

## BAB IV HASIL DAN PEMBAHASAN

### 4.1 Data Description

Data yang digunakan pada penelitian ini merupakan data citra kanker kulit yang didapat dari International Skin Imaging Collaboration Archive. Data citra kanker kulit ini sebanyak 5.500 kelas benign dan 5.105 kelas malignant. Sampel data citra benign dan malignant ditunjukkan pada Gambar 4.1 dan Gambar 4.2 berikut.



Gambar 4.1 *Sample Data citra benign*



Gambar 4.2 *Sample Data Citra Malignant*

### 4.2 Preprocessing Data

Sebagaimana yang telah dipaparkan pada bab 3, *preprocessing data* dilakukan dengan 3 tahap *preprocessing*, yaitu *resize*, *normalization*, dan *augmentation data*.

Pada proses *resize* semua data citra diubah pada ukuran 256x256 dimana satu data citra memiliki tiga lapisan warna yaitu merah (R), hijau (G), dan biru (B). Masing-masing lapisan memiliki nilai piksel dengan intensitas warna yang berkisar antara 0 sampai 255. Nilai piksel lapisan R, G, dan B ditunjukkan pada tabel 4.1, tabel 4.2, tabel 4.3 sebagai contoh sampel dari gambar benign saja.

Tabel 4.1 Nilai Piksel *Benign* pada Lapisan *Red*

	1	2	3	4	5	...	253	254	255	256
1	156	157	160	163	165	...	166	165	163	161
2	156	157	160	163	165	...	167	166	164	162
3	156	158	160	163	165	...	169	167	165	163
4	157	158	161	163	166	...	170	169	166	163
5	157	159	161	164	166	...	171	169	166	163
...	...	...	...	...	...	...	...	...	...	...
253	147	148	153	155	158	...	160	159	155	153
254	146	148	153	156	158	...	163	159	153	153
255	145	147	153	157	159	...	163	157	152	153
256	145	147	153	157	160	...	162	155	152	155

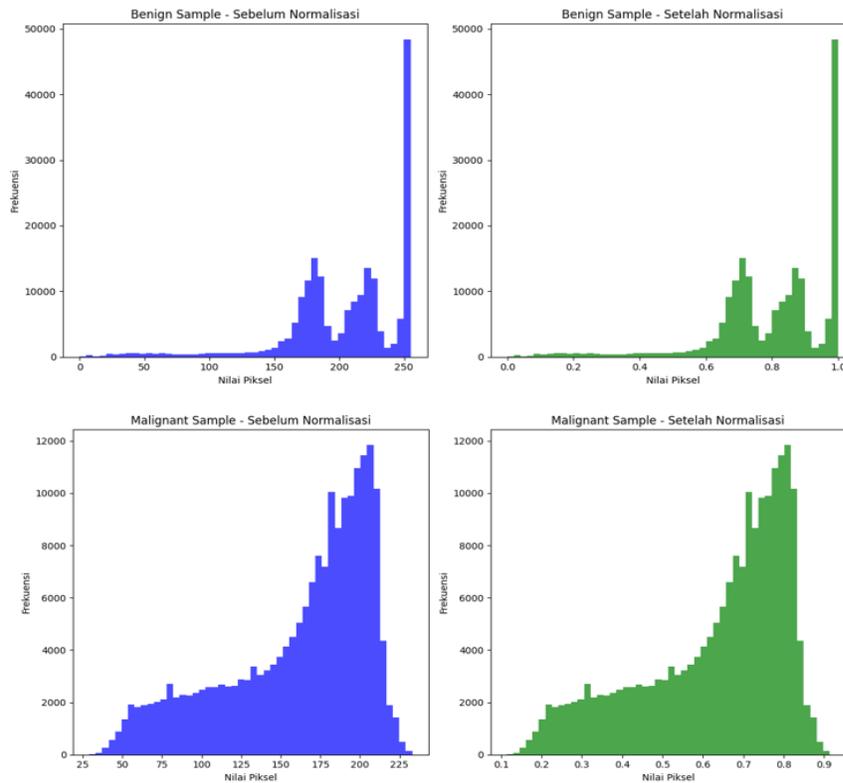
Tabel 4.2 Nilai Piksel *Benign* pada Lapisan *Green*

	1	2	3	4	5	...	253	254	255	256
1	121	122	125	128	130	...	131	130	128	126
2	121	122	125	128	130	...	132	131	129	127
3	121	123	125	128	130	...	134	132	130	128
4	121	123	126	128	131	...	135	134	131	128
5	121	124	126	129	131	...	136	134	131	128
...	...	...	...	...	...	...	...	...	...	...
253	120	121	124	126	129	...	131	130	128	126
254	119	121	124	127	129	...	134	130	126	126
255	118	120	124	128	130	...	134	128	125	126
256	118	120	124	128	131	...	133	126	125	128

