

Perbandingan Kinerja Model *Transfer Learning* *VGG16* dan *ResNet50* Dalam Klasifikasi Tumor Otak Pada Citra *MRI*

Baltasar Yosef Mnaku Gawen
Fakultas Teknik Elektro
Purwokerto, Indonesia
bayugawen@student.telkomuniversity.
ac.id

Sevia Indah Purnama
Fakultas Teknik Elektro
Purwokerto, Indonesia
seviaindah@telkomuniversity.ac.id

Irmayatul Hikmah
Fakultas Teknik Elektro
Purwokerto, Indonesia
Irmayatulh@telkomuniversity.ac.id

Abstrak--Tingginya prevalensi tumor otak dan pentingnya deteksi dini mendorong pengembangan sistem diagnosis otomatis berbasis kecerdasan buatan. Penelitian ini bertujuan membandingkan kinerja dua arsitektur *deep learning*, yaitu *VGG16* dan *ResNet50*, dalam mengklasifikasi empat jenis citra *MRI* tumor otak yakni *glioma*, *meningioma*, *pituitary*, dan *no-tumor*. Pendekatan yang digunakan adalah *transfer learning* dengan dua skenario perbandingan yakni sebelum dan sesudah *fine-tuning*. *Transfer learning* memungkinkan pemanfaatan model yang telah dilatih sebelumnya, sedangkan *fine-tuning* dilakukan untuk menyesuaikan parameter model terhadap karakteristik data baru. Seluruh proses penelitian mengikuti tahapan *Knowledge Discovery in Database (KDD)*, yang meliputi pengumpulan data, *preprocessing*, pelatihan model, dan evaluasi. Hasil evaluasi menunjukkan bahwa *ResNet50* menghasilkan akurasi sebesar 94,80% sebelum *fine-tuning* dan meningkat menjadi 97,54% setelah *fine-tuning*. Sementara itu, *VGG16* memperoleh akurasi sebesar 91,39% sebelum *fine-tuning* dan meningkat signifikan menjadi 98,68% setelah *fine-tuning*. Berdasarkan hasil tersebut, dapat disimpulkan bahwa model *VGG16* dengan *fine-tuning* menunjukkan kinerja terbaik dalam klasifikasi tumor otak pada citra *MRI*. Temuan ini mengindikasikan bahwa pendekatan *deep learning* berbasis *transfer learning* berpotensi meningkatkan efektivitas dan efisiensi diagnosis tumor otak secara otomatis.

Kata Kunci: *fine-tuning*, *transfer learning*, *VGG16*, *ResNet50*, klasifikasi tumor, citra *MRI*

I. PENDAHULUAN

Otak manusia merupakan pusat kendali utama yang mengatur seluruh fungsi tubuh, memungkinkan individu untuk menyesuaikan diri dan merespon berbagai kondisi lingkungan [1]. Sebagai pusat kendali, otak juga rentan terhadap berbagai gangguan, salah satunya adalah tumor otak, yaitu pertumbuhan sel yang tidak terkendali dan menyimpang, yang dapat memengaruhi fungsi vital seperti memori, motorik, serta kontrol kognitif dan emosional [2]. Berdasarkan laporan *International Agency for Research on Cancer (IARC)* tahun 2020, terdapat 308.102 kasus tumor otak ganas dan sistem saraf pusat secara global, dengan 251.329 kasus berujung kematian. Di Indonesia, tercatat 5.738 kasus tumor otak ganas pada tahun 2022, dan 5.259 di

antaranya berakhir dengan kematian, menempatkan Indonesia pada urutan ke-15 kasus terbanyak di dunia [3].

Deteksi dini menjadi langkah krusial dalam proses diagnosis karena memungkinkan pengobatan lebih efektif bila dilakukan pada tahap awal perkembangan tumor [4]. Salah satu metode diagnosis utama adalah pencitraan *Magnetic Resonance Imaging (MRI)*, yang memungkinkan tenaga medis mengidentifikasi lokasi dan jenis tumor, serta merancang terapi yang sesuai [5]. Namun, pembacaan citra *MRI* secara manual masih menghadapi tantangan, seperti keterbatasan tenaga medis, waktu yang lama, dan potensi kesalahan manusia [6].

