

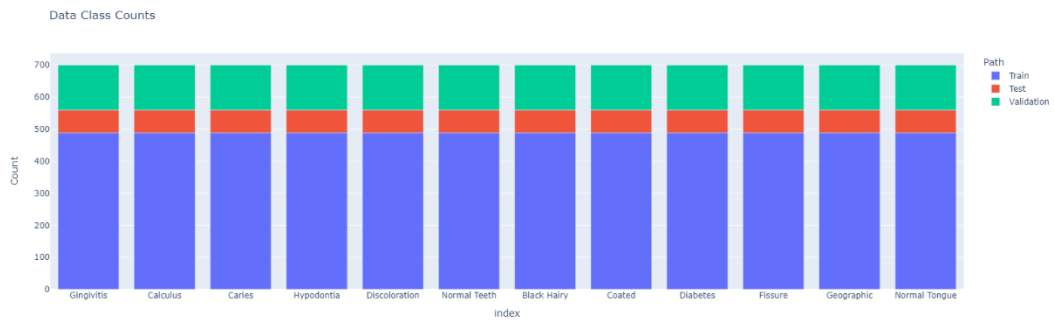
BAB IV HASIL DAN PEMBAHASAN

4.1. Splitting Dataset

Pada penelitian ini dilakukan pembagian dataset menggunakan library dari `splitfolders` dengan proporsi 70% data train, 20% data validation, dan 10% data test. Total data train berjumlah 5.868, data validation 1.680, dan data test 852. Hasil seluruh data setelah dilakukan splitting dapat dilihat pada tabel 7 dan gambar 7.

Tabel 7. Hasil Data Splitting

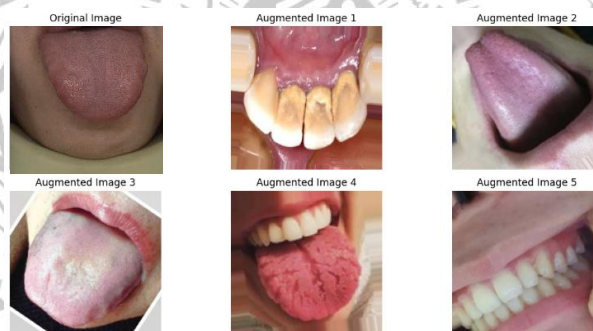
<i>Class</i>	<i>Train</i>	<i>Validation</i>	<i>Test</i>	<i>Total</i>
<i>Caries</i>	489	140	71	700
<i>Calculus</i>	489	140	71	700
<i>Discoloration</i>	489	140	71	700
<i>Gingivitis</i>	489	140	71	700
<i>Hypodontia</i>	489	140	71	700
<i>Normal Teeth</i>	489	140	71	700
<i>Fissure</i>	489	140	71	700
<i>Geographic</i>	489	140	71	700
<i>Diabetic</i>	489	140	71	700
<i>Coated</i>	489	140	71	700
<i>Black Hairy</i>	489	140	71	700
<i>Normal Tongue</i>	489	140	71	700
<i>Total</i>	5.868	1.680	852	



Gambar 7. Grafik Hasil Data Splitting

4.2. Augmentasi Data

Proses augmentasi data digunakan untuk menyeimbangkan jumlah citra setiap kelas yang ada dalam dataset. Proses ini dapat memperbanyak citra dengan tetap mempertahankan inti dari setiap citra. Augmentasi diimplementasikan pada data train dan validation untuk mencegah terjadinya *overfitting* model yang dibangun. Hasil data setelah dilakukan proses augmentasi dapat dilihat pada gambar 8.



Gambar 8. Hasil Augmentasi Data

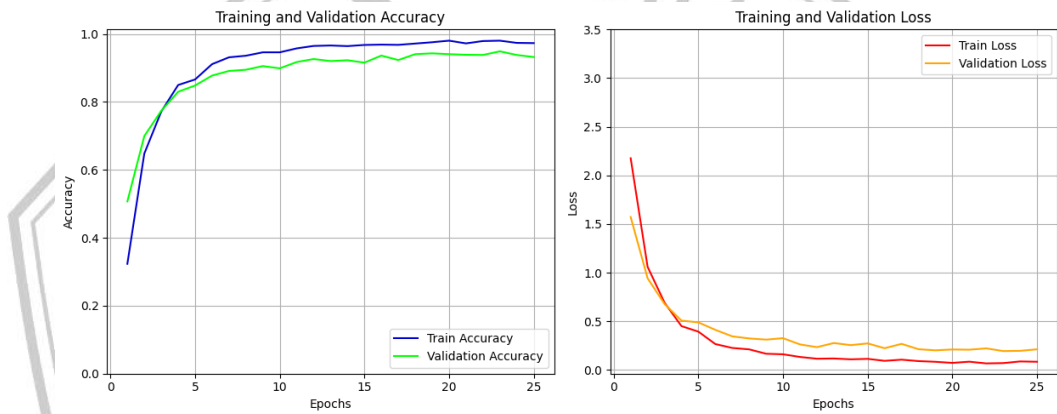
4.3. Pengujian Data

Pada tahapan ini dilakukan pengujian data menggunakan 3 skenario untuk mengetahui pengaruh augmentasi data terhadap hasil kinerja model yang telah diusulkan sebelumnya seperti VGG16, VGG19, dan ResNet50. Dataset yang dikumpulkan bersumber dari situs Kaggle dan Roboflow dengan total data berjumlah 8400 yang terbagi dalam 12 kelas, yaitu 6 kelas untuk gigi dan 6 kelas untuk lidah. Semua skenario pengujian dilakukan dengan menerapkan optimizer SGD, learning rate 0.0001, momentum 0.9, loss “categorical_crossentropy”, batch