Tabel 4.3 Nilai Pixel *Benign* pada Lapisan *Blue*

	1	2	3	4	5	...	253	254	255	256
1	91	92	95	98	100	...	101	100	98	96
2	91	92	95	98	100	...	102	101	99	97
3	91	93	95	98	100	...	104	102	100	98
4	92	93	96	98	101	...	105	104	101	98
5	92	94	96	99	101	...	106	104	101	98
...	...	...	...	...	...	...	...	...	...	...
253	90	91	94	96	99	...	101	100	98	96
254	89	91	94	97	99	...	104	100	96	96
255	88	90	94	98	100	...	104	98	95	96
256	88	90	94	98	101	...	103	96	95	98

Kemudian setelah semua data ukurannya telah diubah menjadi 256x256, maka selanjutnya proses normalisasi data, yang dilakukan dengan mengubah skala nilai piksel dalam rentang 0-255. Berikut gambar distribusi pixel sebelum dan sesudah normalisasi.



Gambar 4.3 Pixel *Distribution Benign* dan *Malignant* Sebelum dan Sesudah *Normalization*

Gambar tersebut menunjukkan distribusi nilai piksel dari kumpulan gambar citra benign dan malignant sebelum dan sesudah normalisasi. Dapat dilihat pola distribusi pada gambar tersebut tetap sama sebelum dan sesudah di normalisasi, hanya saja dapat dilihat bahwa nilai piksel telah berubah ke skala yang lebih kecil dan grafik yang menunjukkan puncak yang tinggi pada nilai tertentu, yang ditunjukkan oleh dominasi warna tertentu dalam kumpulan data benign dan malignant yang beragam.

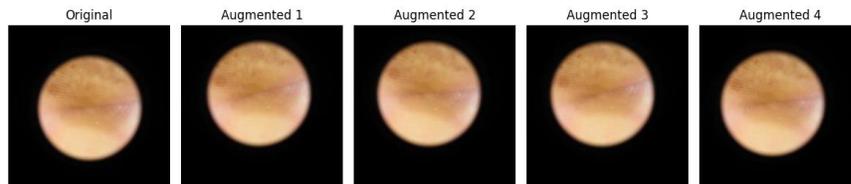
Selanjutnya setelah normalisasi data, maka proses berikutnya yaitu augmentation data. Proses ini untuk memperbanyak variasi data citra dengan memodifikasi data citra asli. Pada penelitian ini metode augmentasi data yang digunakan yaitu rotasi dan menggeser gambar secara horizontal dan vertikal sebesar 0.1. Berikut gambar sampel benign dan malignant pada citra asli dan citra yang telah di augmentasi.

Benign Augmentation Example



Gambar 4.4 *Sample* Citra Asli dan Augmentasi pada Kelas *Benign*

Malignant Augmentation Example



Gambar 4.5 *Sample* Citra Asli dan Augmentasi pada Kelas *Malignant*

## 4.3 CNN Model Implementation

### 4.3.1 Model Architecture

Setelah preprocessing data selesai maka untuk arsitektur model CNN yang diusulkan pada penelitian ini terdapat dua model untuk dilakukan analisis model pertama yaitu model dengan kombinasi max pooling dan global average pooling. Berikut gambar model summary kombinasi max pooling dan global average pooling.

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 256, 256, 32)	896
batch_normalization (BatchNormalization)	(None, 256, 256, 32)	128
conv2d_1 (Conv2D)	(None, 256, 256, 32)	9,248
batch_normalization_1 (BatchNormalization)	(None, 256, 256, 32)	128
max_pooling2d (MaxPooling2D)	(None, 128, 128, 32)	0
dropout (Dropout)	(None, 128, 128, 32)	0
conv2d_2 (Conv2D)	(None, 128, 128, 64)	18,496
batch_normalization_2 (BatchNormalization)	(None, 128, 128, 64)	256
conv2d_3 (Conv2D)	(None, 128, 128, 64)	36,928
batch_normalization_3 (BatchNormalization)	(None, 128, 128, 64)	256
max_pooling2d_1 (MaxPooling2D)	(None, 64, 64, 64)	0
dropout_1 (Dropout)	(None, 64, 64, 64)	0
conv2d_4 (Conv2D)	(None, 64, 64, 128)	73,856
batch_normalization_4 (BatchNormalization)	(None, 64, 64, 128)	512
conv2d_5 (Conv2D)	(None, 64, 64, 128)	147,584
batch_normalization_5 (BatchNormalization)	(None, 64, 64, 128)	512
max_pooling2d_2 (MaxPooling2D)	(None, 32, 32, 128)	0
dropout_2 (Dropout)	(None, 32, 32, 128)	0
global_average_pooling2d (GlobalAveragePooling2D)	(None, 128)	0
dense (Dense)	(None, 256)	33,024
dropout_3 (Dropout)	(None, 256)	0
dense_1 (Dense)	(None, 1)	257