Untuk mengatasi hal tersebut, teknologi kecerdasan buatan seperti *deep learning* mulai diterapkan pada citra *MRI* karena kemampuannya mengenali pola kompleks dan anomali yang mungkin terlewatkan oleh manusia. Salah satu arsitektur yang banyak digunakan dalam pengolahan citra adalah *Convolutional Neural Network (CNN)* [7]. Dalam konteks keterbatasan data dan sumber daya komputasi, *transfer learning* menjadi solusi yang efektif karena memungkinkan pemanfaatan model pra-latih seperti *VGG16* dan *ResNet50* untuk tugas klasifikasi spesifik [8]. Teknik *fine-tuning* diterapkan untuk menyesuaikan parameter model terhadap karakteristik dataset baru guna meningkatkan performa [9].

Penelitian ini bertujuan untuk membandingkan kinerja model *VGG16* dan *ResNet50* dalam mengklasifikasikan empat jenis citra *MRI* tumor otak yakni *glioma*, *meningioma*, *pituitary*, dan *no-tumor*, baik sebelum maupun sesudah penerapan teknik *fine-tuning*. Diharapkan hasil penelitian ini dapat memberikan kontribusi terhadap pengembangan sistem diagnosis otomatis yang lebih akurat dan efisien dalam mendeteksi tumor otak.

II. KAJIAN TEORI

A. Tumor otak

Tumor otak merupakan pertumbuhan sel abnormal pada jaringan otak atau sekitarnya, terbagi menjadi tumor primer dan sekunder (*metastasis*). Berdasarkan asal jaringannya, jenis tumor umum meliputi *glioma*, *meningioma*, *pituitary*,

dan schwannoma. Glioma berasal dari sel glial, meningioma dari meningen, pituitary dari kelenjar hipofisis, dan schwannoma dari sel pelindung saraf [10]

a) MRI

MRI adalah teknik pencitraan *non-invasif* yang menghasilkan gambar anatomi tiga dimensi dengan resolusi tinggi. Prinsip kerja MRI didasarkan pada resonansi magnetik nuklir yang memanfaatkan medan magnet untuk menangkap sinyal dari proton dalam tubuh, sehingga mampu membedakan jenis jaringan dengan akurat [11].

b) Klasifikasi

Klasifikasi merupakan proses untuk mengidentifikasi dan memetakan data ke dalam kelas tertentu berdasarkan pola yang dipelajari. Tahap klasifikasi umumnya mencakup pelatihan (*training*) dan pengujian (*testing*) menggunakan data berlabel untuk membentuk dan menguji model klasifikasi [12].

c) Deep learning

Deep learning merupakan subbidang *machine learning* yang memanfaatkan jaringan saraf dalam berlapis (*deep neural network*) untuk mengenali pola kompleks dalam data besar. Keunggulannya adalah kemampuannya memproses data tidak terstruktur secara otomatis tanpa fitur buatan manusia [13][14].

d) Transfer learning

Transfer learning adalah metode yang memanfaatkan model pra-latih untuk menyelesaikan tugas baru pada domain target. Hal ini efektif ketika jumlah data terbatas atau daya komputasi rendah. Transfer learning terdiri dari tiga jenis utama yakni *inductive*, *transductive*, dan *unsupervised* [15].

e) Fine-tuning

Fine-tuning merupakan proses pelatihan ulang sebagian atau seluruh parameter model pra-latih agar menyesuaikan dengan dataset baru. Proses ini dilakukan dengan *learning rate* rendah untuk menjaga stabilitas parameter awal dan meningkatkan akurasi pada tugas klasifikasi spesifik [16].

f) ResNet50 & VGG16

ResNet50 adalah arsitektur CNN dengan 50 lapisan dan *shortcut connection*, yang mencegah *vanishing gradient* pada jaringan dalam. Sedangkan VGG16 adalah model CNN dengan 16 lapisan yang menggunakan filter kecil (3×3) dan konfigurasi sederhana, efektif untuk deteksi fitur visual dengan kontras tinggi [17][18].

