

BAB IV

HASIL DAN PEMBAHASAN

Bab ini membahas lingkungan dan perangkat pengujian yang digunakan dalam penelitian ini, serta hasil dan pembahasan yang telah diperoleh dari penelitian skripsi ini.

4.1 Perangkat Pengujian

Spesifikasi perangkat yang digunakan dalam melakukan penelitian ini adalah sebagai berikut.

1. CPU : AMD Ryzen 5 5600H with Radeon Graphics
2. GPU : NVIDIA GeForce RTX 3060
3. RAM : 8 GB
4. Memori : 512 GB SSD
5. Sistem Operasi : Windows 11 Home Single Language 64-bit

Spesifikasi perangkat diatas digunakan untuk melakukan proses pelatihan model menggunakan platform *Google Colab* dengan bahasa pemrograman *Python* serta *framework TensorFlow* dan *Keras*. *Library* tambahan yang digunakan meliputi *Albumentations* untuk augmentasi data, *Matplotlib* untuk visualisasi, dan *Scikit-learn* untuk evaluasi performa model.

4.2 Skenario Pelatihan Model

Penelitian ini menggunakan model *EfficientNet-B0* dan *ResNet50* yang dievaluasi menggunakan berbagai kombinasi *hyperparameter* pada empat skenario dataset, yaitu data original, data augmentasi, data dengan *class weight*, serta kombinasi antara data augmentasi dan *class weight*. *Hyperparameter* yang digunakan meliputi tiga variasi *learning rate*, yaitu 0,001; 0,0001; dan 0,00001, dua variasi *dropout* sebesar 0,3 dan 0,5, serta dua variasi *batch size*, yaitu 16 dan 32. Dengan demikian, pada setiap skenario dataset dihasilkan 12 kombinasi percobaan pelatihan.

Proses pelatihan dilakukan selama 50 *epoch* dengan menerapkan mekanisme *early stopping* apabila nilai *validation accuracy* tidak mengalami peningkatan selama lima *epoch* berturut-turut. Penerapan *early stopping* ini bertujuan untuk mencegah terjadinya *overfitting*

serta meningkatkan efisiensi waktu komputasi. Dengan mempertimbangkan empat skenario dataset dan dua arsitektur model yang digunakan, maka total eksperimen dalam penelitian ini berjumlah 96 percobaan pelatihan.

Skenario pertama menggunakan data original, yaitu data citra MRI tanpa penerapan teknik tambahan seperti augmentasi. Pada skenario ini, model hanya mempelajari pola yang terdapat pada data asli. Tujuan dari skenario ini adalah untuk memperoleh gambaran performa dasar model sebelum dilakukan penerapan teknik peningkatan performa.

Skenario kedua menerapkan teknik data augmentasi pada data pelatihan untuk meningkatkan variasi citra. Proses augmentasi dilakukan melalui beberapa transformasi, seperti rotasi, *zoom*, serta penyesuaian *contrast* dan *brightness*. Penerapan teknik ini bertujuan untuk meningkatkan kemampuan generalisasi model serta mengurangi risiko *overfitting*.

Skenario ketiga menggunakan teknik *class weight* untuk mengatasi ketidakseimbangan jumlah data antar kelas. Pada pendekatan ini, kelas dengan jumlah data yang lebih sedikit diberikan bobot yang lebih besar, sehingga kesalahan prediksi pada kelas minoritas memberikan kontribusi *loss* yang lebih signifikan. Dengan demikian, model diharapkan mampu mempelajari pola dari setiap kelas secara lebih seimbang serta mengurangi bias terhadap kelas mayoritas.

Skenario keempat merupakan kombinasi antara data augmentasi dan *class weight*. Kombinasi ini bertujuan untuk meningkatkan variasi data sekaligus mengatasi permasalahan ketidakseimbangan kelas. Dengan penerapan kedua teknik tersebut, diharapkan model dapat menghasilkan performa yang lebih optimal, baik dari segi *accuracy* maupun kemampuan generalisasi.

Konfigurasi *hyperparameter* yang digunakan pada masing-masing skenario pelatihan disajikan pada Tabel 4.1.

Tabel 4. 1 Tabel *Hyperparameter* untuk Pelatihan

No	<i>Learning rate</i>	<i>Batch size</i>	<i>dropout</i>
1.	0,001	16	0,3
2.	0,001	16	0,5

No	<i>Learning rate</i>	<i>Batch size</i>	<i>dropout</i>
3.	0,0001	16	0,3
4.	0,0001	16	0,5
5.	0,00001	16	0,3
6.	0,00001	16	0,5
7.	0,001	32	0,3
8.	0,001	32	0,5
9.	0,0001	32	0,3
10.	0,0001	32	0,5
11.	0,00001	32	0,3
12.	0,00001	32	0,5

4.3 Hasil Pelatihan Model

Proses pelatihan model pada penelitian ini dilakukan menggunakan dua arsitektur *Convolutional Neural Network* (CNN), yaitu EfficientNet-B0 dan ResNet50, untuk mengklasifikasikan citra MRI otak ke dalam empat kelas stadium Alzheimer. Evaluasi kinerja model dilakukan menggunakan metrik *accuracy*, *precision*, dan *recall* yang diperoleh dari *confusion matrix* pada data pengujian. Selain itu, proses pelatihan juga dianalisis melalui grafik *training* dan *validation accuracy* serta *training* dan *validation loss* pada setiap *epoch*.

Evaluasi ini bertujuan untuk mengetahui kemampuan model dalam mempelajari pola citra MRI, serta menganalisis pengaruh penerapan data augmentation dan class weight terhadap peningkatan performa klasifikasi. Hasil dari seluruh skenario pelatihan kemudian digunakan untuk menentukan konfigurasi model terbaik yang akan diinterpretasikan menggunakan metode Grad-CAM.

4.3.1 EfficientNet-B0

Pelatihan model EfficientNet-B0 dilakukan selama maksimal 50 *epoch* dengan menerapkan mekanisme *early stopping* apabila nilai *validation accuracy* tidak mengalami peningkatan selama lima *epoch* berturut-turut. Selama proses pelatihan, dilakukan pemantauan terhadap metrik *accuracy* dan *loss* pada data latih dan data validasi untuk mengevaluasi kinerja model. Hasil pelatihan disajikan dalam bentuk grafik *loss* dan

accuracy, *confusion matrix*, serta nilai evaluasi berupa *accuracy*, *precision*, *recall*. Ringkasan hasil terbaik dari masing-masing skenario ditampilkan pada Tabel 4.2.

Tabel 4. 2 Tabel Hasil Pelatihan Terbaik Masing-Masing Skenario Model EfficientNet-B0

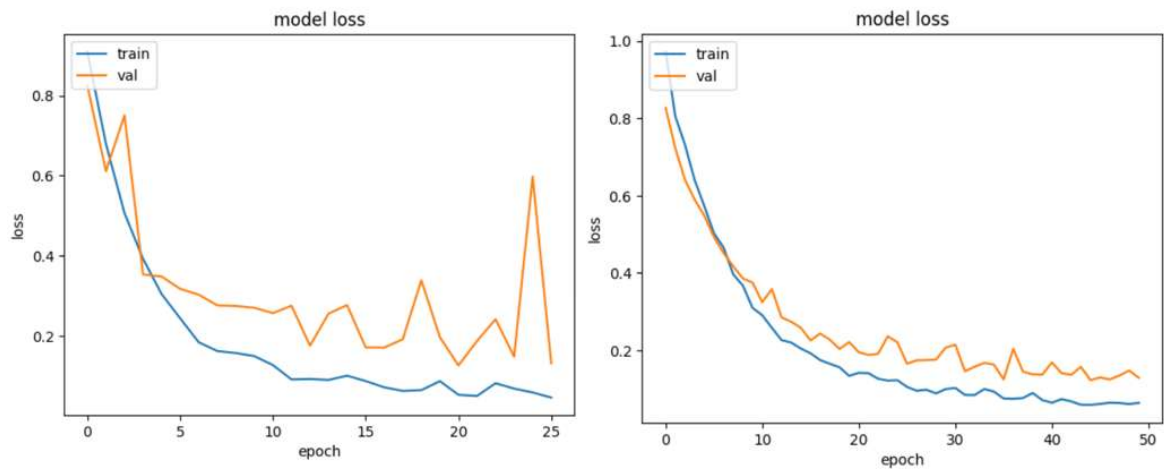
Nama Skenario	<i>Learning Rate</i>	<i>Dropout</i>	<i>Batch Size</i>	<i>Epoch</i>	<i>Validation Accuracy</i>	<i>Accuracy</i>	<i>Recall</i>
Data Original	0,001	0,3	32	21	0,962	0,96	0,96
Data dengan Augmentasi	0,0001	0,3	16	47	0,965	0,95	0,95
Data dengan Class Weight	0,001	0,3	16	21	0,94	0,95	0,95
Data dengan Augmentasi dan Class Weight	0,0001	0,3	16	36	0,945	0,947	0,93

Hasil pelatihan skenario data augmentasi pada Tabel 4.2 menunjukkan performa terbaik dari sisi kemampuan generalisasi model dengan nilai *validation accuracy* sebesar 0,965. Hasil ini diperoleh menggunakan kombinasi *learning rate* 0,0001, *dropout* 0,3, *batch size* 16, dan *epoch* optimal pada iterasi ke-47. Model pada skenario ini juga menghasilkan *accuracy* dan *recall* sebesar 0,95.

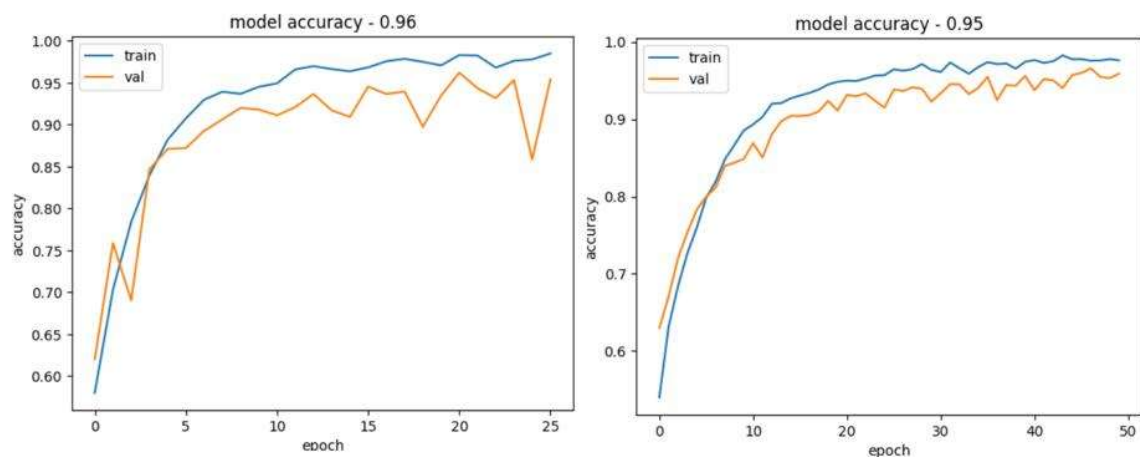
Skenario data original menunjukkan performa yang tidak jauh berbeda dengan *validation accuracy* sebesar 0,962 serta *accuracy* dan *recall* sebesar 0,96. Meskipun nilai ini lebih tinggi secara numerik, terdapat indikasi *overfitting* karena model dilatih tanpa variasi data tambahan. Hal ini terlihat dari perbedaan pola antara data latih dan validasi selama proses pelatihan.