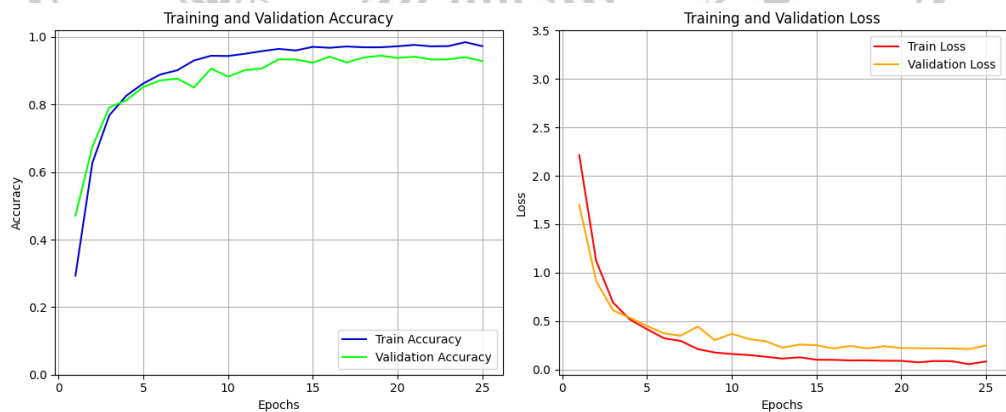
size 32, dan step_per_epoch 100. Hasil pengujian dilakukan untuk membandingkan kinerja paling optimal dari beberapa model yang telah diusulkan, yaitu VGG16, VGG19, dan ResNet50.

4.3.1. Skenario Pengujian 1

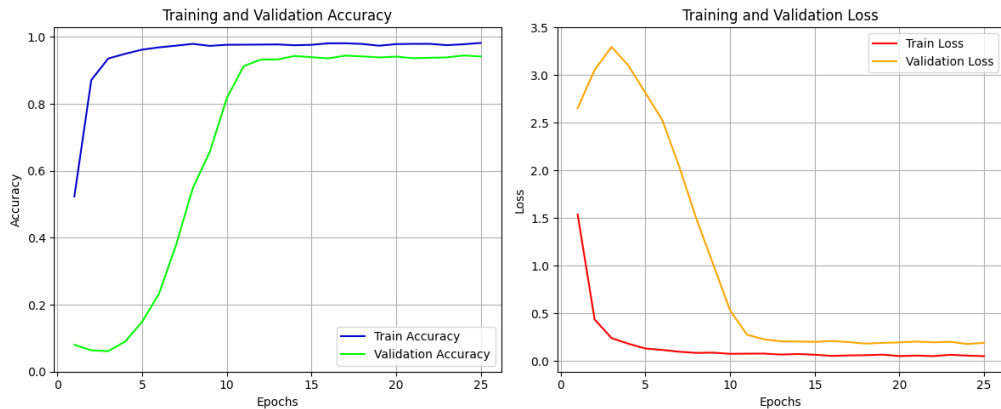
Pada skenario pertama, model yang diusulkan diuji dengan menggunakan dataset original atau tanpa melalui proses augmentasi dengan jumlah epoch sebesar 25. Grafik hasil *train acc* dan *loss* dari pengujian skenario pertama dapat dilihat pada gambar 9, 10, dan 11.