Gambar 4.6 Model Summary Kombinasi Max Pooling dan Global Average Pooling

Model dimulai dengan lapisan convolutional layer pertama (Conv2D) dengan 32 filter dan ukuran kernel 3x3, menggunakan fungsi aktivasi ReLU dan diikuti oleh lapisan batch normalization untuk menstabilkan proses pelatihan. Kemudian, lapisan konvolusi kedua (Conv2D) dengan konfigurasi yang sama diterapkan, diikuti oleh batch normalization lagi untuk lebih lanjut meningkatkan stabilitas. Setelah dua lapisan konvolusi ini, digunakan lapisan max pooling (MaxPooling2D) dengan ukuran pool 2x2 untuk mengurangi dimensi spasial dari fitur yang diekstraksi, serta lapisan dropout dengan tingkat 0.2 untuk mengurangi overfitting. Selanjutnya pada blok kedua, model memiliki dua lapisan konvolusi dengan 64 filter dan ukuran kernel 3x3, masing-masing diikuti oleh lapisan batch normalization untuk menstabilkan aktivasi. Setelah itu, lapisan max pooling dengan ukuran pool 2x2 kembali diterapkan, diikuti oleh lapisan dropout dengan tingkat 0.3 untuk mengurangi overfitting lebih lanjut. Blok ketiga mencakup dua lapisan konvolusi dengan 128 filter dan ukuran kernel 3x3, menggunakan fungsi aktivasi ReLU dan diikuti oleh lapisan batch normalization. Setelah dua lapisan konvolusi ini, digunakan lapisan max pooling dengan ukuran pool 2x2 untuk lebih lanjut mengurangi dimensi spasial, diikuti oleh lapisan dropout dengan tingkat 0.2.

Setelah proses konvolusi dan pooling selesai, model menggunakan lapisan Global Average Pooling (GAP) untuk meratakan output menjadi vektor satu dimensi dengan ukuran 128. Proses ini membantu menangkap informasi penting dari seluruh fitur yang diekstraksi dan mengurangi dimensi spasial secara signifikan. Kemudian, terdapat lapisan dense dengan 256 unit yang menggunakan fungsi aktivasi ReLU dan regularisasi L2 untuk memperkuat representasi fitur, diikuti oleh lapisan dropout dengan tingkat 0.5. Akhirnya, model ini memiliki lapisan output dengan 1 unit yang menggunakan fungsi aktivasi sigmoid untuk melakukan klasifikasi biner, yaitu malignant dan benign.

Selanjutnya rancangan arsitektur model kedua yaitu hanya max pooling saja tanpa kombinasi dapat dilihat pada gambar berikut.

Layer (type)	Output Shape	Param #
conv2d_24 (Conv2D)	(None, 256, 256, 32)	896
batch_normalization_24 (BatchNormalization)	(None, 256, 256, 32)	128
conv2d_25 (Conv2D)	(None, 256, 256, 32)	9,248
batch_normalization_25 (BatchNormalization)	(None, 256, 256, 32)	128
max_pooling2d_12 (MaxPooling2D)	(None, 128, 128, 32)	0
dropout_16 (Dropout)	(None, 128, 128, 32)	0
conv2d_26 (Conv2D)	(None, 128, 128, 64)	18,496
batch_normalization_26 (BatchNormalization)	(None, 128, 128, 64)	256
conv2d_27 (Conv2D)	(None, 128, 128, 64)	36,928
batch_normalization_27 (BatchNormalization)	(None, 128, 128, 64)	256
max_pooling2d_13 (MaxPooling2D)	(None, 64, 64, 64)	0
dropout_17 (Dropout)	(None, 64, 64, 64)	0
conv2d_28 (Conv2D)	(None, 64, 64, 128)	73,856
batch_normalization_28 (BatchNormalization)	(None, 64, 64, 128)	512
conv2d_29 (Conv2D)	(None, 64, 64, 128)	147,584
batch_normalization_29 (BatchNormalization)	(None, 64, 64, 128)	512
max_pooling2d_14 (MaxPooling2D)	(None, 32, 32, 128)	0
dropout_18 (Dropout)	(None, 32, 32, 128)	0
flatten_4 (Flatten)	(None, 131072)	0
dense_8 (Dense)	(None, 256)	33,554,688
dropout_19 (Dropout)	(None, 256)	0
dense_9 (Dense)	(None, 1)	257