Analisis grafik *loss* menunjukkan bahwa pada data original terdapat fluktuasi yang cukup tinggi, terutama pada *validation loss* setelah pertengahan *epoch*, serta jarak yang semakin melebar antara *training loss* dan *validation loss*. Kondisi ini mengindikasikan bahwa model cenderung menghafal data latih dan kurang mampu melakukan generalisasi. Sebaliknya, pada skenario data augmentasi, grafik *loss* menunjukkan pola yang lebih stabil dengan fluktuasi yang lebih kecil dan jarak antar kurva yang relatif konsisten, menandakan

proses pembelajaran yang lebih baik. Gambar perbandingan grafik *loss* dapat dilihat pada Gambar 4.1.



Gambar 4. 1 Grafik Loss Model EfficientNet-B0 Data Original (Kiri) Data Augmentasi (Kanan)

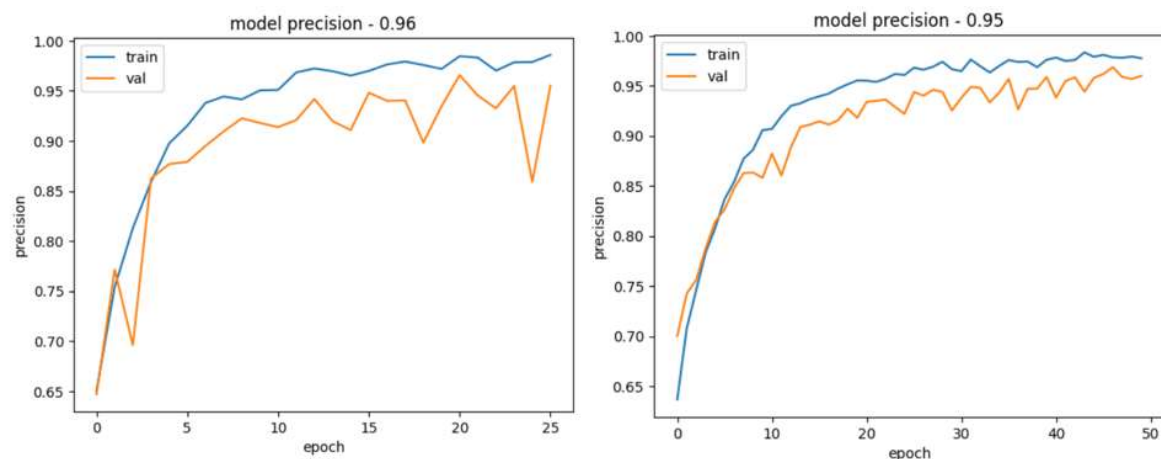


Gambar 4. 2 Grafik *Accuracy* Model EfficientNet-B0 Data Original (Kiri) Data Augmentasi (Kanan)

Hasil yang serupa juga terlihat pada grafik *accuracy*. Pada data original, *validation accuracy* menunjukkan fluktuasi yang cukup signifikan dan tidak konsisten, meskipun mencapai nilai maksimum yang tinggi. Sementara itu, pada data augmentasi, kurva *accuracy* menunjukkan pola yang lebih stabil dan konvergen, dengan jarak yang lebih kecil antara *training* dan *validation accuracy*, sehingga mencerminkan kemampuan generalisasi yang lebih baik. Grafik *loss* dapat dilihat pada Gambar 4.2

Analisis pada grafik *precision* menunjukkan bahwa pada skenario data original, meskipun nilai maksimum mencapai 0,96, kurva validation precision mengalami fluktuasi

yang cukup tinggi, terutama pada *epoch* akhir. Selain itu, terdapat perbedaan yang cukup besar antara kurva *training* dan *validation precision*, yang mengindikasikan terjadinya *overfitting* serta peningkatan risiko kesalahan prediksi positif (*false positive*). Sebaliknya, pada skenario data augmentasi, nilai *precision* sedikit lebih rendah, yaitu sekitar 0,95, namun kurva menunjukkan pola yang lebih stabil dan konsisten. Kedekatan antara kurva *training* dan *validation precision* mencerminkan kemampuan model yang lebih baik dalam menghasilkan prediksi positif yang akurat secara konsisten.

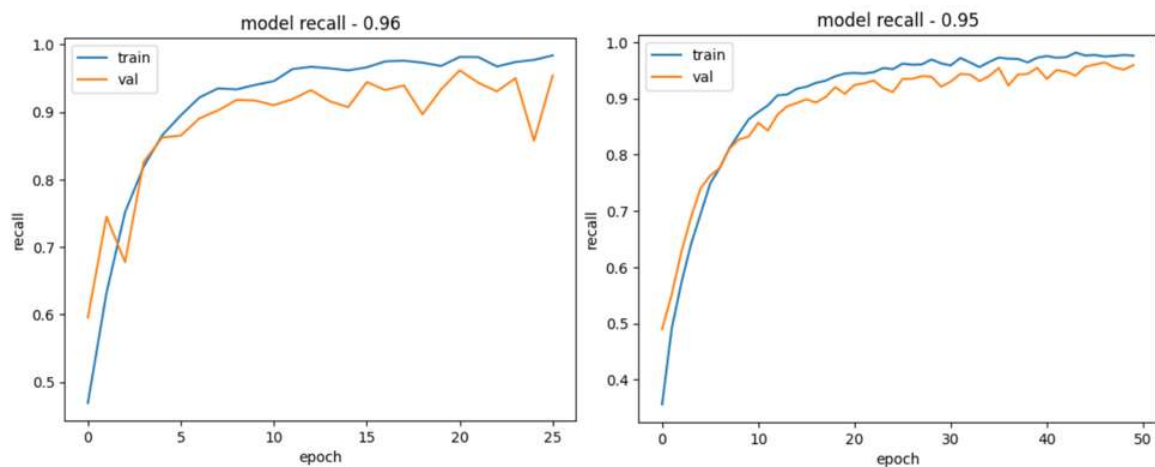


Gambar 4. 3 Grafik *Precision* Model EfficientNet-B0 Data Original (Kiri) Data Augmentasi (Kanan)

Sementara itu, analisis terhadap grafik *recall* pada Gambar 4.4 menunjukkan bahwa pada skenario data original, meskipun nilai *recall* maksimum mampu mencapai 0,96, kurva *validation recall* cenderung mengalami fluktuasi dan mengalami penurunan pada beberapa *epoch*. Kondisi ini menunjukkan bahwa kemampuan model dalam mendeteksi sampel positif belum sepenuhnya konsisten selama proses pelatihan. Penurunan nilai *validation recall* tersebut mengindikasikan adanya peningkatan kesalahan *false negative*, yaitu kondisi ketika data yang sebenarnya termasuk ke dalam kelas positif gagal dikenali atau terklasifikasi sebagai kelas lain oleh model. Dalam konteks diagnosis penyakit Alzheimer, kesalahan semacam ini perlu diperhatikan karena dapat menyebabkan kasus yang seharusnya terdeteksi menjadi terlewat.

Selain itu, terdapat jarak yang cukup terlihat antara kurva *training recall* dan *validation recall*, yang mengindikasikan bahwa model mulai mempelajari karakteristik spesifik pada data latih sehingga kemampuan generalisasinya terhadap data yang belum pernah dilihat menjadi menurun. Fenomena ini menunjukkan adanya indikasi *overfitting*

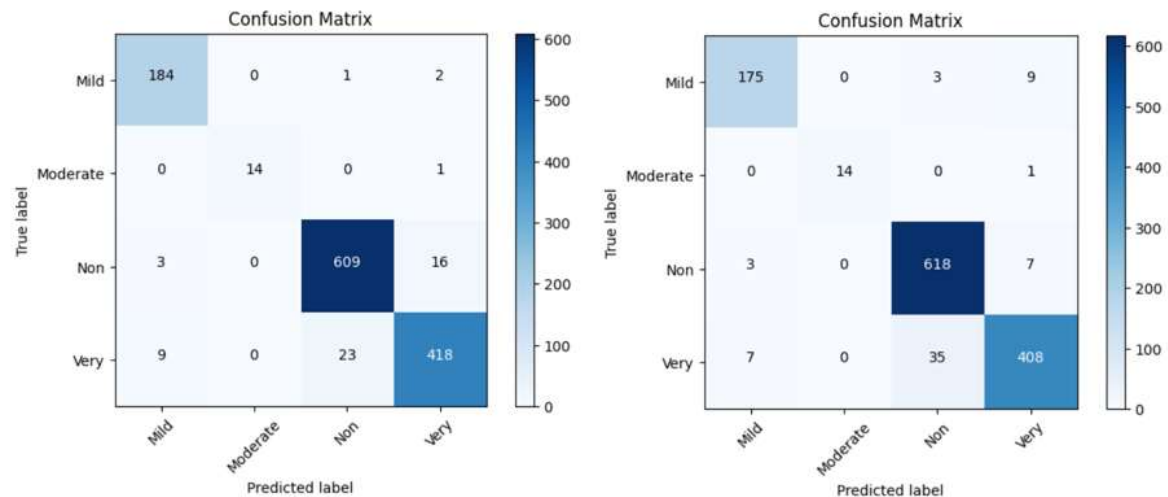
pada skenario data original. Sebaliknya, pada skenario data augmentasi, nilai *recall* berada pada kisaran 0,95 dengan pola yang lebih stabil dan konsisten sepanjang proses pelatihan. Kurva *validation recall* yang mengikuti pola *training recall* dengan selisih yang relatif kecil menunjukkan bahwa model mampu mempertahankan performa deteksi pada data yang belum pernah dilihat sebelumnya. Hasil ini mengindikasikan bahwa penerapan augmentasi data membantu meningkatkan kemampuan generalisasi model dalam mengenali sampel positif secara lebih konsisten, sehingga berpotensi meningkatkan keandalan model pada tugas klasifikasi medis.



Gambar 4. 4 Grafik *Recall* Model EfficientNet-B0 Data Original (Kiri) Data Augmentasi (Kanan)

Berdasarkan *confusion matrix* pada Gambar 4.5, skenario data original model menunjukkan prediksi yang baik pada sebagian besar kelas, namun masih terdapat kesalahan klasifikasi pada kelas dengan karakteristik yang mirip, khususnya *antara Non Demented* dan *Very Mild Demented*. Selain itu, masih ditemukan beberapa kasus *false positive* dan *false negative* yang dapat mempengaruhi keandalan model.