Gambar 9. Grafik Accuracy dan Loss Model VGG16 Skenario 1



Gambar 10. Grafik Accuracy dan Loss Model VGG19 Skenario 1

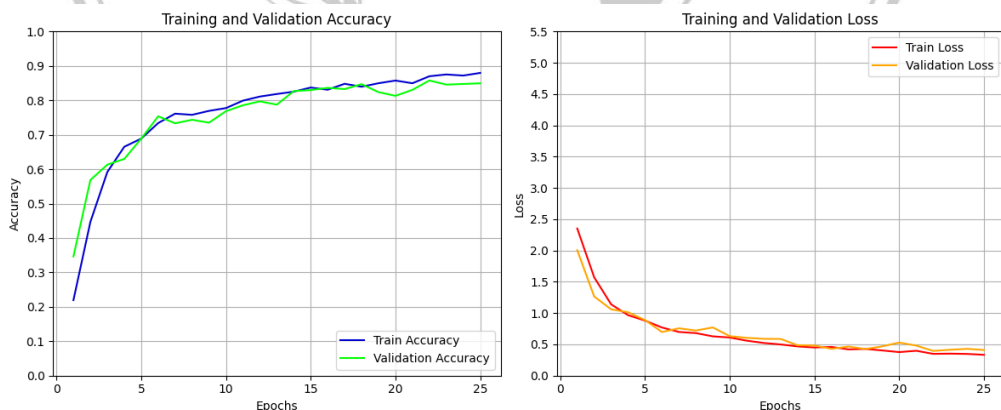


Gambar 11. Grafik Accuracy dan Loss Model ResNet50 Skenario 1

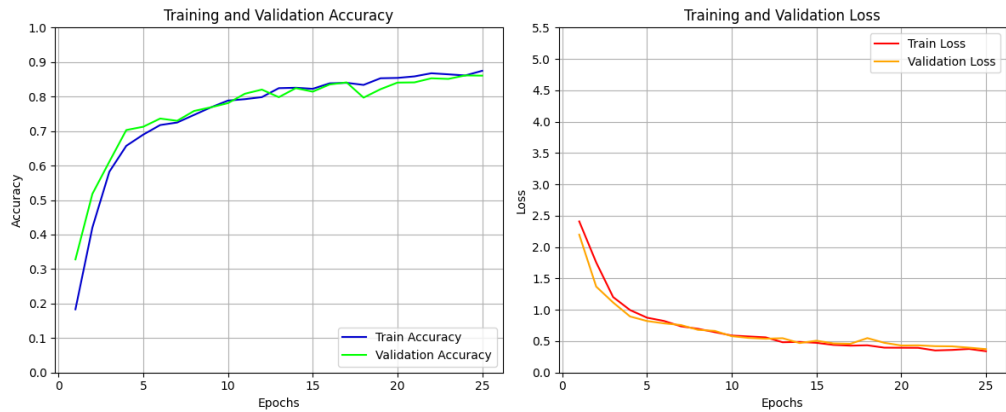
Berdasarkan hasil gambar grafik diatas pada pengujian skenario 1, kedua model VGG, baik VGG16 maupun VGG19 mendapatkan hasil yang stabil pada kurva *train* dan *loss*. Sedangkan model ResNet50 pada awal proses pelatihan terdapat gap antara kurva *train* dan *loss* atau *overfit*, namun pada pertengahan proses, kurva dapat mendekat hingga akhir epoch pengujian.

4.3.2. Skenario Pengujian 2

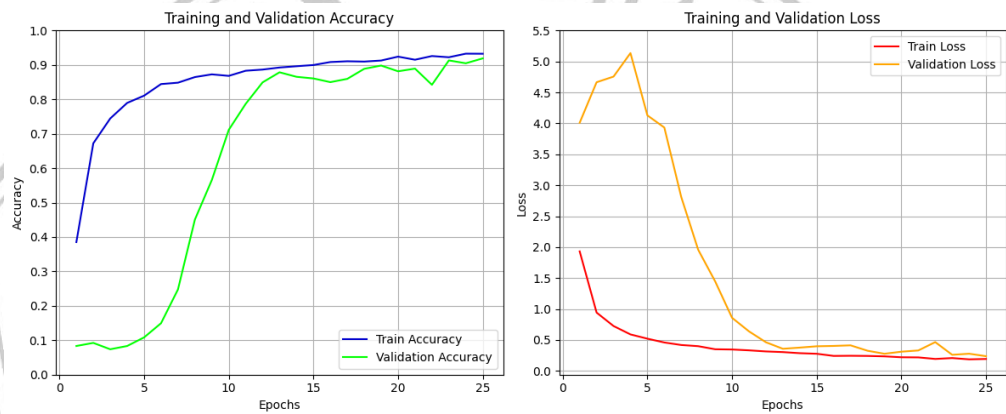
Pada skenario pengujian kedua, dilakukan dengan menerapkan proses augmentasi pada seluruh dataset dengan jumlah epoch sebesar 25 yang diujikan pada setiap model, seperti VGG16, VGG19, dan ResNet50. Grafik hasil *train acc* dan *loss* dari pengujian skenario kedua dapat dilihat pada gambar 12, 13, dan 14.



Gambar 12. Grafik Accuracy dan Loss Model VGG16 Skenario 2



Gambar 13. Grafik Accuracy dan Loss Model VGG19 Skenario 2



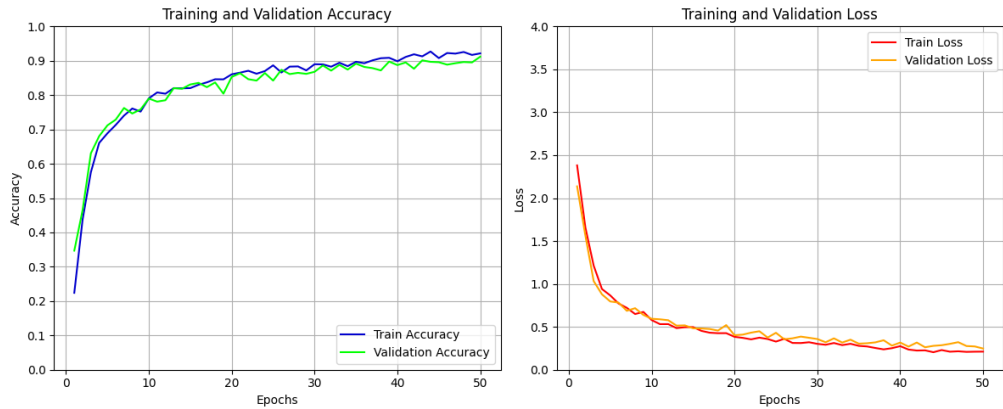
Gambar 14. Grafik Accuracy dan Loss Model ResNet50 Skenario 2

Pengujian pada skenario 2 dilakukan dengan data yang sudah melalui proses augmentasi, didapatkan kemiripan grafik dengan hasil pada skenario 1. Model VGG16 dan VGG19 memiliki grafik kurva yang lebih stabil baik dari awal proses pelatihan hingga akhir dibandingkan model ResNet50 yang pada awal proses pelatihan mengalami *overfit* yang cukup tinggi, namun pada epoch ke 10 hingga akhir kurva dapat stabil.