Gambar 4.7 Model Summary Max Pooling Only

Pada gambar tersebut model summary max pooling saja tanpa kombinasi dimulai dengan lapisan konvolusi (Conv2D) pertama yang memiliki 32 filter dengan ukuran kernel 3x3, menggunakan fungsi aktivasi ReLU. Lapisan ini diikuti oleh batch normalization untuk menstabilkan distribusi aktivasi selama pelatihan. Lapisan konvolusi kedua juga memiliki 32 filter dan ukuran kernel 3x3, diikuti oleh batch normalization. Setelah itu, lapisan max pooling (MaxPooling2D) dengan ukuran pool 2x2 diterapkan untuk mengurangi dimensi spasial dari fitur yang diekstraksi. Kemudian pada blok kedua, model menggunakan lapisan konvolusi dengan 64 filter dan ukuran kernel 3x3, diikuti oleh batch normalization untuk menjaga stabilitas

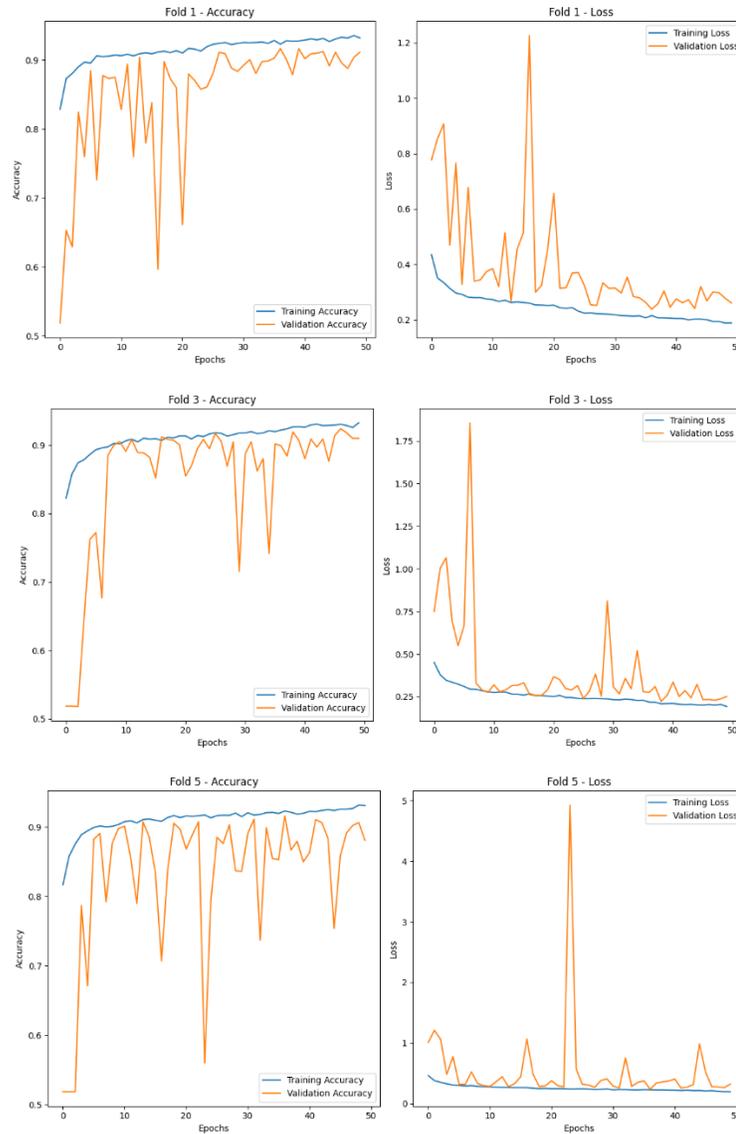
pelatihan. Lapisan konvolusi berikutnya juga memiliki 64 filter dengan ukuran kernel 3x3, diikuti oleh batch normalization. Setelah itu, lapisan max pooling dengan ukuran pool 2x2 kembali diterapkan. Blok ketiga mencakup lapisan konvolusi dengan 128 filter dan ukuran kernel 3x3, dilengkapi dengan batch normalization. Lapisan konvolusi berikutnya juga memiliki 128 filter dengan ukuran kernel 3x3, diikuti oleh batch normalization. Setelah itu, lapisan max pooling dengan ukuran pool 2x2 diterapkan untuk mengurangi dimensi spasial lebih lanjut.