Confusion matrix pada skenario data augmentasi menunjukkan terdapat sedikit penurunan jumlah prediksi benar pada beberapa kelas, model menunjukkan performa yang lebih seimbang. Jumlah kesalahan pada kelas tertentu menjadi lebih terkendali, terutama penurunan *false positive* pada kelas *Non Demented*, sehingga mengurangi risiko kesalahan diagnosis pada pasien sehat. Selain itu, performa pada kelas dengan jumlah data terbatas tetap terjaga dengan baik.



Gambar 4. 5 *Confusion Matrix* Data Original (Kiri) Data Augmentasi (Kanan)

Penerapan data *augmentation* terbukti mampu meningkatkan stabilitas proses pelatihan serta kemampuan generalisasi model dalam mengenali pola citra MRI. Hal ini terlihat dari kurva pelatihan yang lebih konsisten, fluktuasi yang lebih kecil, serta kesesuaian yang lebih baik antara performa data latih dan data validasi. Meskipun model dengan augmentasi tidak selalu menghasilkan nilai metrik tertinggi secara numerik, performa yang dihasilkan cenderung lebih stabil dan tidak bergantung pada kondisi tertentu selama pelatihan.

Model juga menunjukkan kemampuan yang lebih baik dalam menangani variasi data baru, sehingga mengurangi risiko *overfitting* yang umum terjadi pada data original. Dengan demikian, model yang dihasilkan menjadi lebih robust dan memiliki tingkat keandalan yang lebih tinggi dalam melakukan klasifikasi. Oleh karena itu, dalam konteks klasifikasi medis yang menuntut konsistensi dan *accuracy* dalam berbagai kondisi, penerapan augmentasi data dapat dianggap sebagai pendekatan yang lebih tepat.

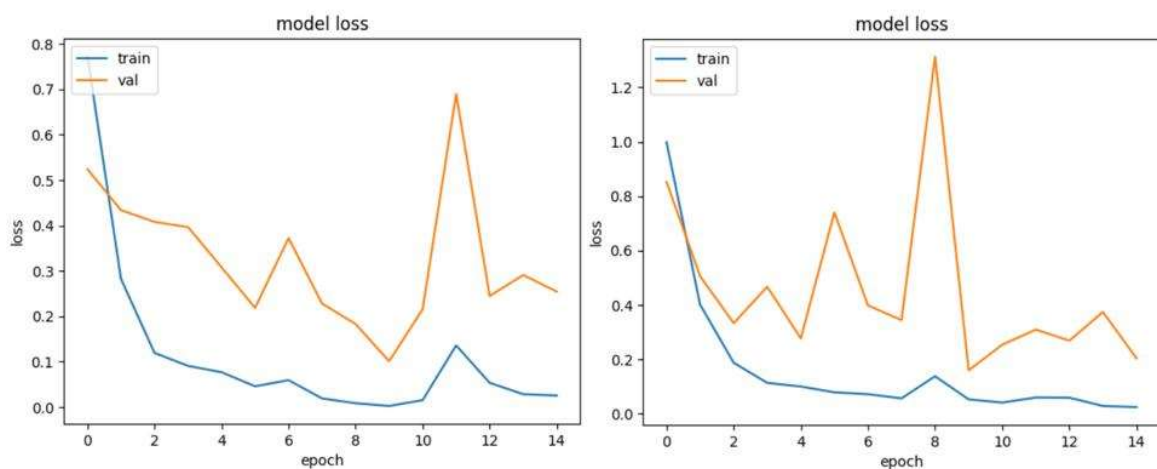
4.3.2 ResNet50

Sesuai dengan skenario, dilakukan proses pelatihan model menggunakan arsitektur ResNet50 dengan 50 epoch serta penerapan *early stopping* apabila nilai *validation accuracy* tidak mengalami peningkatan selama lima *epoch* berturut-turut. Selama proses pelatihan, dilakukan pemantauan terhadap metrik *accuracy* dan *loss* pada data latih dan data validasi untuk mengevaluasi kinerja model. Hasil pelatihan kemudian disajikan dalam bentuk grafik *loss* dan *accuracy*, *confusion matrix*, serta nilai evaluasi berupa *accuracy*, *precision*, dan *recall*. Ringkasan hasil terbaik dari skenario pelatihan ditampilkan pada Tabel 4.3.

Tabel 4. 3 Tabel Hasil Pelatihan Terbaik Masing-Masing Skenario Model ResNet50

Nama Skenario	<i>Learning Rate</i>	<i>Dropout</i>	<i>Batch Size</i>	<i>Epoch</i>	<i>Validation Accuracy</i>	<i>Accuracy</i>	<i>Recall</i>
Data Original	0,0001	0,3	32	10	0,97	0,92	0,92
Data dengan Augmentasi	0,001	0,3	16	23	0,95	0,90	0,90
Data dengan Class Weight	0,0001	0,3	16	10	0,945	0,93	0,93
Data dengan Augmentasi dan Class Weight	0,0001	0,5	16	20	0,96	0,89	0,89

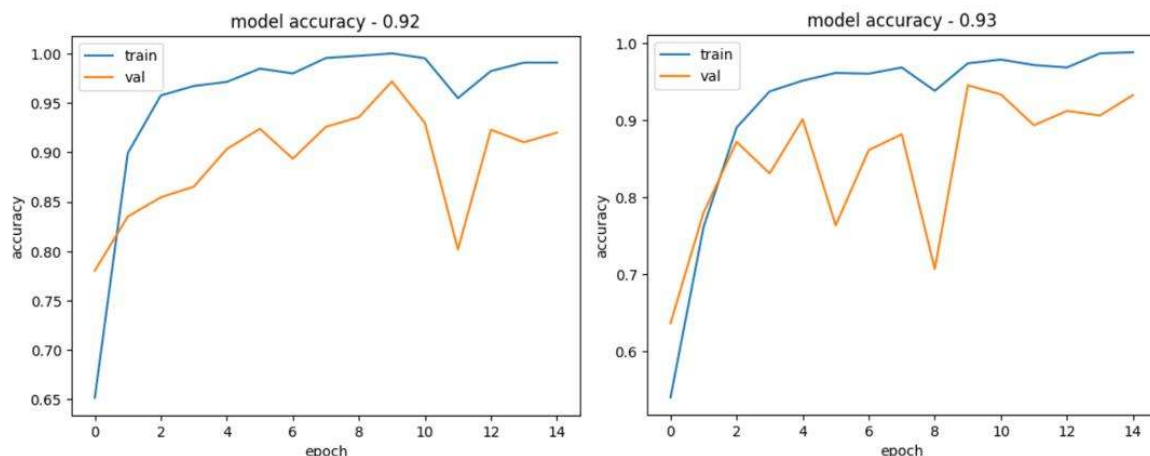
Tabel 4.3 pada skenario data original dan data dengan penerapan *class weight* menunjukkan performa terbaik. Skenario data original memiliki nilai *validation accuracy* tertinggi sebesar 0,97, yang mencerminkan performa model yang sangat baik secara umum. Sementara itu, skenario dengan *class weight* menghasilkan nilai *recall* tertinggi sebesar 0,93, yang menjadi metrik penting dalam konteks medis karena berkaitan dengan kemampuan model dalam mendeteksi seluruh kasus positif. Dengan demikian, kedua skenario dipilih untuk analisis lebih lanjut karena mewakili keseimbangan antara performa umum dan sensitivitas terhadap penyakit.



Gambar 4. 6 Grafik *Loss* Model ResNet50 Data Original (Kiri) Data dengan *Class Weight* (Kanan)

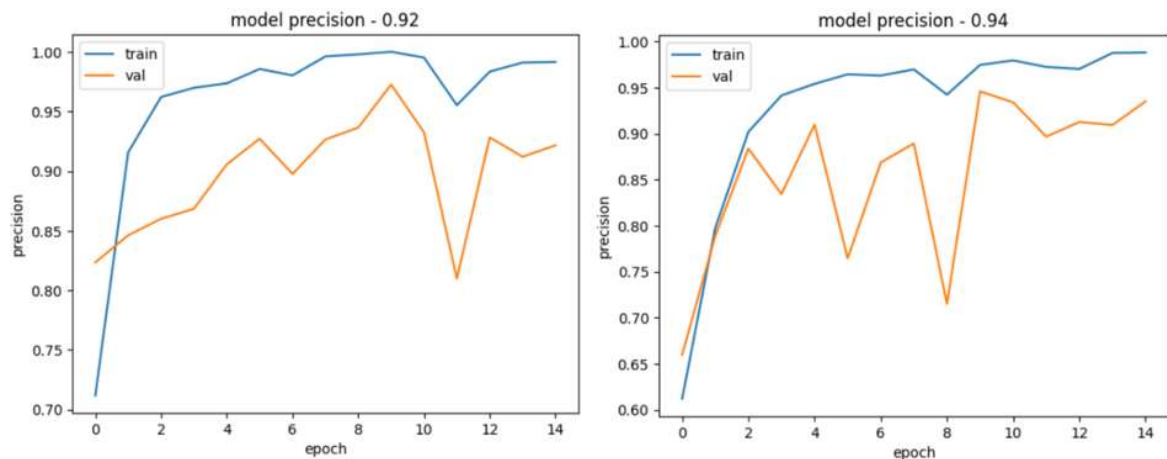
Analisis grafik *loss* pada Gambar 4.6 menunjukkan bahwa skenario data original mengalami penurunan *training loss* dengan sangat cepat hingga mendekati nol, yang menandakan model mampu mempelajari data latih dengan baik. Namun, *validation loss* menunjukkan fluktuasi yang cukup tajam, terutama setelah *epoch* ke-10, serta terdapat kesenjangan yang cukup besar antara *training* dan *validation loss*. Hal ini mengindikasikan terjadinya *overfitting*, di mana model kurang mampu melakukan generalisasi terhadap data baru.

Skenario dengan *class weight*, *validation loss* menunjukkan fluktuasi yang lebih tinggi, terutama pada fase awal pelatihan. Hal ini disebabkan oleh penalti yang lebih besar terhadap kesalahan pada kelas minoritas. Meskipun demikian, pada akhir pelatihan terlihat adanya penurunan *loss* yang signifikan, yang menunjukkan bahwa model mulai mencapai keseimbangan dalam pembelajaran dan mampu menangani distribusi data yang tidak seimbang dengan lebih baik.