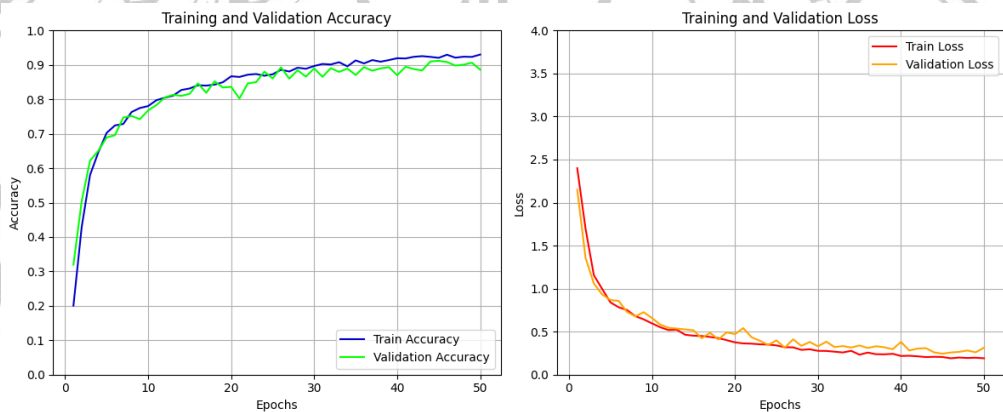
4.3.3 Skenario Pengujian 3

Pada skenario ketiga, pengujian dilakukan dengan menerapkan proses augmentasi pada seluruh dataset serta jumlah epoch yang ditingkatkan sebanyak 2 kali sebesar 50 dan 75. Grafik hasil *train acc* dan *loss* terhadap

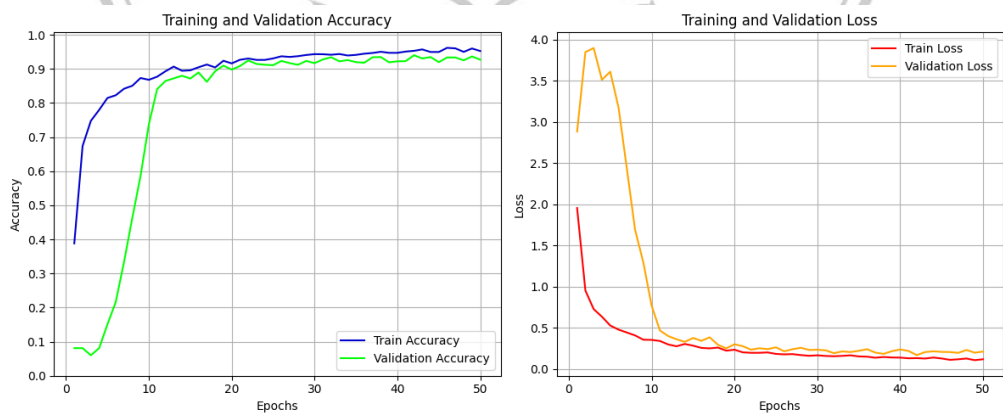
setiap model yang diterapkan dari pengujian skenario ketiga dapat dilihat pada gambar 15, 16, 17, 18, 19, dan 20.



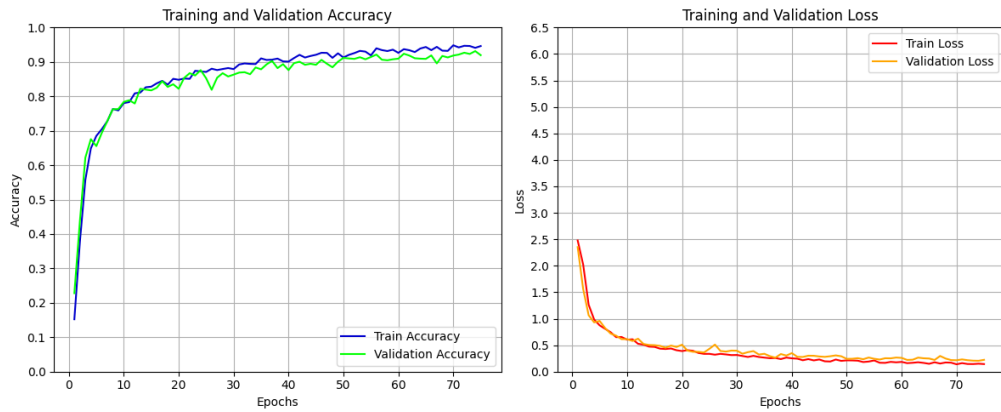
Gambar 15. Grafik Accuracy dan Loss VGG16 Epoch 50 Skenario 3



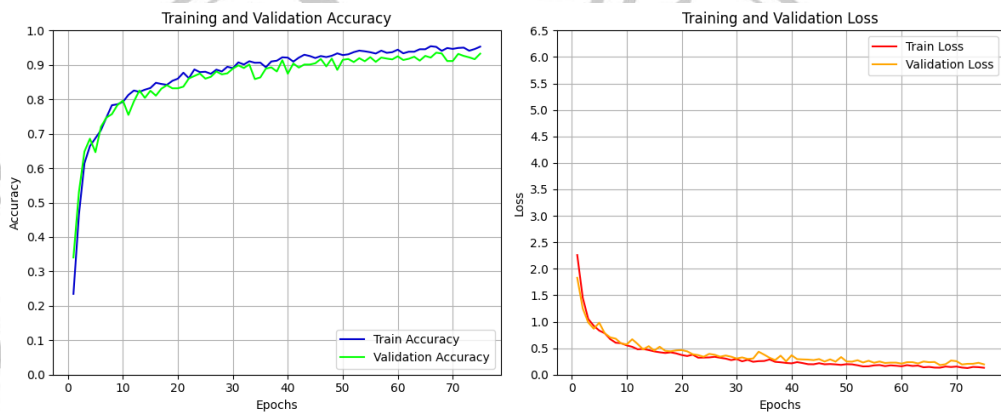
Gambar 16. Grafik Accuracy dan Loss VGG19 Epoch 50 Skenario 3