Setelah proses konvolusi dan pooling selesai, model meratakan (flatten) output menjadi vektor satu dimensi. Vektor ini kemudian diteruskan ke lapisan dense dengan 256 unit yang menggunakan fungsi aktivasi ReLU dan regularisasi L2 untuk memperkuat representasi fitur. Selanjutnya, diterapkan lapisan dropout dengan tingkat 0.5 untuk mengurangi overfitting. Akhirnya, model ini memiliki lapisan output yang menggunakan fungsi aktivasi sigmoid untuk klasifikasi biner, yaitu malignant dan benign.

#### **4.3.2 Training and Testing**

Pada proses training dan testing yang dilakukan pada kedua model CNN yang diusulkan sebelumnya menggunakan k-fold cross validation. Teknik ini melibatkan pembagian data menjadi lima subset atau fold dimana  $k = 5$ . Pada setiap iterasi, empat subset digunakan sebagai data pelatihan, dan sisanya sebagai data validasi, jadi dataset tersebut dibagi secara seimbang antara kategori benign dan malignant pada tiap iterasi fold untuk menjaga proporsi yang sama seperti data asli.

Selama proses pelatihan, setiap fold dilatih selama 50 epoch. Data pelatihan diambil dari citra benign dan malignant yang telah di preprocessing. Berikut hasil kinerja model kombinasi max pooling dan global average pooling selama proses training dan testing.



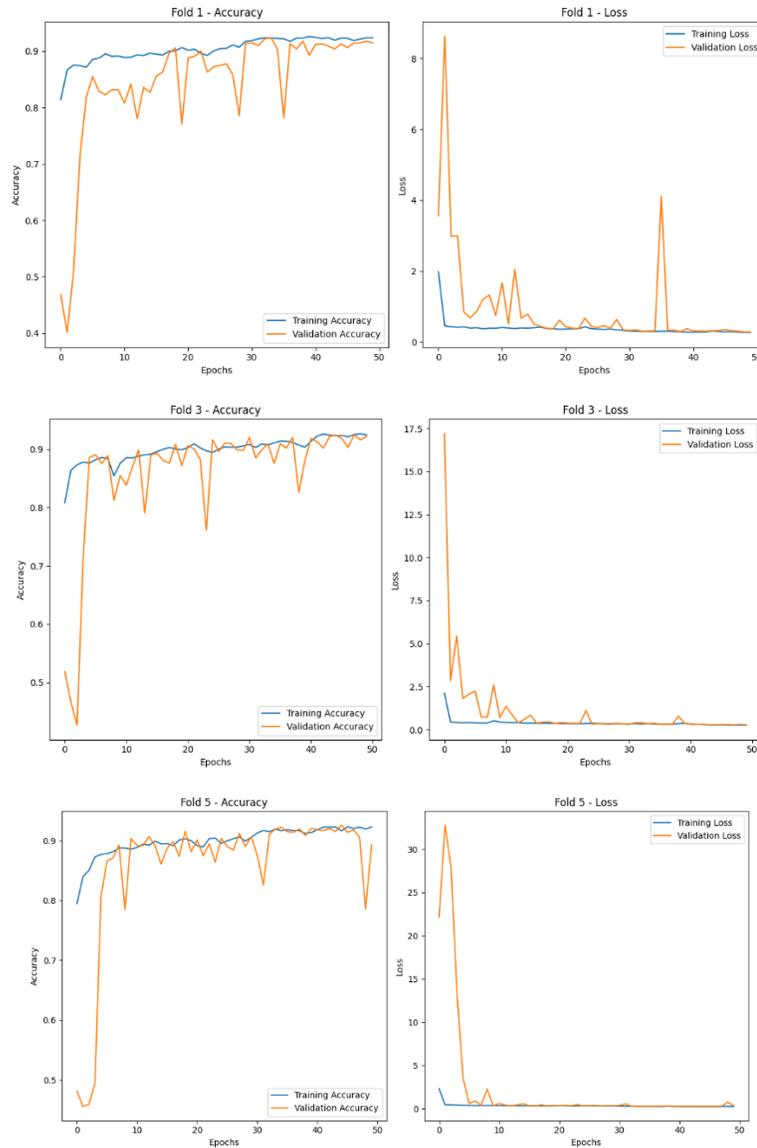
Gambar 4.8 Grafik Akurasi dan *Loss Model* Kombinasi *Fold 1, 3, dan 5*