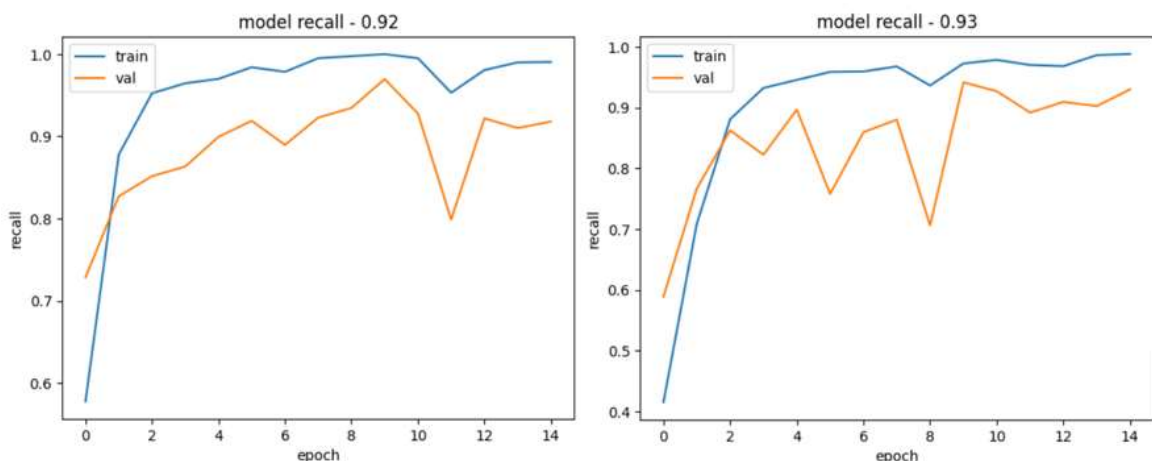
Gambar 4. 7 Grafik *Accuracy* Model ResNet50 Data Original (Kiri) Data dengan *Class Weight* (Kanan)

Analisis grafik *accuracy* pada Gambar 4.7 menunjukkan bahwa *training accuracy* pada data original meningkat dengan cepat hingga mendekati nilai maksimum, sementara *validation accuracy* cenderung menurun pada beberapa *epoch* tertentu. Perbedaan yang cukup besar antara kedua kurva menunjukkan adanya *overfitting*. Sebaliknya, pada skenario *class weight*, kurva *accuracy* terlihat lebih fluktuatif akibat penyesuaian bobot yang lebih agresif. Meskipun demikian, model mampu mencapai *accuracy* yang relatif baik serta menunjukkan peningkatan sensitivitas terhadap distribusi data yang tidak seimbang.



Gambar 4. 8 Grafik *Precision* Model *ResNet50* Data Original (Kiri) Data dengan *Class Weight* (Kanan)

Grafik *precision* pada Gambar 4.8, skenario data original menunjukkan nilai yang tinggi pada data latih, namun *validation precision* mengalami fluktuasi yang cukup besar dan penurunan pada beberapa *epoch*, yang mengindikasikan peningkatan *false positive* serta adanya *overfitting*. Sebaliknya, pada skenario *class weight*, nilai *precision* meningkat hingga sekitar 0,94. Meskipun kurva terlihat fluktuatif, pada akhir pelatihan *precision* menjadi lebih stabil, yang menunjukkan bahwa model mulai mampu menyeimbangkan prediksi antar kelas.



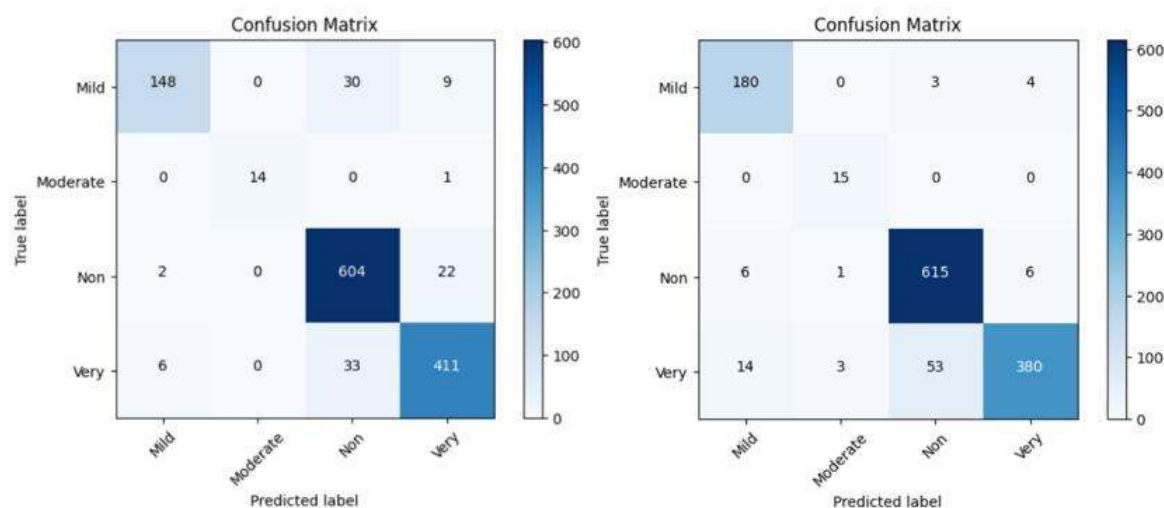
Gambar 4. 9 Grafik *Recall* Model *ResNet50* Data Original (Kiri) Data dengan *Class Weight* (Kanan)

Sementara itu, grafik *recall* pada Gambar 4.9 menunjukkan bahwa pada data original terdapat fluktuasi yang cukup signifikan pada *validation recall*, bahkan mengalami penurunan pada beberapa *epoch*, yang mengindikasikan tingginya risiko *false negative*. Sebaliknya, pada skenario *class weight*, model mampu mencapai *recall* sebesar 0,93, yang

menunjukkan peningkatan kemampuan dalam mendeteksi kasus positif. Meskipun kurva awalnya fluktuatif, pada akhir pelatihan menunjukkan pola yang lebih stabil dan meningkat, sehingga lebih sesuai untuk kebutuhan klasifikasi medis.

Berdasarkan *confusion matrix* pada Gambar 4.10, menunjukkan bahwa skenario data original model cenderung bias terhadap kelas mayoritas (*Non Demented*), yang ditunjukkan oleh tingginya prediksi benar pada kelas tersebut, namun diikuti dengan tingginya kesalahan pada kelas penderita, khususnya *Mild Demented* dan *Very Mild Demented* yang sering diklasifikasikan sebagai normal. Hal ini menunjukkan tingginya *false negative* yang berisiko dalam konteks medis.

Sebaliknya, pada skenario dengan *class weight*, model menunjukkan peningkatan sensitivitas terhadap kelas minoritas. Hal ini ditandai dengan peningkatan jumlah prediksi benar pada kelas *Mild Demented* serta keberhasilan mendeteksi seluruh data pada kelas *Moderate Demented*. Selain itu, kesalahan klasifikasi pada kelas tertentu mengalami penurunan yang cukup signifikan. Namun demikian, masih terdapat tantangan dalam membedakan kelas dengan perbedaan fitur yang sangat tipis, terutama antara *Very Mild Demented* dan *Non Demented*.



Gambar 4. 10 *Confusion Matrix* Model ResNet50 Data Original (Kiri) Data dengan *Class Weight* (Kanan)

Penerapan *class weight* terbukti mampu meningkatkan kemampuan model dalam menangani permasalahan *class imbalance* serta meningkatkan sensitivitas model terhadap kelas yang jumlah datanya lebih sedikit. Dengan memberikan bobot yang lebih besar pada kelas minoritas, model menjadi lebih memperhatikan kesalahan prediksi pada kelas tersebut

selama proses pelatihan. Meskipun penerapan *class weight* menyebabkan fluktuasi performa yang lebih tinggi pada beberapa metrik selama proses pelatihan, pendekatan ini membantu mengurangi kecenderungan model untuk lebih berfokus pada kelas mayoritas. Akibatnya, model yang dihasilkan menjadi lebih seimbang dan adil dalam mempelajari karakteristik setiap kelas. Selain itu, peningkatan kemampuan dalam mendeteksi kasus positif menunjukkan bahwa model memiliki sensitivitas yang lebih baik terhadap kondisi penyakit, sehingga berpotensi mengurangi terjadinya kesalahan *false negative*. Karakteristik tersebut menjadikan model lebih andal dan lebih sesuai untuk diterapkan pada konteks klasifikasi medis, di mana kemampuan mendeteksi kasus penyakit secara akurat memiliki peran yang sangat penting.

4.3.3 Perbandingan Performa *Baseline* (*Hyperparameter* Identik)

Evaluasi keunggulan masing-masing arsitektur secara objektif serta mengukur tingkat sensitivitasnya terhadap *hyperparameter* diperlukan. Perbandingan dilakukan dengan membandingkan dua kondisi pengujian, yaitu kondisi *baseline* (kontrol) dan optimal (hasil tuning). Tabel 4.4, menunjukkan kondisi optimal yang merepresentasikan performa terbaik yang dicapai setiap model dengan konfigurasi *hyperparameter* masing-masing, sedangkan kondisi *baseline* diperoleh dengan menukar *hyperparameter* optimal antar model. Evaluasi difokuskan pada dua metrik utama yang divisualisasikan dalam grafik, yaitu *validation accuracy* untuk menilai kemampuan generalisasi model dan *recall* (sensitivitas) untuk mengukur kemampuan model dalam mengenali kelas positif.

