dan *normal*. Pada kelas *immature*, model berhasil mengklasifikasikan 88 sampel secara benar (True Positive), dengan hanya 1 sampel yang salah diklasifikasikan sebagai kelas *mature* dan tidak ada sampel yang diklasifikasikan sebagai kelas *normal*. Pada kelas *mature*, sebanyak 87 sampel berhasil diklasifikasikan dengan tepat, sementara terdapat 2 sampel yang keliru diklasifikasikan sebagai kelas *immature* dan 1 sampel yang keliru diklasifikasikan sebagai kelas *normal*. Adapun pada kelas *normal*, seluruh 66 sampel berhasil diklasifikasikan dengan benar tanpa adanya kesalahan klasifikasi ke kelas lain. Secara keseluruhan, hasil ini mengindikasikan bahwa model *ensemble stacking* memiliki kemampuan generalisasi yang tinggi dengan jumlah *misclassification* yang sangat minimal, yakni hanya sebesar 4 sampel dari total 245 sampel uji, sehingga menghasilkan akurasi keseluruhan sebesar 98,37%. Rendahnya tingkat kesalahan klasifikasi pada setiap kelas mencerminkan bahwa model mampu mengenali pola fitur dari masing-masing kategori secara efektif dan konsisten.

Classification Report (Ensemble Stacking):				
	precision	recall	f1-score	support
immature	0.98	0.99	0.98	89
mature	0.99	0.97	0.98	90
normal	0.99	1.00	0.99	66
accuracy			0.98	245
macro avg	0.98	0.99	0.98	245
weighted avg	0.98	0.98	0.98	245

Gambar 13. Matrix Evaluasi Stacking Ensemble.

Berdasarkan hasil evaluasi model stacking ensemble pada data uji yang terdiri dari 245 sampel, model menunjukkan performa klasifikasi yang sangat baik dengan akurasi keseluruhan sebesar 98%. Pada kelas *immature*, model mencapai nilai *precision* sebesar 0,98, *recall* sebesar 0,99, dan *f1-score* sebesar 0,98 dari total 89 sampel, yang mengindikasikan bahwa model mampu mengidentifikasi kelas ini dengan sangat tepat dan hampir tidak melewatkan sampel yang seharusnya terdeteksi. Pada kelas *mature* dengan 90 sampel, model memperoleh nilai *precision* sebesar 0,99, *recall* sebesar 0,97, dan *f1-score* sebesar 0,98, menunjukkan bahwa meskipun terdapat sedikit sampel *mature* yang tidak terdeteksi dengan benar, tingkat presisi model terhadap kelas ini tetap sangat tinggi. Sementara itu, kelas *normal* mencatatkan performa terbaik di antara ketiga kelas dengan nilai *precision* sebesar 0,99, *recall* sempurna sebesar 1,00, dan *f1-score* sebesar 0,99 dari 66 sampel, yang berarti seluruh sampel kelas *normal* berhasil diklasifikasikan dengan benar tanpa ada yang terlewat.

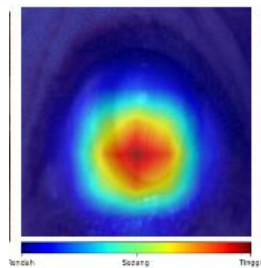
Secara keseluruhan, nilai *macro average* dan *weighted average* untuk ketiga metrik evaluasi masing-masing berada pada angka 0,98, yang mencerminkan konsistensi performa model di seluruh kelas meskipun terdapat ketidakseimbangan jumlah sampel antar kelas. Hasil ini menunjukkan bahwa pendekatan *stacking ensemble* yang mengombinasikan ResNet18 dan EfficientNet-B0 dengan *meta learner* Logistic Regression berhasil memanfaatkan kekuatan komplementer kedua *base learner* secara efektif, sehingga menghasilkan model klasifikasi yang robust dan generalisasi tinggi terhadap data yang belum pernah dilihat sebelumnya.

4.8. Interpretasi Hasil

Untuk memperoleh pemahaman yang lebih mendalam mengenai proses pengambilan keputusan model dalam melakukan klasifikasi, dilakukan visualisasi menggunakan metode *Gradient-weighted Class Activation Mapping* (Grad-CAM). Teknik ini memungkinkan identifikasi region atau area pada citra yang paling berpengaruh terhadap prediksi yang dihasilkan oleh model, sehingga meningkatkan interpretabilitas dan transparansi sistem klasifikasi yang dibangun.