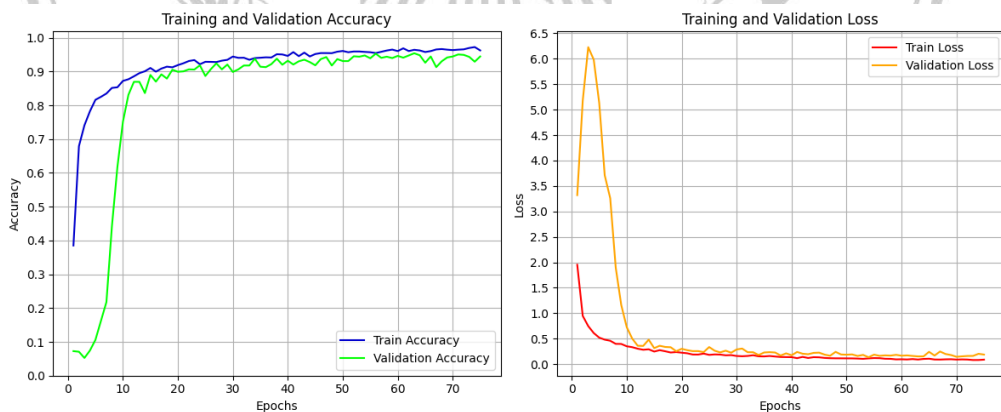
Gambar 17. Grafik Accuracy dan Loss ResNet50 Epoch 50 Skenario 3



Gambar 18. Grafik Accuracy dan Loss VGG16 Epoch 75 Skenario 3



Gambar 19. Grafik Accuracy dan Loss VGG19 Epoch 75 Skenario 3



Gambar 20. Grafik Accuracy dan Loss ResNet50 Epoch 75 Skenario 3

Pada skenario 3, terdapat 2 pengujian berbeda dimana pengujian pertama dilakukan dengan proses pelatihan 50 epoch dan pada pengujian

kedua menggunakan 75 epoch. Berdasarkan hasil yang diperoleh pada pengujian pertama dengan 50 epoch, ketiga model mendapatkan hasil yang cukup baik, namun pada model ResNet50 masih terdapat *overfit* pada awal proses pelatihan. Pada pengujian kedua dengan 75 epoch, didapatkan hasil lebih baik dari semua skenario yang telah diujikan terhadap setiap model. Namun pada pengujian kali ini model ResNet 50 mengalami overfit yang paling tinggi dari semua skenario yang diujikan.

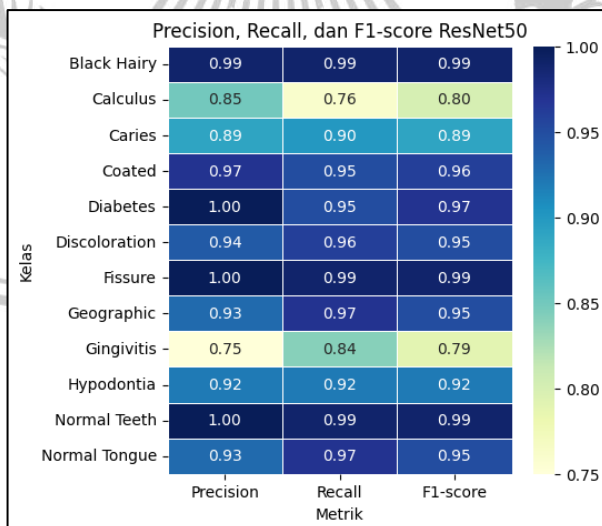
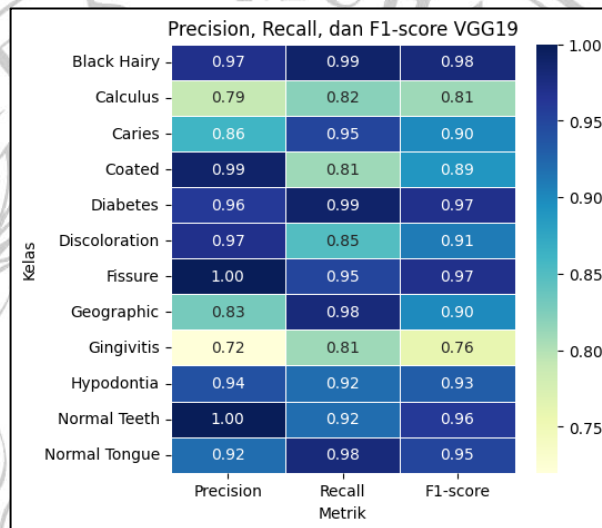
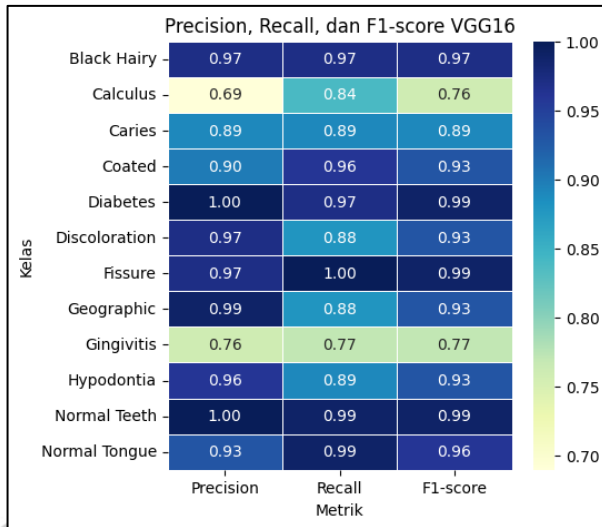
Berdasarkan hasil grafik train dan loss yang didapatkan pada semua skenario pengujian, ResNet cenderung mengalami overfitting pada awal proses pelatihan. Hal ini karena struktur yang dalam dengan banyak layer dapat menyebabkan model lebih mudah untuk mempelajari detail-detail kecil dari data latih yang kompleks, bahkan tidak relevan, dimana mengakibatkan penyesuaian yang berlebihan pada data tersebut. Namun, ketika jumlah epoch bertambah, ResNet mampu menyesuaikan dan menangkap pola-pola yang lebih kompleks, sehingga kinerjanya menjadi lebih baik dan *overfitting* menurun. Sementara itu, VGG cenderung lebih stabil pada awal proses pelatihan karena memiliki struktur yang lebih sederhana dibandingkan ResNet, sehingga tidak rentan terhadap *overfitting*. Namun, meskipun stabil, hasil akhir VGG ternyata masih sedikit di bawah atau setara dengan ResNet, menunjukkan bahwa meskipun mengalami *overfitting* pada awalnya, ResNet mampu mengejar dan bahkan melampaui kinerja VGG pada akhir proses pelatihan karena kemampuannya untuk menangkap pola yang lebih kompleks.