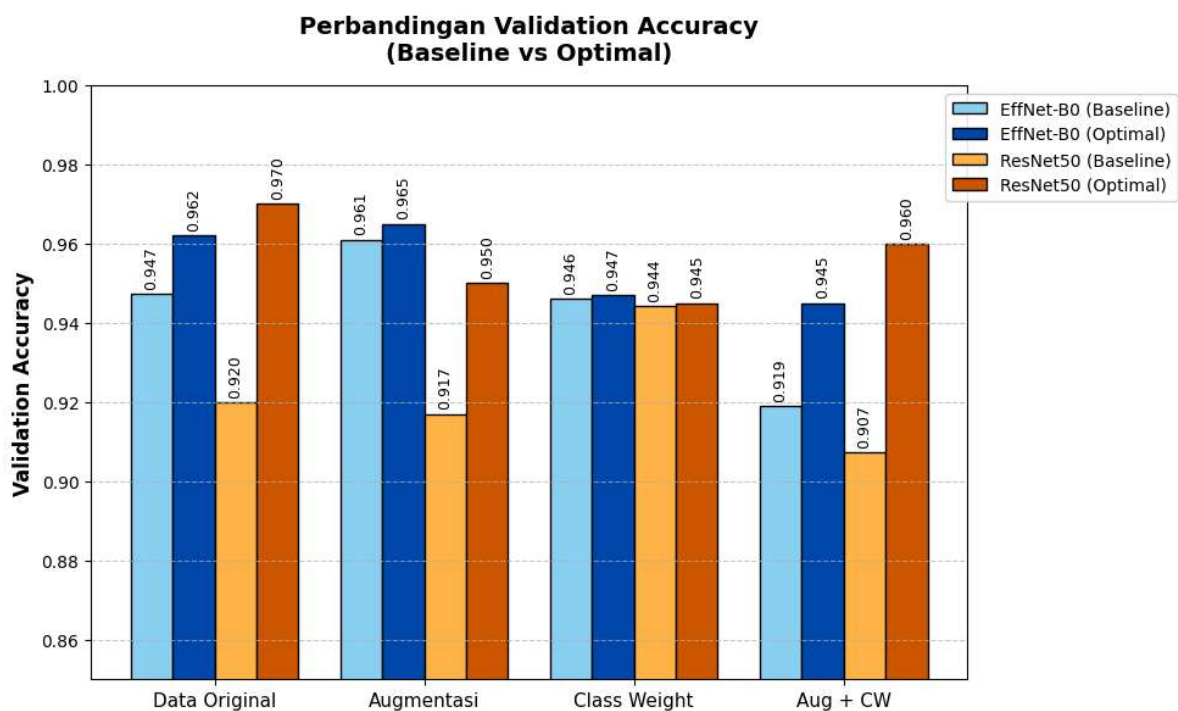
Pengaruh variasi *hyperparameter* ini dapat dianalisis lebih lanjut melalui Tabel 4.4. ResNet50 terbukti sangat sensitif terhadap nilai *learning rate*, di mana penggunaan nilai yang lebih besar (0,001) pada kondisi *baseline* menyebabkan proses pelatihan kurang stabil dan sulit mencapai konvergensi optimal. Sebaliknya, penggunaan *learning rate* yang lebih kecil (0,0001) pada kondisi optimal mampu meningkatkan stabilitas dan performa model secara signifikan. Di sisi lain, EfficientNet-B0 menunjukkan fleksibilitas yang lebih tinggi terhadap variasi *learning rate*. Selain itu, penggunaan *batch size* yang lebih kecil (16) pada skenario augmentasi dan *class weight* membantu kedua model dalam menangkap variasi data yang lebih kompleks. Khusus pada ResNet50, peningkatan nilai *dropout* menjadi 0,5 pada kondisi optimal terbukti dapat mengurangi *overfitting* dan berkontribusi terhadap pencapaian *accuracy* maksimum.

Tabel 4. 4 Tabel Perbandingan *Baseline Hyperparameter*

Nama Skenario	Kondisi	Arsitektur Model	<i>Learning Rate</i>	<i>Dropout</i>	<i>Batch Size</i>	<i>Epoch</i>	<i>Validation Accuracy</i>	<i>Recall</i>
Data Original	Optimal	ResNet50	0,0001	0,3	32	10	0,970	0,920
	<i>Baseline</i>	EfficientNetB0	0,0001	0,3	32	38	0,947	0,973
	Optimal	EfficientNetB0	0,0010	0,3	32	21	0,962	0,960
	<i>Baseline</i>	ResNet50	0,0010	0,3	32	6	0,920	0,950
Data dengan Augmentasi	Optimal	ResNet50	0,0010	0,3	16	23	0,950	0,900
	<i>Baseline</i>	EfficientNetB0	0,0010	0,3	16	20	0,961	0,970
	Optimal	EfficientNetB0	0,0001	0,3	16	47	0,965	0,950
	<i>Baseline</i>	ResNet50	0,0001	0,3	16	5	0,917	0,940
Data dengan <i>Class Weight</i>	Optimal	ResNet50	0,0001	0,3	16	10	0,945	0,930
	<i>Baseline</i>	EfficientNetB0	0,0001	0,3	16	42	0,946	0,960

Nama Skenario	Kondisi	Arsitektur Model	<i>Learning Rate</i>	<i>Dropout</i>	<i>Batch Size</i>	<i>Epoch</i>	<i>Validation Accuracy</i>	<i>Recall</i>
	Optimal	EfficientNetB0	0,0010	0,3	16	21	0,947	0,950
	<i>Baseline</i>	ResNet50	0,0010	0,3	16	18	0,944	0,960
Data dengan Augmentasi dan <i>Class Weight</i>	Optimal	ResNet50	0,0001	0,5	16	20	0,960	0,890
	<i>Baseline</i>	EfficientNetB0	0,0001	0,5	16	31	0,918	0,932
	Optimal	EfficientNetB0	0,0001	0,3	16	36	0,945	0,930
	<i>Baseline</i>	ResNet50	0,0001	0,3	16	8	0,907	0,970

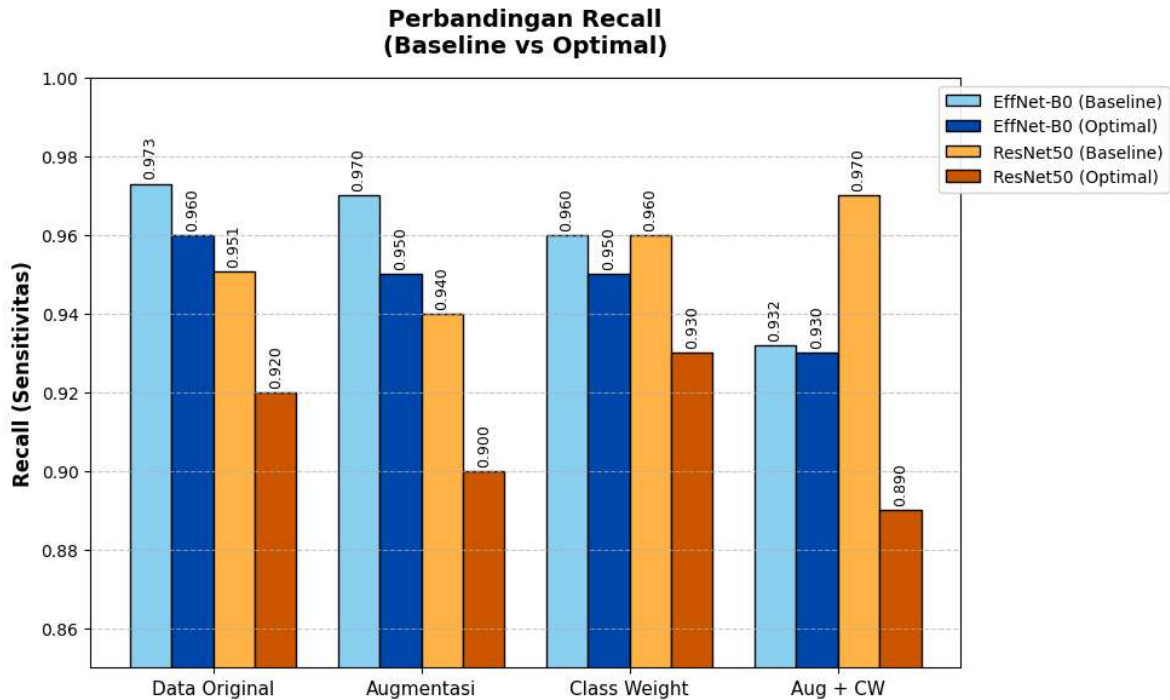
Grafik *validation accuracy* pada Gambar 4.11, terlihat perbedaan karakteristik yang kontras antara kedua arsitektur. EfficientNet-B0 menunjukkan tingkat stabilitas (*robustness*) yang tinggi, di mana model ini tetap mampu menghasilkan *accuracy* yang kompetitif meskipun menggunakan konfigurasi *baseline* yang bukan merupakan parameter optimalnya. Selisih *accuracy* antara kondisi baseline dan optimal pada EfficientNet-B0 relatif kecil; bahkan pada skenario data original, *accuracy baseline* telah mencapai 0,947. Sebaliknya, ResNet50 menunjukkan sensitivitas yang tinggi terhadap perubahan *hyperparameter*, namun memiliki potensi performa maksimum yang lebih besar. Pada kondisi *baseline*, performanya mengalami penurunan signifikan, misalnya menjadi 0,920 pada data original dan 0,907 pada skenario kombinasi augmentasi dan *class weight*. Namun, setelah dilakukan optimasi, performa ResNet50 meningkat tajam dan melampaui EfficientNet-B0 dengan *validation accuracy* mencapai 0,970 dan 0,960 pada masing-masing skenario tersebut.



Gambar 4. 11 Perbandingan *Validation Accuracy Baseline* dan Optimal

Sementara itu, analisis terhadap grafik recall pada Gambar 4.12 menunjukkan adanya fenomena *trade-off* yang konsisten setelah proses *hyperparameter tuning*. Pada hampir seluruh skenario pengujian, kondisi optimal justru menghasilkan nilai *recall* yang lebih rendah dibandingkan *baseline*, baik pada EfficientNet-B0 maupun ResNet50. Penurunan paling signifikan terjadi pada ResNet50 dalam skenario augmentasi dan *class weight*, di

mana *recall* menurun dari 0,970 menjadi 0,890. Temuan ini mengindikasikan bahwa optimasi yang berfokus pada peningkatan *accuracy* keseluruhan cenderung membuat model lebih konservatif dalam memprediksi kelas positif, sehingga sensitivitasnya terhadap kelas tersebut menurun.



Gambar 4. 12 Perbandingan *Recall Baseline* dan Optimal

Hasil pengujian tersebut menunjukkan bahwa EfficientNet-B0 merupakan arsitektur yang lebih stabil dan tidak terlalu bergantung pada proses *tuning hyperparameter*, sehingga cocok digunakan dalam kondisi keterbatasan sumber daya komputasi. Sebaliknya, ResNet50 merupakan arsitektur yang masih sangat bergantung pada konfigurasi *hyperparameter*, namun memiliki potensi performa yang lebih tinggi apabila dilakukan optimasi secara tepat.

4.4 Interpretasi Model Grad-CAM

Pada tahap ini dilakukan interpretasi model menggunakan metode Grad-CAM untuk memahami bagaimana model dalam mengambil keputusan klasifikasi. Meskipun evaluasi kuantitatif seperti *accuracy*, *precision*, dan *recall* menunjukkan performa model yang baik, metrik tersebut belum sepenuhnya mampu menjelaskan apakah model benar-benar mempelajari fitur yang relevan secara klinis. Hal ini menjadi sangat penting dalam domain medis, di mana keputusan model harus dapat dipertanggungjawabkan dan tidak hanya bergantung pada nilai performa numerik. Oleh karena itu, diperlukan pendekatan interpretasi

visual seperti Grad-CAM untuk mengidentifikasi area citra yang menjadi fokus model dalam melakukan prediksi (Selvaraju dkk., 2017). Dengan demikian, dapat dianalisis apakah model benar-benar memperhatikan bagian otak yang relevan dengan indikasi penyakit Alzheimer atau justru terfokus pada area yang tidak berkaitan.