Gambar 14. Data yang diuji



Gambar 15. Hasil Visualisasi Grad-CAM

Berdasarkan hasil visualisasi Grad-CAM yang ditampilkan pada Gambar 14, terlihat bahwa model secara konsisten memusatkan perhatiannya (*attention*) pada area lensa mata, khususnya bagian sentral kornea yang menampakkan kekeruhan (*opacity*). Peta aktivasi yang dihasilkan menunjukkan gradasi warna dari biru (aktivasi rendah) hingga merah (aktivasi tinggi), di mana area berwarna merah dan oranye yang terkonsentrasi di tengah lensa mengindikasikan bahwa model memberikan bobot perhatian tertinggi pada region tersebut. Hal ini sejalan secara klinis, mengingat kekeruhan pada bagian sentral lensa merupakan karakteristik visual utama yang membedakan kondisi katarak *immature*, *mature*, maupun *normal*.

5. KESIMPULAN DAN SARAN

Penelitian ini bertujuan untuk membandingkan performa model *deep learning* berbasis arsitektur ResNet-18 dan EfficientNet-B0 sebagai *single model*, serta pendekatan *ensemble stacking* dalam klasifikasi citra *slit-lamp* mata ke dalam tiga kelas, yaitu *normal*, *immature*, dan *mature*, sekaligus mengintegrasikan

metode Grad-CAM untuk meningkatkan interpretabilitas model secara visual. Berdasarkan hasil eksperimen dan evaluasi yang telah dilakukan, dapat disimpulkan bahwa model *ensemble stacking* yang menggabungkan ResNet-18 dan EfficientNet-B0 dengan *meta-learner* Logistic Regression menghasilkan performa klasifikasi terbaik dengan akurasi sebesar 98,37%, melampaui performa model tunggal ResNet-18 (92,15%) maupun EfficientNet-B0 (94,01%). Peningkatan performa ini menunjukkan bahwa pendekatan *stacking* mampu memanfaatkan keunggulan masing-masing arsitektur secara komplementer, di mana ResNet-18 unggul dalam menangkap fitur melalui mekanisme *residual connection*, sementara EfficientNet-B0 mengoptimalkan representasi fitur melalui pendekatan *compound scaling*.

Hasil evaluasi *confusion matrix* pada model *ensemble stacking* memperlihatkan bahwa hanya terdapat 4 kesalahan klasifikasi dari total 245 sampel uji, dengan kebingungan klasifikasi utama terjadi antara kelas *immature* dan *mature* pada ketiga model, yang mengindikasikan adanya kemiripan fitur visual yang tinggi di antara kedua kelas tersebut. Selain itu, visualisasi Grad-CAM yang diintegrasikan dalam penelitian ini membuktikan bahwa model tidak hanya menghasilkan prediksi yang akurat, tetapi juga berfokus pada area anatomis yang relevan secara klinis, yaitu bagian sentral lensa mata yang mengalami kekeruhan, sehingga memperkuat validitas dan transparansi keputusan model. Secara keseluruhan, penelitian ini memberikan kontribusi berupa sistem klasifikasi katarak otomatis berbasis *deep learning* yang akurat dan interpretatif, dengan pendekatan *ensemble stacking* yang terbukti efektif dalam meningkatkan performa dibandingkan arsitektur tunggal.

Berdasarkan hasil penelitian yang telah dilakukan, beberapa saran dapat dikemukakan untuk pengembangan penelitian selanjutnya. Pertama, perluasan dataset perlu dilakukan dengan mengumpulkan citra *slit-lamp* dari berbagai sumber institusi medis yang berbeda, termasuk penambahan kelas katarak pada stadium *hypermature* agar model dapat mengenali spektrum perkembangan katarak yang lebih lengkap dan representatif secara klinis. Kedua, eksplorasi arsitektur *base learner* yang lebih beragam, seperti Vision Transformer (ViT) atau DenseNet, dapat dipertimbangkan untuk menggantikan atau melengkapi ResNet-18 dan EfficientNet-B0 dalam kerangka *stacking ensemble*, sehingga potensi komplementaritas antar model dapat lebih dimaksimalkan. Ketiga, mengingat masih terdapat kebingungan klasifikasi antara kelas *immature* dan *mature* pada ketiga model yang diuji, penelitian lanjutan dapat mengeksplorasi teknik peningkatan fitur yang lebih spesifik, seperti penerapan *attention mechanism* atau *fine-grained feature extraction*, untuk mempertajam kemampuan model dalam membedakan kedua kelas tersebut.