4.4. Evaluasi Hasil

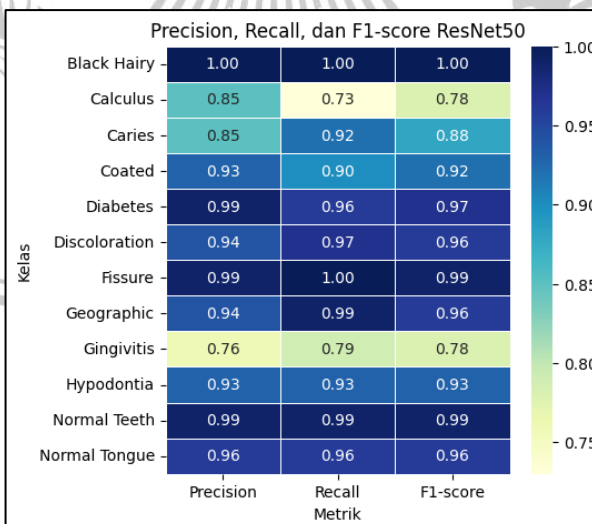
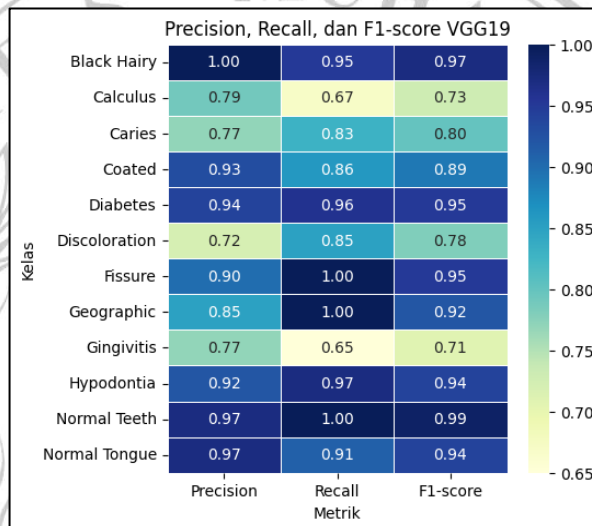
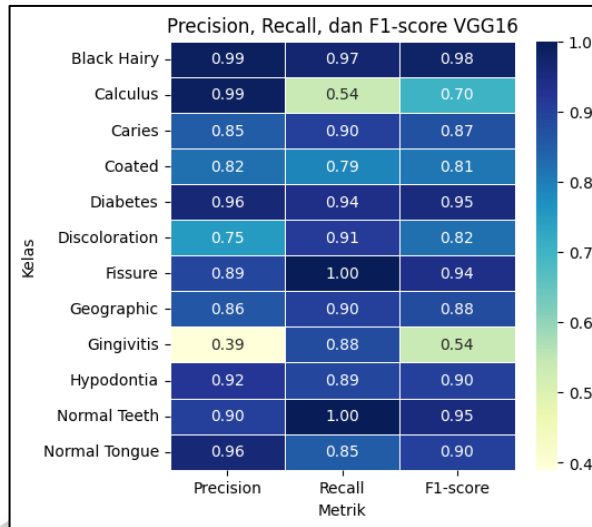
Pada penelitian ini hasil pengujian dievaluasi dengan menggunakan metode classification report untuk mengetahui performa dari beberapa model yang telah diujikan dalam semua skenario.