Visualisasi Grad-CAM pada penelitian ini disajikan dalam tiga tingkatan proses untuk memberikan pemahaman yang lebih komprehensif. Tahap *raw* merupakan *heatmap* mentah yang dihasilkan dari ekstraksi gradien pada lapisan konvolusi terakhir. Selanjutnya, tahap *smooth* merupakan hasil penerapan teknik penghalusan (*smoothing*) pada *heatmap* untuk mengurangi *noise* visual. Adapun tahap *clean* merupakan representasi akhir di mana *heatmap* telah melalui proses *masking*, sehingga area aktivasi hanya terfokus pada bagian anatomi otak. Tahap *masking* pada visualisasi *clean* dilakukan dengan menerapkan ambang batas (*thresholding*) biner untuk memisahkan intensitas piksel jaringan otak dari latar belakang citra. Melalui perkalian matriks antara *heatmap* dan masker biner tersebut, seluruh aktivasi di luar area anatomi klinis dieliminasi untuk memastikan fokus interpretasi hanya tertuju pada struktur otak yang relevan. Ketiga tahapan tadi bertujuan untuk mempermudah identifikasi fitur-fitur anatomis yang relevan yang digunakan model dalam proses klasifikasi.

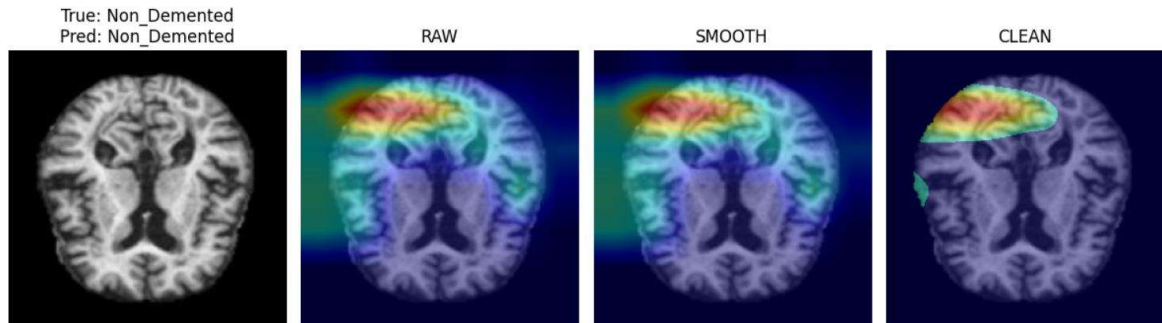
4.4.1 EfficientNet-B0

4.4.1.1 Data Original

Visualisasi Grad-CAM pada model EfficientNet-B0 untuk skenario data original pada Gambar 4.13 menunjukkan bahwa area aktivasi masih relatif luas dan cenderung menyebar, bahkan sebagian mencakup area di luar jaringan otak. Hal ini mengindikasikan bahwa model masih dipengaruhi oleh *noise* atau fitur non-klinis dalam proses pengambilan keputusan. Pada tahap *raw* dan *smooth*, aktivasi belum sepenuhnya terlokalisasi dengan baik dan masih menunjukkan adanya *activation leakage* ke latar belakang citra.

Proses *cleaning* dilakukan melalui *masking* yang menunjukkan area aktivasi menjadi lebih terfokus pada struktur anatomi otak, khususnya pada lobus frontal. Temuan ini menunjukkan bahwa meskipun model pada tahap awal belum sepenuhnya selektif, sebagian fitur yang dipelajari kemungkinan telah berkaitan dengan struktur anatomi otak yang relevan. Dengan kata lain, proses pemurnian visualisasi mengungkap bahwa model mulai

menangkap pola yang berkaitan dengan struktur otak, meskipun masih terdapat ketergantungan terhadap area non-esensial.

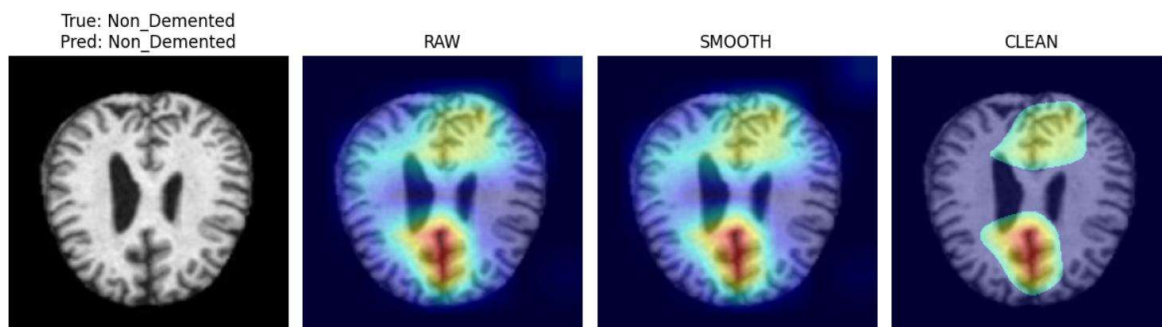


Gambar 4. 13 Grad-CAM Model EfficientNet-B0 pada Data Original

Jika dikaitkan dengan konteks klinis, fokus pada lobus frontal dapat menjadi indikasi awal keterlibatan area korteks asosiasi, meskipun pada tahap ini pola aktivasi belum cukup spesifik untuk dijadikan dasar interpretasi yang kuat. Oleh karena itu, pada skenario data original, model dapat dikatakan masih berada pada tahap awal pembelajaran fitur, di mana kemampuan dalam membedakan struktur anatomi yang relevan dan *noise* belum sepenuhnya optimal.

4.4.1.2 Data dengan Class Weight

Gambar 4.14 memperlihatkan visualisasi Grad-CAM pada citra MRI yang termasuk kelas *Non-Demented* dan berhasil diklasifikasikan dengan benar oleh model pada skenario *class weight*. Pada visualisasi *raw*, area aktivasi tertinggi ditunjukkan oleh warna merah hingga kuning yang terkonsentrasi pada korteks serebral bagian superior serta area di sekitar ventrikel lateral. Setelah dilakukan proses *smoothing*, pola perhatian model menjadi lebih jelas dan tetap berfokus pada wilayah yang sama.

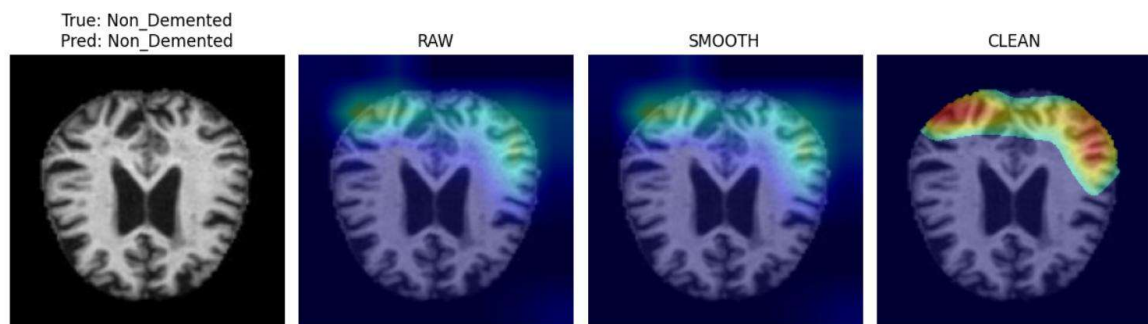


Gambar 4. 14 Grad-CAM Model EfficientNet-B0 pada Data dengan Class Weight

Visualisasi *clean* pada Gambar 2.4 menunjukkan bahwa model terutama memanfaatkan informasi dari korteks frontal-parietal dan area sekitar ventrikel sebagai dasar pengambilan keputusan. Aktivasi yang terdistribusi pada beberapa struktur anatomi otak mengindikasikan bahwa model mulai mampu mempertimbangkan fitur anatomis otak yang relevan dalam mengenali kondisi normal.

4.4.1.3 Data dengan Augmentasi

Gambar 4.15 menunjukkan visualisasi Grad-CAM pada citra MRI kelas *Non-Demented* yang diklasifikasikan dengan benar oleh model pada skenario *data augmentation*. Pada visualisasi *raw* dan *smooth*, area aktivasi tertinggi tampak terkonsentrasi pada korteks serebral bagian superior, khususnya pada wilayah frontal dan parietal. Setelah dilakukan proses *cleaning*, terlihat bahwa model terutama memanfaatkan informasi dari permukaan korteks yang membentang pada bagian atas otak sebagai dasar pengambilan keputusan. Fokus aktivasi yang dominan pada area korteks menunjukkan bahwa area tersebut memberikan kontribusi yang relatif besar terhadap proses klasifikasi pada kategori *Non Demented*. Hasil ini menunjukkan bahwa penerapan augmentasi membantu model mempelajari fitur yang lebih umum dan konsisten sehingga perhatian model terfokus pada struktur otak yang relevan untuk proses klasifikasi.

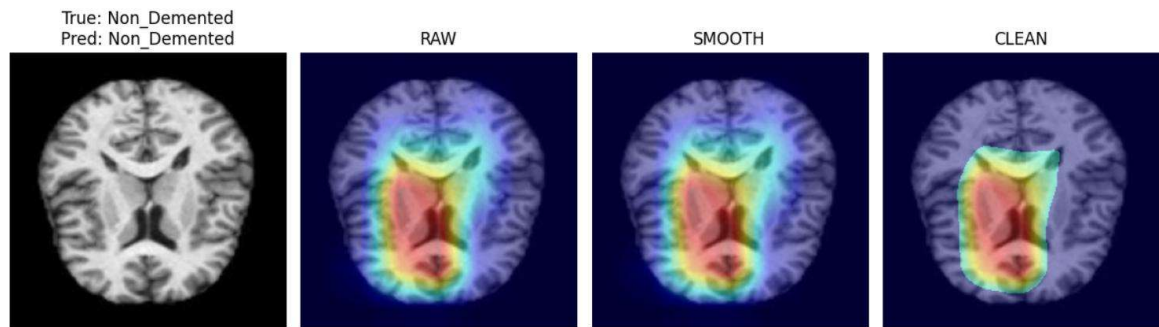


Gambar 4. 15 Grad-CAM Model EfficientNet-B0 pada Data dengan Augmentasi

4.4.1.4 Data dengan Augmentasi dan Class Weight

Visualisasi Grad-CAM pada model EfficientNet-B0 dengan kombinasi augmentasi dan *class weight* pada Gambar 4.16 menampilkan fokus aktivasi yang lebih terpusat pada area anatomi otak dibandingkan skenario lainnya. Pada tahap *raw* dan *smooth*, area aktivasi terlihat sangat terpusat di dalam struktur otak dengan hampir tidak adanya *activation leakage*

ke latar belakang. Hal ini menandakan bahwa model telah mampu membedakan secara jelas antara fitur anatomis dan *noise*, serta memiliki konsistensi fokus yang tinggi.