g) Confusion matrix

Confusion matrix digunakan untuk menilai performa model klasifikasi melalui metrik seperti *accuracy*, *precision*, *recall*, dan *F1-score*. Evaluasi ini membantu memahami kekuatan dan kelemahan model dalam mengklasifikasikan data positif maupun negatif secara tepat. Confusion matrix memiliki elemen-elemen, diantaranya adalah True Positive (TP), False Positive (FP), False Negative (FN), dan True Negative (TN). Berikut ini persamaan untuk menghitung metrik confusion matrix [19].

Accuracy:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (20)$$

Precision:

$$Precision = \frac{TP}{TP + FP} \quad (20)$$

$$Precision = \frac{TP}{TP + FP}$$

Recall:

$$Recall = \frac{TP}{TP + FN} \quad (20)$$

F1-score:

$$F1 - Score = \frac{Precision \times Recall}{Precision + Recall} \quad (20)$$

Keterangan:

TP: True Positive

TN: True Negative

FP: False Positive

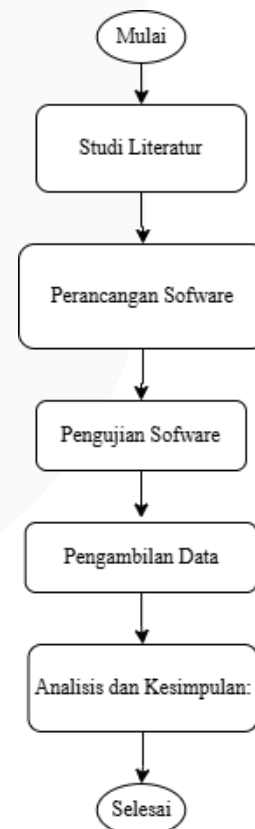
FN: False Negative

h) Knowledge Discovery in Database

adalah proses ekstraksi pengetahuan dari data dalam lima tahap utama yakni pengumpulan data, *preprocessing*, transformasi, data mining, evaluasi, dan interpretasi hasil. Metode ini mendasari struktur kerja penelitian klasifikasi berbasis data citra medis [21].

III. METODE

Penelitian ini menggunakan pendekatan *transfer learning* dalam proses klasifikasi citra MRI tumor otak dengan dua arsitektur CNN, yaitu VGG16 dan ResNet50, dan dua skenario perbandingan yakni sebelum dan sesudah *fine-tuning*. Proses penelitian mengacu pada tahapan *Knowledge Discovery in Database (KDD)*. Seluruh alur tahapan penelitian dijelaskan secara visual dalam gambar 3.1.



GAMBAR 1
(DIAGRAM ALIR PENELITIAN)

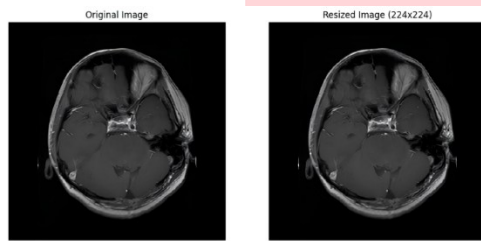
Data citra MRI tumor otak diperoleh dari *Mendeley Data* dan *Kaggle*, yang terdiri atas empat kelas yakni *glioma*, *meningioma*, *pituitary*, dan *no-tumor*. Dataset tersebut

dikumpulkan menjadi dua folder utama yakni *data training* (70%) dan *data testing* (30%) yang kemudian akan dibagi oleh model menjadi 3 folder yakni *data training* (70%), *data validation* (15%) dan *data testing* (30%) sebagaimana ditampilkan dalam tabel 3.1.

TABEL 1
(JUMLAH DATASET MASING-MASING KELAS SETELAH PROSES DATA SPLIT)

No	Kelas	Training	Testing	validation
1	<i>Meningioma</i>	1151	247	247
2	<i>Glioma</i>	1135	243	243
3	<i>Pituitary</i>	1230	264	263
4	<i>No-Tumor</i>	1414	303	303

Selanjutnya, *preprocessing* dilakukan melalui penyesuaian ukuran gambar (*resize 224×224 piksel*), *padding*, serta konversi *grayscale* ke format *ImageNet* seperti pada gambar 3.2.