Berdasarkan grafik akurasi dan loss untuk tiap-tiap fold pada model kombinasi max pooling dan global average pooling dapat dilihat pada fold1, model menunjukkan peningkatan akurasi pelatihan yang konsisten sekitar 0.8 hingga 0.95. akurasi validasi terlihat fluktuasi pada awal pelatihan namun stabil di sekitar 0.85 pada akhir pelatihan. Untuk loss menunjukkan penurunan yang konsisten dari 0.8 hingga 0.2, sementara loss validasi fluktuatif namun stabil di sekitar 0.4 pada akhir pelatihan. Kemudian pada gambar grafik fold3, model menunjukkan peningkatan akurasi pelatihan dari sekitar

0.85 sampai 0.9. akurasi validasi dimulai dari 0.6 dan berakhir stabil di sekitar 0.85. untuk loss pada pelatihan menurun dari 0.5 hingga 0.25, tetapi loss validasi fluktuatif dengan beberapa yang tinggi dan turun pada grafik, kemudian stabil di sekitar 0.5 pada akhir pelatihan. Dan pada gambar terakhir fold5, akurasi pelatihan meningkat dari 0.75 hingga lebih dari 0.9. akurasi validasi terlihat fluktuatif, dimulai dari 0.5 dan mencapai puncak diatas 0.9, yang menunjukkan ketidakstabilan yang signifikan. Loss pelatihan menunjukkan penurunan yang stabil, sementara loss validasi terdapat lonjakan yang signifikan.

Hal tersebut menunjukkan untuk kinerja pelatihan pada model kemampuannya baik dalam menangkap pola dari data pelatihan dengan akurasi pelatihan yang meningkat dan loss pelatihan yang menurun konsisten di semua fold. Kemudian kinerja pada akurasi validasi umumnya stabil pada akhir pelatihan walaupun ada fluktuasi di awal. Tetapi, fluktuasi yang signifikan pada loss validasi menunjukkan adanya potensi masalah overfitting atau ketidakstabilan saat pelatihan [45].

Selanjutnya untuk hasil kinerja selama proses training dan testing pada model max pooling saja tanpa kombinasi dapat dilihat pada gambar grafik akurasi dan loss berikut.



Gambar 4.9 Grafik Akurasi dan *Loss Model Max Pooling Only* Fold 1, 3, dan 5

Berdasarkan grafik akurasi dan loss model max pooling saja tersebut pada fold1 menunjukkan peningkatan akurasi pelatihan yang konsisten dari sekitar 0.7 hingga mencapai sekitar 0.9 seiring bertambahnya jumlah epoch. Akurasi validasi juga mengikuti peningkatan yang sama, walaupun mengalami fluktuasi yang signifikan pada awalnya. Pada akhir pelatihan untuk akurasi validasi stabil sekitar 0.9. grafik loss pelatihan stabil dari sekitar 2 hingga mendekati 0, dan loss validasi juga kurang lebih

sama. Pada fold3 terdapat peningkatan akurasi pelatihan sekitar 0.7 hingga mencapai 0.9. lalu akurasi validasi juga serupa, walaupun mengalami fluktuasi yang signifikan pada awalnya tetapi diakhir validasi stabil di sekitar 0.9. grafik loss pelatihan menunjukkan penurunan yang stabil dari sekitar 2 hingga mendekati 0, sementara loss validasi terlihat fluktuasi yang signifikan, namun akhirnya stabil. Selanjutnya untuk fold5, peningkatan akurasi pelatihan yang konsisten dari 0.7 hingga lebih dari 0.9, sementara untuk akurasi validasi mengalami fluktuasi yang signifikan, namun akhirnya stabil di sekitar 0.9. grafik loss pelatihan menunjukkan penurunan yang stabil dari sekitar 2 hingga mendekati 0, sementara loss validasi terlihat fluktuasi yang signifikan.

Berdasarkan dari hasil grafik proses kinerja training dan testing kedua model tersebut. Perbandingan antara model kombinasi max pooling dan global average pooling dengan model max pooling saja tanpa kombinasi yaitu, model kombinasi terdapat fluktuasi loss validasi diawal dan pertengahan setelah itu lebih cenderung stabil akhir pelatihan. Sementara model yang hanya max pooling saja menunjukkan stabilisasi yang cepat setelah fluktuasi awal yang tinggi, dengan akurasi validasi yang akhirnya mendekati akurasi pelatihan serta loss yang tetap rendah setelah beberapa epoch awal. Hal tersebut cukup bagus, karena setelah awal yang fluktuatif, model dengan cepat stabil dan mempertahankan performa yang baik. Sejalan dengan pengamatan pada penelitian yang dilakukan Prechelt [46], kurva error validasi seringkali memiliki beberapa titik minimum sebelum mencapai hasil terbaik. Model kombinasi mengalami lebih banyak fluktuasi sebelum akhirnya stabil, sementara model max pooling saja lebih cepat stabil setelah awal pelatihan yang kurang baik.