Gambar 4. 16 Grad-CAM Model EfficientNet-B0 pada Data dengan Augmentasi dan Class Weight

Pada tahap *clean*, model memfokuskan perhatian utamanya pada area di sekitar ventrikel dan beberapa bagian subkortikal. Pergeseran fokus ini berbeda dari skenario sebelumnya yang cenderung menyoroti lobus frontal, dan memiliki relevansi klinis yang kuat. Area di sekitar ventrikel diketahui sering digunakan dalam studi neuroimaging karena perubahan ukuran ventrikel berkaitan dengan perubahan volume jaringan otak pada berbagai tingkat keparahan demensia (Centro dkk., 2011). Aktivasi pada area tersebut mengindikasikan bahwa bagian tersebut memberikan kontribusi yang relatif besar terhadap proses pengambilan keputusan model.

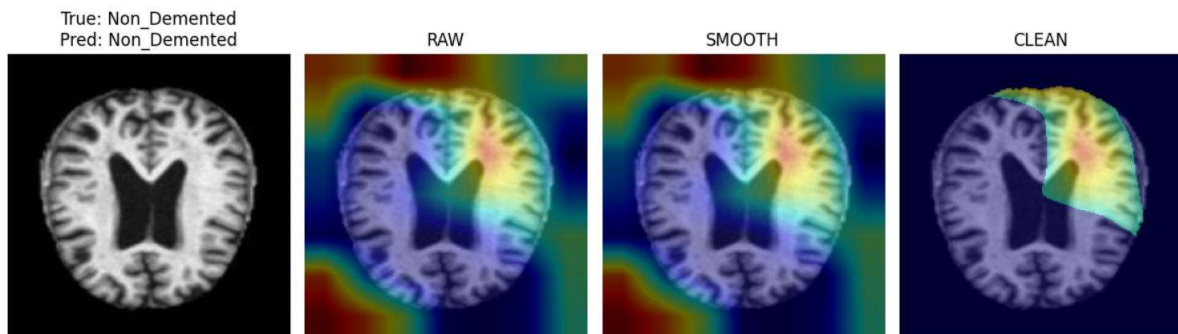
Jika dibandingkan dengan skenario augmentasi maupun *class weight* secara terpisah, kombinasi kedua teknik ini menghasilkan pola aktivasi yang lebih selektif sekaligus stabil. Augmentasi membantu model menjadi lebih robust terhadap variasi spasial citra, sementara *class weight* meningkatkan sensitivitas terhadap fitur-fitur penting pada kelas minoritas. Hasil visualisasi menunjukkan bahwa kombinasi kedua teknik tersebut berkorelasi dengan pola aktivasi yang lebih konsisten dan terfokus dibandingkan penggunaan masing-masing teknik secara terpisah.

Dengan demikian, skenario ini tidak hanya memberikan performa yang baik secara kuantitatif, tetapi juga menghasilkan interpretasi visual yang paling terlokalisasi dan konsisten. Hal ini menunjukkan bahwa kombinasi augmentasi dan *class weight* merupakan pendekatan yang dapat meningkatkan kualitas pembelajaran model, baik dari sisi *accuracy* maupun dari sisi akuntabilitas dalam konteks diagnosis medis.

4.4.2 ResNet50

4.4.2.1 Data Original

Visualisasi Grad-CAM pada model ResNet50 dengan skenario data original pada Gambar 4.17 menunjukkan ketidakmampuan model dalam mengidentifikasi fitur anatomis yang relevan. Pada tahap *raw*, area aktivasi justru terkonsentrasi pada sudut-sudut latar belakang citra, bukan pada jaringan otak. Fenomena ini mengindikasikan adanya bias spasial serta *shortcut learning*. Dalam konteks *deep learning*, *shortcut learning* terjadi ketika model memanfaatkan strategi pengambilan keputusan yang tidak diharapkan, seperti mengenali noise atau latar belakang yang membuatnya tampak berkinerja baik pada data pengujian, namun sebenarnya gagal mempelajari target klinis yang sesungguhnya (Geirhos dkk., 2020).



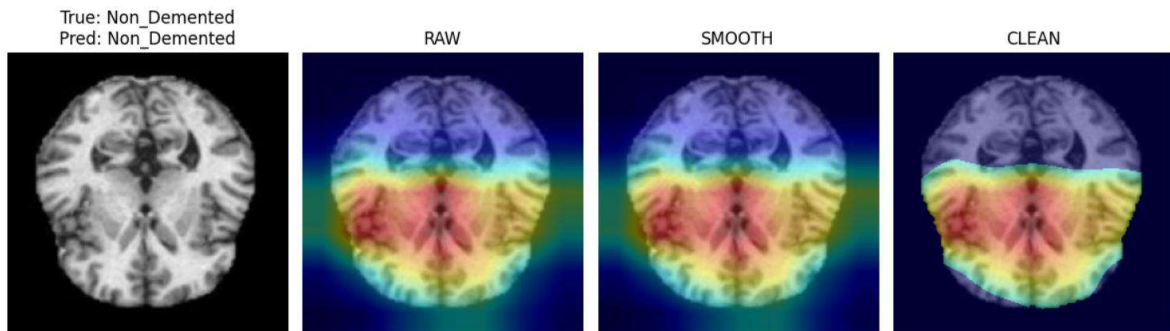
Gambar 4. 17 Grad-CAM Model ResNet50 pada Data Original

Fenomena *shortcut learning* sering terjadi dalam klasifikasi citra medis, di mana model secara keliru menggunakan fitur di luar organ tubuh sebagai dasar prediksi alih-alih mempelajari patologi penyakit itu sendiri. Kondisi ini diperkuat oleh hasil visualisasi yang menunjukkan minimnya aktivasi pada area parenkim otak, sehingga model belum mampu memusatkan perhatian pada struktur otak yang relevan untuk proses klasifikasi. Dengan demikian, pada skenario data original, meskipun menghasilkan validation accuracy tertinggi, hasil interpretasi Grad-CAM menunjukkan bahwa model belum memanfaatkan fitur yang relevan secara konsisten sehingga performa kuantitatif tersebut perlu ditinjau secara hati-hati.

4.4.2.2 Data dengan Class Weight

Penerapan *class weight* pada model ResNet50 yang dapat dilihat pada Gambar 4.18 menunjukkan perbaikan yang signifikan dalam kualitas fokus dibandingkan skenario data

original. Area aktivasi yang sebelumnya terpusat pada latar belakang mulai bergeser ke dalam jaringan otak, dengan area aktivasi yang mulai terkonsentrasi pada struktur anatomi otak. Hal ini mengindikasikan bahwa model mulai mampu membedakan antara fitur anatomis yang relevan dan noise.



Gambar 4. 18 Grad-CAM Model ResNet50 pada Data dengan Class Weight

Visualisasi Grad-CAM pada model ResNet50 untuk skenario *class weight* menunjukkan citra MRI kelas *Non Demented* yang berhasil diklasifikasikan dengan benar oleh model. Pada visualisasi *raw*, area aktivasi tertinggi terlihat terkonsentrasi pada bagian tengah hingga bawah struktur otak, yang ditunjukkan oleh warna merah dan kuning. Setelah dilakukan proses *smoothing*, pola aktivasi menjadi lebih halus namun tetap mempertahankan fokus pada wilayah yang sama. Visualisasi *clean* menunjukkan bahwa sebagian besar area aktivasi berada di dalam jaringan otak dengan penyebaran yang relatif simetris pada kedua hemisfer serta minim aktivasi pada latar belakang citra.

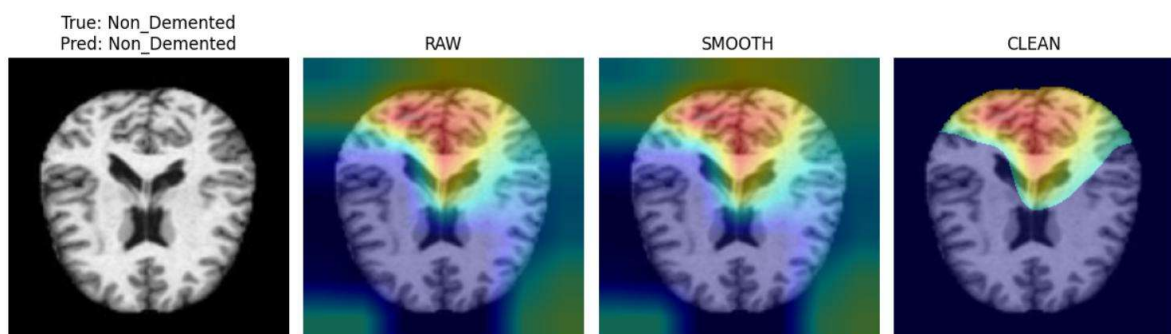
Pola aktivasi tersebut mengindikasikan bahwa model memanfaatkan informasi dari beberapa region anatomi otak secara bersamaan dalam proses pengambilan keputusan. Selain itu, berkurangnya *activation leakage* pada area di luar otak menunjukkan bahwa penerapan *class weight* membantu model memusatkan perhatian pada fitur yang lebih relevan dibandingkan *noise* di sekitar citra. Secara keseluruhan, hasil visualisasi ini menunjukkan bahwa model mampu menghasilkan fokus aktivasi yang relatif konsisten dan terlokalisasi pada struktur otak yang berkontribusi terhadap klasifikasi kategori *Non Demented*.

4.4.2.3 Data dengan Augmentasi

Gambar 4.19 menunjukkan visualisasi Grad-CAM pada model ResNet50 untuk skenario augmentasi data pada citra MRI kelas *Non Demented* yang berhasil diklasifikasikan

dengan benar oleh model. Pada visualisasi *raw*, area aktivasi tertinggi yang ditunjukkan oleh warna merah hingga kuning terkonsentrasi pada bagian superior dan tengah struktur otak. Setelah dilakukan proses *smoothing*, pola aktivasi tetap mempertahankan lokasi yang sama dengan distribusi yang lebih halus dan konsisten. Hasil visualisasi *clean* memperlihatkan bahwa sebagian besar perhatian model terfokus pada area kortikal bagian atas serta bagian di sekitar garis tengah otak, sementara aktivasi pada area lain relatif lebih rendah.

Pola aktivasi yang terlokalisasi ini menunjukkan bahwa model memanfaatkan informasi dari bagian anatomi tertentu yang memberikan kontribusi terbesar terhadap proses klasifikasi. Dibandingkan dengan skenario *class weight*, aktivasi pada skenario augmentasi tampak lebih terpusat dan menunjukkan *activation leakage* yang lebih rendah pada area di luar bagian utama. Hasil ini mengindikasikan bahwa penerapan augmentasi data membantu model mempelajari representasi fitur yang lebih konsisten sehingga perhatian model menjadi lebih terarah pada struktur otak yang berkontribusi terhadap proses pengambilan keputusan.