4.4.1. Classification Report

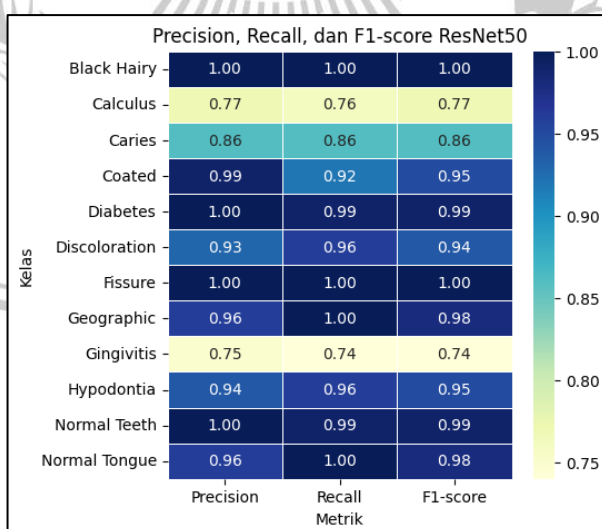
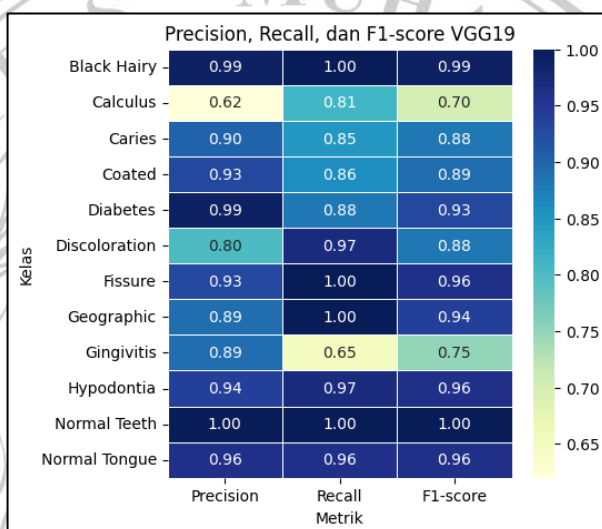
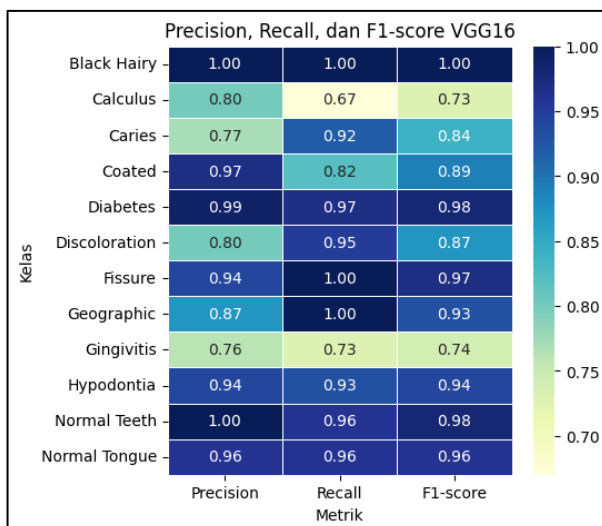
Hasil classification report yang didapatkan pada pengujian skenario 1, 2, dan 3 untuk setiap model, yaitu VGG16, VGG19, dan ResNet50 dapat dilihat pada gambar 21, 22, 23 dan 24.



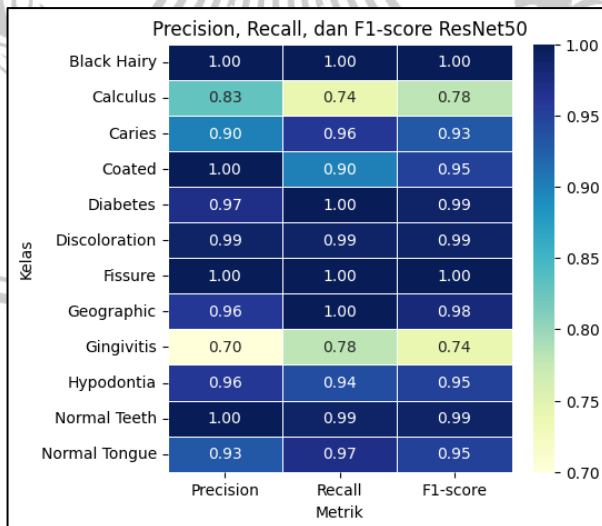
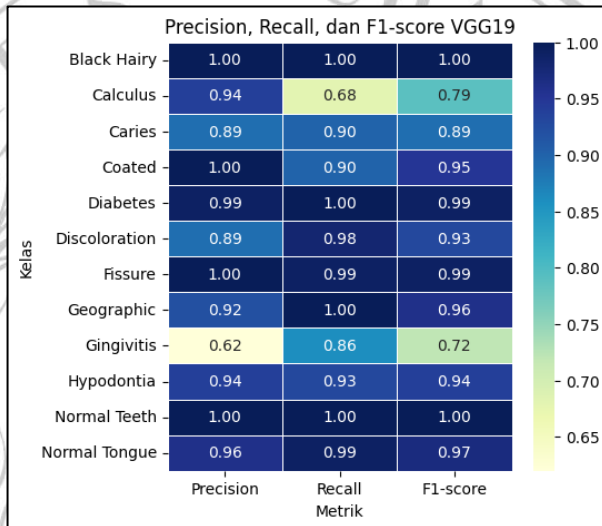
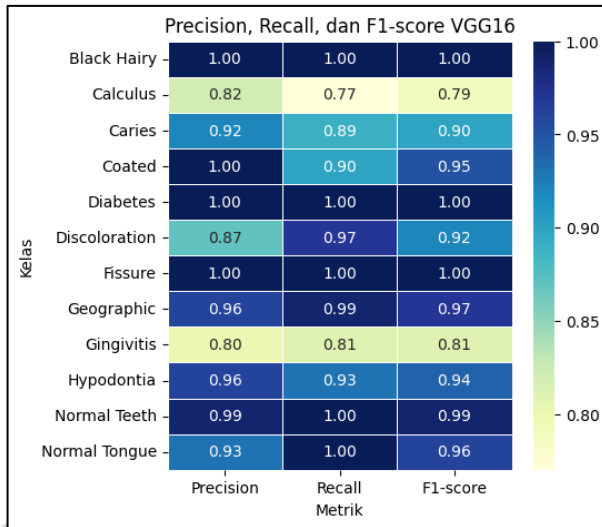
Gambar 21. Classification Report Skenario 1