Secara keseluruhan, model kombinasi memiliki pola belajar yang lebih bervariasi tetapi tetap stabil di akhir pelatihan, sedangkan model max pooling saja menunjukkan proses stabilisasi yang lebih cepat setelah fluktuasi awal yang tinggi. Hal ini menunjukkan bahwa model max pooling tidak mengalami overfitting parah, tetapi justru mencapai keseimbangan antara *training* dan generalisasi. Ini dikarenakan penggunaan teknik regulasi lain seperti Dropout atau Batch Normalization, kemudian data pelatihan yang cukup beragam dan besar sehingga overfitting dapat dikendalikan oleh max pooling [47], [48], [49].

### 4.3.3 Evaluation

Tahapan evaluasi pada penelitian ini dilakukan dengan menghitung tiap-tiap fold dengan menggunakan metrik akurasi, presisi, recall, dan F1-Score. Nilai metrik pada tiap fold kemudian dirata-ratakan untuk memberikan gambaran hasil keseluruhan. Akurasi mengukur seberapa baik model dalam membuat prediksi yang benar, presisi mengukur proporsi prediksi positif yang benar-benar positif, recall mengukur proporsi kasus positif yang berhasil dideteksi oleh model, dan F1-Score adalah rata-rata harmonis dari presisi dan recall. Berikut tabel perbandingan rata-rata matrik kombinasi max pooling dan global average pooling dengan max pooling saja.

Tabel 4.4 *Comparison of Average Result Model Evaluation*

No	Model CNN	Average Result			
		Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
1	Max Pooling + GAP	91.68%	96.35%	85.99%	90.87%
2	Max Pooling	91.79%	93.60%	89.04%	91.26%

Berdasarkan dari hasil evaluasi tersebut, dapat diamati bahwa model max pooling tanpa kombinasi memiliki nilai rata-rata akurasi yang sedikit lebih tinggi dibandingkan model kombinasi max pooling dan global average pooling (GAP). Namun, perbedaannya sangatlah kecil yaitu sekitar 0.11% (91.79% vs 91.68%). Fenomena ini dapat dijelaskan melalui karakteristik teknik pooling yang berbeda dalam arsitektur CNN.

Max Pooling cenderung mempertahankan fitur yang paling dominan dengan mengambil nilai maksimum pada tiap *feature map* [16], sehingga memungkinkan model untuk mempertahankan detail penting pada area yang signifikan dalam citra kulit. Hal tersebut berkontribusi pada kemampuan model max pooling untuk mencapai recall yang lebih tinggi sebesar 89.04% dibandingkan 85.99% pada model kombinasi, yang menunjukkan bahwa model ini lebih baik dalam mendeteksi kasus kanker kulit positif.

Tetapi, model kombinasi max pooling dan GAP menunjukkan keunggulan signifikan pada nilai presisi (96.35% vs 93.60%). R. L. Kumar et al. (2021) menjelaskan bahwa GAP membantu mengurangi overfitting dengan memberikan regularisasi implisit dalam model, yang mengarah pada pengurangan prediksi positif palsu [17]. GAP mengambil rata-rata dari seluruh *feature map*, sehingga memberikan perspektif global yang lebih baik tentang keberadaan fitur di seluruh citra. Pendekatan ini sangat bermanfaat dalam konteks deteksi kanker kulit dimana presisi tinggi diperlukan untuk menghindari diagnosis yang tidak perlu (prediksi positif palsu).

Selain itu, nilai F1-Score pada model kombinasi max pooling dan GAP adalah 90.87% sedikit lebih rendah dari 91.26% pada model max pooling saja. F1-Score merupakan rata-rata harmonis dari presisi dan recall, sehingga menggambarkan keseimbangan antara kemampuan model dalam mendeteksi dan mengurangi prediksi positif palsu. Penggunaan teknik pooling yang berbeda menciptakan representasi fitur yang berbeda, di mana Max Pooling lebih fokus pada fitur lokal yang menonjol, sementara kombinasi dengan GAP memberikan keseimbangan antara fitur lokal dan global.

Selain itu, *confusion matrix* digunakan untuk setiap fold untuk memvisualisasikan jumlah prediksi benar dan salah untuk masing-masing kategori benign dan malignant. *Confusion matrix* membantu dalam memahami dimana kesalahan prediksi terjadi dan dalam mengevaluasi keseimbangan kinerja model antara dua kategori. Berikut tabel 4.5 dan tabel 4.6 *confusion matrix* tiap fold pada model kombinasi dan tidak.