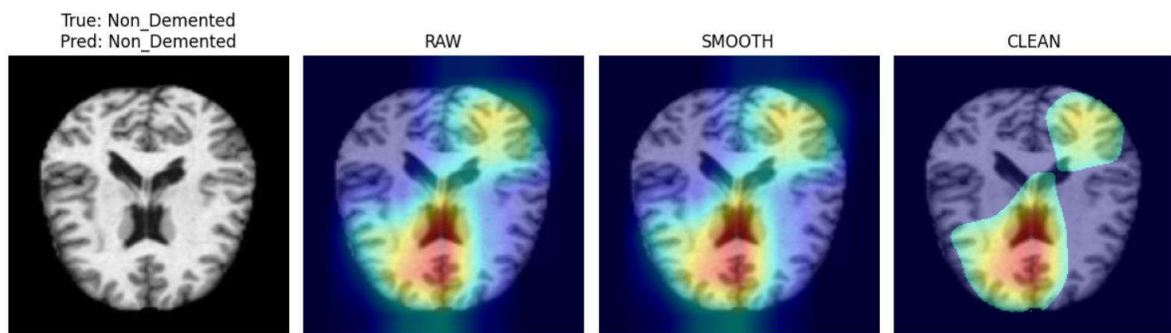


Gambar 4. 19 Grad-CAM Model ResNet50 pada Data dengan Augmentasi

4.4.2.4 Data dengan Augmentasi dan Class Weight

Gambar 4.20 menunjukkan visualisasi Grad-CAM pada model ResNet50 untuk skenario kombinasi augmentasi data dan *class weight* pada citra MRI kelas *Non Demented* yang berhasil diklasifikasikan dengan benar oleh model. Pada visualisasi *raw*, area aktivasi tertinggi terlihat terkonsentrasi pada bagian tengah-bawah otak serta sebagian area kortikal pada sisi kanan atas citra. Setelah dilakukan proses *smoothing*, pola aktivasi tetap berada pada wilayah yang sama dengan distribusi yang lebih halus dan konsisten. Hasil visualisasi *clean* memperlihatkan bahwa perhatian model terfokus pada dua region utama, yaitu area di sekitar bagian tengah otak dan area kortikal pada bagian superior. Aktivasi yang relatif terlokalisasi tersebut menunjukkan bahwa model memanfaatkan beberapa struktur anatomi otak yang memberikan kontribusi terbesar terhadap prediksi model.

Dibandingkan dengan skenario *class weight* maupun *augmentasi* secara terpisah, pola aktivasi pada kombinasi kedua teknik ini tampak lebih selektif dan memiliki *activation leakage* yang lebih rendah terhadap area di luar region utama. Konsistensi lokasi aktivasi antara visualisasi *raw*, *smooth*, dan *clean* mengindikasikan bahwa model memiliki fokus perhatian yang lebih stabil selama proses pengambilan keputusan. Hasil ini menunjukkan bahwa penerapan *data augmentation* dan *class weight* secara bersamaan membantu model mempelajari representasi fitur yang lebih terarah, sehingga perhatian model lebih terpusat pada area anatomi yang berkontribusi terhadap klasifikasi kategori *Non Demented*.



Gambar 4. 20 Grad-CAM Model ResNet50 pada Data dengan Augmentasi dan Class Weight

4.4.3 Analisis Komparatif Model EfficientNet-B0 dan ResNet50

Untuk memperjelas perbedaan karakteristik dan performa kedua model secara lebih komprehensif dan terstruktur, disajikan tabel perbandingan yang merangkum tidak hanya hasil evaluasi kuantitatif, seperti *accuracy* dan *recall*, tetapi juga hasil interpretasi visual menggunakan metode Grad-CAM. Tabel 4.5 ini dirancang untuk memberikan gambaran menyeluruh mengenai bagaimana masing-masing model bekerja, baik dari sisi kinerja numerik maupun dari sisi kemampuan dalam mengidentifikasi dan memfokuskan area penting pada citra, sehingga perbedaan pola pembelajaran dan pengambilan keputusan antar model dapat dianalisis secara lebih mendalam.

Tabel 4. 5 Tabel Perbandingan EfficientNet-B0 dan ResNet50

No	Aspek Evaluasi	EfficientNet-B0	ResNet50
1.	Stabilitas Proses Pelatihan	Menunjukkan kurva loss dan <i>accuracy</i> yang lebih stabil serta konvergen pada sebagian besar skenario	Cenderung fluktuatif, terutama pada <i>validation loss</i> dan <i>validation accuracy</i>

No	Aspek Evaluasi	EfficientNet-B0	ResNet50
2.	Sensitivitas terhadap <i>Hyperparameter</i>	Relatif lebih <i>robust</i> , dengan perubahan performa yang tidak terlalu signifikan antara kondisi <i>baseline</i> dan optimal	Sangat sensitif, performa meningkat signifikan hanya pada kombinasi <i>hyperparameter</i> tertentu
3.	Kemampuan Generalisasi	Baik, terutama pada skenario data augmentasi yang menghasilkan kurva validasi yang stabil	Kurang stabil pada data original, namun meningkat setelah penerapan augmentasi dan <i>class weight</i>
4.	Risiko <i>Overfitting</i>	Lebih rendah, ditunjukkan oleh selisih kecil antara <i>training</i> dan <i>validation</i>	Tinggi pada data original, dengan gap besar antara <i>training</i> dan <i>validation</i>
5.	Performa Kuantitatif	Konsisten tinggi pada berbagai skenario (<i>validation accuracy</i> hingga 0,965)	Dapat mencapai performa tertinggi (0,97), namun tidak konsisten antar skenario
6.	Sensitivitas terhadap Kelas Minoritas	Cukup baik, meningkat dengan augmentasi dan <i>class weight</i>	Meningkat signifikan dengan <i>class weight</i> (<i>recall</i> hingga 0,93)
7.	Interpretasi Grad-CAM	Aktivasi lebih konsisten dan terlokalisasi pada area citra yang relevan terhadap prediksi model	Skenario awal menunjukkan indikasi shortcut learning, namun fokus aktivasi menjadi lebih terlokalisasi setelah penerapan augmentasi dan <i>class weight</i>
8.	Kualitas Interpretasi Klinis	Menunjukkan pola aktivasi yang lebih konsisten dan terlokalisasi pada area utama citra	Meningkat setelah augmentasi dan <i>class weight</i> , namun masih terdapat sedikit <i>noise</i>
9.	Kebutuhan Teknik Tambahan	Sudah cukup optimal dengan augmentasi	Memperoleh peningkatan kualitas fokus aktivasi setelah penerapan augmentasi dan <i>class weight</i> .
10.	Kesimpulan Umum	Lebih unggul dalam stabilitas, generalisasi, dan interpretabilitas	Unggul pada performa maksimum, namun memerlukan <i>tuning</i> yang lebih kompleks

Hasil evaluasi kuantitatif dan interpretasi visual menggunakan Grad-CAM, terdapat perbedaan mendasar antara model EfficientNet-B0 dan ResNet50 dalam memahami pola citra medis. EfficientNet-B0 terbukti lebih tangguh (*robust*) terhadap variasi *hyperparameter* dan mampu menghasilkan kurva pelatihan yang stabil dengan risiko

overfitting yang minim. Sebaliknya, meskipun ResNet50 mampu meraih skor *validation accuracy* puncak (0,97), arsitektur ini menuntut konfigurasi parameter yang sangat spesifik dan lebih rentan terhadap fluktuasi selama proses pelatihan. Meskipun kedua model menunjukkan performa yang tinggi berdasarkan metrik seperti *accuracy*, *precision*, dan *recall*, analisis lebih lanjut mengungkap bahwa nilai tersebut tidak selalu mencerminkan kemampuan model dalam menangkap fitur yang relevan secara klinis.

Model ResNet50, khususnya pada skenario data original, visualisasi Grad-CAM menunjukkan adanya fenomena *shortcut learning* (Geirhos dkk., 2020), di mana model cenderung memanfaatkan pola latar belakang sebagai dasar prediksi. Hal ini menyebabkan model memperoleh *validation accuracy* yang tinggi, namun tanpa didukung oleh lokalisasi fitur anatomis yang benar. Dengan kata lain, performa yang tinggi tidak sepenuhnya didukung oleh fokus aktivasi yang sesuai dan tidak dapat dipertanggungjawabkan dalam konteks medis. Meskipun penerapan augmentasi dan *class weight* mampu memperbaiki fokus model secara signifikan, masih ditemukan sedikit aktivasi pada area non-anatomis yang menunjukkan bahwa model belum sepenuhnya bebas dari pengaruh *noise*.

Sebaliknya, model EfficientNet-B0 menunjukkan konsistensi yang lebih baik dalam mengarahkan perhatian pada struktur otak yang relevan di berbagai skenario. Pada kondisi data original, model memang masih menunjukkan penyebaran aktivasi yang cukup luas. Namun, setelah diterapkan augmentasi dan kombinasi augmentasi dengan *class weight*, model mampu menghasilkan aktivasi yang lebih terlokalisasi, stabil, dan mengikuti kontur anatomi otak. Bahkan, pada skenario terbaik, model tidak hanya berfokus pada area kortikal seperti lobus frontal, tetapi juga mampu menunjukkan fokus aktivasi pada area di sekitar ventrikel yang diketahui sering digunakan dalam studi *neuroimaging* Alzheimer (Centro dkk., 2011).

Perbedaan kemampuan ini tidak lepas dari karakteristik fundamental masing-masing arsitektur. Keunggulan EfficientNet-B0 dalam memfilter *noise* dan memusatkan perhatian pada fitur anatomis sangat dipengaruhi oleh penggunaan blok *Mobile Inverted Bottleneck Convolution* (MBConv) yang dilengkapi dengan modul *Squeeze-and-Excitation* (SE). Modul ini bertindak sebagai mekanisme *attention* yang secara adaptif memberikan bobot lebih pada fitur (saluran) yang informatif, sekaligus menekan fitur latar belakang yang tidak

relevan. Di sisi lain, arsitektur standar ResNet50 tidak memiliki mekanisme *attention* bawaan ini, sehingga lebih rentan mengekstraksi fitur yang salah.

Dengan demikian, dapat disimpulkan bahwa EfficientNet-B0 lebih unggul dibandingkan ResNet50, tidak hanya dari sisi stabilitas dan kemampuan generalisasi, tetapi juga dalam hal interpretabilitas dan akuntabilitas klinis. Temuan ini menegaskan bahwa dalam pengembangan sistem klasifikasi berbasis *deep learning* untuk domain medis, pemilihan model tidak boleh hanya didasarkan pada performa kuantitatif semata, tetapi juga harus mempertimbangkan kemampuan model dalam mempelajari dan memfokuskan perhatian pada fitur yang benar-benar relevan secara anatomis. Oleh karena itu, pendekatan *Explainable AI* seperti Grad-CAM menjadi komponen penting dalam memastikan bahwa keputusan model dapat dipercaya dan sesuai dengan pengetahuan medis yang berlaku (Selvaraju dkk., 2017).