DAFTAR PUSTAKA

- [1] W. Health Organization, "World report on vision," World Health Organization, Geneva, 2019.
- [2] Kementerian Kesehatan Republik Indonesia, "Profil Kesehatan Indonesia 2022," Kementerian Kesehatan RI, Jakarta, 2022.
- [3] A. Mardjuki, A. Larasandi, D. H. Silaen, and R. A. Ramadani, "Improving Knowledge and Behavior through Cataract Education in the Pre-Prosperous Community of Tangerang," *Jurnal Abmas Negeri (JAGRI)*, vol. 6, no. 1, pp. 101–109, Jun. 2025, doi: 10.36590/jagri.v6i1.1525.
- [4] M. Khairallah *et al.*, "Number of people blind or visually impaired by cataract worldwide and in world regions, 1990 to 2010," *Invest. Ophthalmol. Vis. Sci.*, vol. 56, no. 11, pp. 6762–6769, Oct. 2015, doi: 10.1167/iovs.15-17201.
- [5] R. Klein and B. E. K. Klein, "Epidemiology of Ocular Functions and Diseases in Persons With Diabetes," 2018, *National Institute of Diabetes and Digestive and Kidney Diseases (US), Bethesda (MD)*. [Online]. Available: <http://europepmc.org/books/NBK567971>
- [6] C. Liu, Y. Zhang, X. Cao, and Z. Chen, "Association between metabolic syndrome and cataract: a meta-analysis," *Eye (Basingstoke)*, vol. 39, no. 13, pp. 2555–2564, Sep. 2025, doi: 10.1038/s41433-025-03910-2.
- [7] G. Kourounis, A. A. Elmahmudi, B. Thomson, J. Hunter, H. Ugail, and C. Wilson, "Computer image analysis with artificial intelligence: a practical introduction to convolutional neural networks for medical professionals," *Postgrad. Med. J.*, vol. 99, no. 1178, pp. 1287–1294, Dec. 2023, doi: 10.1093/postmj/qgad095.
- [8] Y. K. Wong, "Understanding the Features of Deep Learning," *International Journal of Information Technology (IJIT)*, vol. 7, no. 5, pp. 1–6, Sep. 2021.
- [9] J. Li *et al.*, "Automatic Cataract Diagnosis by Image-Based Interpretability," in *Proceedings - 2018 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2018*, Institute of Electrical and Electronics Engineers Inc., Jul. 2018, pp. 3964–3969. doi: 10.1109/SMC.2018.00672.
- [10] M. F. Fayyad and Mustakim, "Application of AlexNet, EfficientNetV2B0, and VGG19 with Explainable AI for Cataract and Glaucoma Image Classification," in *2024 International Electronics Symposium: Shaping the Future: Society 5.0 and Beyond, IES 2024 - Proceeding*, Institute of Electrical and Electronics Engineers Inc., 2024, pp. 406–412. doi: 10.1109/IES63037.2024.10665856.
- [11] S. S. Aishwarya, C. N. Arpitha, C. Gopalkrishna, S. B. Kurtakoti, and G. Chandra Naik, "Cataract Early Detection Using ResNet150 for Eye Fundus Images," in *Proceedings of 3rd IEEE International Conference on Knowledge Engineering and Communication Systems, ICKECS 2025*, Institute of Electrical and Electronics Engineers Inc., 2025. doi: 10.1109/ICKECS65700.2025.11035694.
- [12] J. D. Steinmetz *et al.*, "Causes of blindness and vision impairment in 2020 and trends over 30 years, and prevalence of avoidable blindness in relation to VISION 2020: The Right to Sight: An analysis for the Global Burden of Disease Study," *Lancet Glob. Health*, vol. 9, no. 2, pp. e144–e160, Feb. 2021, doi: 10.1016/S2214-109X(20)30489-7.
- [13] J. K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, 2016, pp. 770–778.
- [14] K. He, X. Zhang, S. Ren, and J. Sun, "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 1026–1034.
- [15] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," in *Proceedings of the 36th International Conference on Machine Learning (ICML)*, vol. 97, 2019, pp. 6105–6114.
- [16] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 4510–4520.
- [17] E. R. Putri *et al.*, "KLASIFIKASI PERULANGAN KANKER TIROID MENGGUNAKAN STACK ENSEMBLE DAN SMOTE," 2025.
- [18] X. Li, W. Lv, I. U. Khan, B. Xie, and R. Zhu, "Explainable Electricity Theft Detection With Gradient-Weighted Class Activation Mapping," *Electron. Lett.*, vol. 61, no. 1, Jan. 2025, doi: 10.1049/ell2.70264.
- [19] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization," Dec. 2019, doi: 10.1007/s11263-019-01228-7.
- [20] J. Görtler *et al.*, "Neo: Generalizing Confusion Matrix Visualization to Hierarchical and Multi-Output Labels," in *Conference on Human Factors in Computing Systems - Proceedings*, Association for Computing Machinery, Apr. 2022. doi: 10.1145/3491102.3501823.