Tabel 4.5 *Confusion Matrix Model Max Pooling and Global Average Pooling*

Fold	True Positive (TP)	True Negative (TN)	False Positive (FP)	False Negative (FN)
1	927	1016	84	94
2	1073	852	27	169
3	1070	890	30	131
4	1075	874	25	147
5	1075	880	25	141

Tabel 4.6 *Confusion Matrix Model Only Max Pooling*

Fold	True Positive (TP)	True Negative (TN)	False Positive (FP)	False Negative (FN)
1	1034	906	66	115
2	1047	917	53	104
3	1053	909	47	112
4	1027	895	73	126
5	1028	919	72	102

Berdasarkan dari perbandingan tabel hasil dari *confusion matrix* pada kedua model, terlihat perbedaan pola yang signifikan dalam kinerja deteksi kanker kulit. Pada model kombinasi (Max Pooling + GAP), nilai *true positive* TP menunjukkan jumlah kasus kanker kulit ganas (malignant) yang berhasil terdeteksi dengan benar oleh model. Misalnya, pada fold 1, angka 927 mengindikasikan bahwa sebanyak 927 sampel kanker ganas berhasil diidentifikasi dengan tepat. Nilai TP ini cenderung bervariasi di seluruh fold, dengan fold 1 memiliki nilai TP yang lebih rendah dibandingkan fold lainnya, menunjukkan konsistensi deteksi yang berbeda tergantung pada distribusi data dalam fold tersebut.

Model kombinasi menunjukkan keunggulan dalam hal True Negative (TN), seperti pada fold 1 dengan nilai 1016, yang berarti 1016 kasus kulit jinak berhasil dikategorikan dengan benar sebagai jinak. Keunggulan signifikan model kombinasi terlihat pada nilai False Positive (FP) yang secara konsisten lebih rendah di seluruh fold. Contohnya, pada fold 2, model kombinasi hanya menghasilkan 27 kasus FP (kasus jinak yang salah diklasifikasikan sebagai ganas), sementara model Max Pooling saja menghasilkan 53 kasus FP. Rendahnya nilai FP ini menunjukkan presisi yang lebih baik, sejalan dengan nilai presisi yang lebih tinggi pada tabel evaluasi (96,35% vs 93,60%).

Namun, trade-off yang terjadi terlihat jelas pada nilai False Negative (FN), di mana model kombinasi memiliki angka yang lebih tinggi di sebagian besar fold. Pada fold 2, terdapat 169 kasus FN (kasus ganas yang tidak terdeteksi) dibandingkan dengan hanya 104 kasus pada model Max Pooling saja. Nilai FN yang lebih tinggi ini

mengindikasikan bahwa model kombinasi cenderung lebih konservatif dalam mengklasifikasikan sampel sebagai kanker ganas, yang berkorelasi dengan nilai recall yang lebih rendah (85,99% vs 89,04%).

Pada model Max Pooling saja, terlihat pola yang berbeda dengan nilai TP yang lebih konsisten dan umumnya lebih tinggi di semua fold. Misalnya, pada fold 1, model ini berhasil mengidentifikasi 1034 kasus ganas dengan benar, dibandingkan 927 pada model kombinasi. Nilai TN juga cukup stabil, seperti pada fold 3 dengan 909 kasus jinak yang teridentifikasi dengan benar. Namun, kelemahan utama model ini adalah nilai FP yang lebih tinggi, dengan contoh pada fold 4 sebanyak 73 kasus jinak salah dikategorikan sebagai ganas, yang bisa berimplikasi pada tingkat intervensi medis yang tidak perlu dalam konteks klinis.

Perbedaan-perbedaan ini menjelaskan mengapa model Max Pooling memiliki recall yang lebih baik karena kemampuannya mendeteksi lebih banyak kasus positif, tetapi dengan konsekuensi presisi yang lebih rendah akibat FP yang lebih tinggi. Nilai F1-Score yang sedikit lebih tinggi untuk model Max Pooling saja (91,26% vs 90,87%) mengindikasikan bahwa model ini mencapai keseimbangan yang lebih baik antara presisi dan recall, meskipun perbedaannya tidak signifikan.

Variasi kinerja antar fold juga menunjukkan sensitivitas model terhadap distribusi data dan memberikan gambaran komprehensif tentang stabilitas model dalam berbagai subset data, yang merupakan keuntungan dari penggunaan metode k-fold cross-validation dalam evaluasi penelitian ini.