Gambar 22. Classification Report Skenario 2



Gambar 23. Classification Report Skenario 3 (50 Epoch)



Gambar 24. Classification Report Skenario 3 (75 Epoch)

Berdasarkan classification report hasil terbaik didapatkan oleh skenario 3 dengan pengujian 75 epoch yang dapat dilihat pada gambar 24. Dimana hampir semua nilai *precision*, *recall*, dan *f1-score* berhasil mendapatkan angka diatas 0.90, namun terdapat 2 kelas dalam dataset yang selalu mendapatkan hasil kurang memuaskan yaitu Calculus dan Gingivitis. 2 kelas tersebut dalam setiap skenario pengujian terhadap semua model selalu mendapatkan hasil paling rendah dari kelas lain dalam dataset. Hal ini kemungkinan karena citra dalam 2 kelas tersebut atau penyakit Calculus dan Gingivitis terlihat mirip, sehingga model terlalu sulit dalam membedakan 2 kelas tersebut.

4.5. Perbandingan Hasil

Pada tahapan ini dilakukan perbandingan hasil dari semua skenario yang diujikan pada seluruh model yang diusulkan seperti, VGG16, VGG19, dan ResNet50. Perbandingan yang dilakukan ditujukan untuk mengetahui perbedaan performa setiap model dalam masing-masing skenario yang diujikan.

4.5.1 Perbandingan Augmentasi & Non Augmentasi

Berikut ini merupakan tabel hasil perbandingan dari setiap model yang diujikan dengan data melalui proses augmentasi maupun tanpa augmentasi dan jumlah epoch sebesar 25. Hasil dapat dilihat pada tabel 8.

Tabel 8. Perbandingan Non Augmentasi & Augmentasi

	Model	Precision	Recall	F1-Score	Accuraccy
Non Augmentasi	VGG16	0.92	0.92	0.92	0.92
	VGG19	0.91	0.91	0.91	0.91
	ResNet50	0.93	0.93	0.93	0.93
Augmentasi	VGG16	0.86	0.88	0.85	0.85
	VGG19	0.88	0.89	0.88	0.88
	ResNet50	0.93	0.93	0.93	0.93

Dari hasil perbandingan performa model dengan dan tanpa augmentasi pada Tabel 9, dapat dilihat bahwa model ResNet50 memberikan hasil terbaik di antara ketiga model yang diujikan. Dengan nilai *precision*, *recall*, *F1-Score*, dan *accuracy* yang sama yaitu 0.93. Hal ini menunjukkan bahwa ResNet50 memiliki konsistensi performa yang tinggi dalam kedua kondisi. Di sisi lain, model VGG16 dan VGG19 mengalami penurunan performa setelah dilakukan augmentasi, ditunjukkan dengan nilai metrik yang lebih rendah dibandingkan tanpa augmentasi. Penurunan ini dapat disebabkan oleh kompleksitas tambahan dari data hasil augmentasi, dimana mungkin memerlukan lebih banyak epoch untuk mencapai performa optimal.

4.5.2 Perbandingan Augmentasi Dengan Meningkatkan Epoch

Selanjutnya merupakan perbandingan hasil dari pengujian terhadap setiap model dengan data yang sudah dilakukan proses augmentasi. Dimana pada setiap pengujian, epoch ditingkatkan menjadi 25, 50, dan 75. Hasil perbandingan dapat dilihat pada tabel 9.

Tabel 9. Perbandingan Model Setiap Peningkatan Epoch

	Model	Precision	Recall	F1-Score	Accuracy
25 Epoch	VGG16	0.86	0.88	0.85	0.85
	VGG19	0.88	0.89	0.88	0.88
	ResNet50	0.93	0.93	0.93	0.93
50 Epoch	VGG16	0.90	0.91	0.90	0.90
	VGG19	0.90	0.91	0.90	0.90
	ResNet50	0.93	0.93	0.93	0.93
75 Epoch	VGG16	0.94	0.94	0.94	0.94
	VGG19	0.93	0.94	0.93	0.93
	ResNet50	0.94	0.94	0.94	0.94

Pada Tabel 10, dengan peningkatan jumlah epoch dari 25 hingga 75, terlihat bahwa semua model mengalami peningkatan performa seiring

bertambahnya epoch. Model VGG16 dan VGG19 menunjukkan peningkatan yang signifikan dalam semua metrik ketika epoch ditingkatkan. VGG16 mencapai performa terbaik pada 75 epoch dengan nilai *precision*, *recall*, *F1-Score*, dan *accuracy* sebesar 0.94 sedangkan VGG 19 memperoleh nilai *precision* sebesar 0.93, *recall* sebesar 0.94, *F1-Score* sebesar 0.93, dan *accuracy* sebesar 0.94. Hal ini menunjukkan bahwa kedua model tersebut membutuhkan lebih banyak epoch untuk mencapai performa optimal pada data yang diaugmentasi. Di sisi lain, ResNet50 mempertahankan performa stabil dengan nilai metrik yang konsisten dari 25 hingga 75 epoch, menunjukkan bahwa model ini sudah cukup optimal dengan 25 epoch dan augmentation tidak memberikan perbaikan signifikan lebih lanjut.