GAMBAR 2
(RESIZED CITRA MRI)

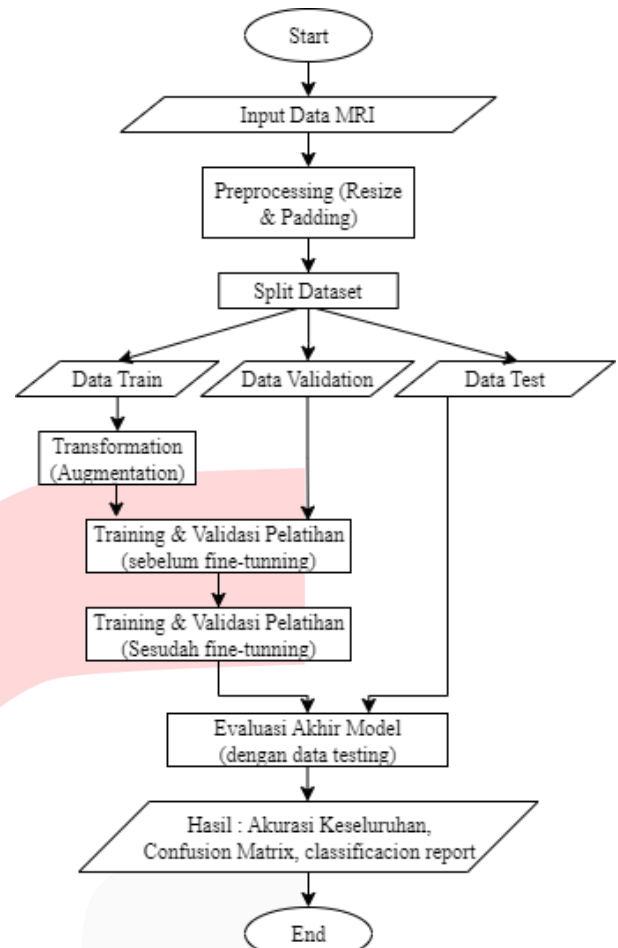
Data kemudian ditransformasikan dengan teknik *augmentasi*, meliputi *rotation 10°*, *zooming*, dan *flipping* untuk memperbesar keragaman data dan mencegah *overfitting*. Model *VGG16* dan *ResNet50* diinisialisasi dengan layer pra-latih dari *ImageNet*. Pada tahap pertama, model dilatih (*training*) hanya dengan *costum classifier* yang disesuaikan seperti pada tabel 3.2.

TABEL 2
(LAPISAN COSTUM CLASSIFIER VGG16 dan ResNet50)

No	Lapisan
1	<i>GlobalAveragePooling2D ()</i>
2	<i>Dropout (0.5)</i>
3	<i>BatchNormalization ()</i>
4	<i>Dropout (0.3)</i> <i>Dense(len(train_ds.class_names),</i> <i>activation='softmax')</i>

Selanjutnya pada tahap kedua, dilakukan *fine-tuning* dengan membuka seluruh layer model dari *base model* dan disatukan dengan *costum classifier* untuk pelatihan ulang, agar model lebih adaptif terhadap dataset *MRI*.

Evaluasi kinerja model dilakukan menggunakan *data testing* dengan metrik *accuracy*, *precision*, *recall*, dan *F1-score* yang dihitung dari *confusion matrix*. Seluruh proses sistem klasifikasi yang dibangun dijelaskan melalui Gambar 3.3.



GAMBAR 3
(DIAGRAM ALIR SISTEM)

Perangkat keras yang digunakan terdiri dari laptop dengan spesifikasi *Intel Core i3*, *RAM 8 GB*, dan sistem operasi *Windows 11*. Perangkat lunak yang digunakan meliputi *Google Colaboratory*, bahasa pemrograman *Python*, dan library *TensorFlow*, *Keras*, *OpenCV*, *NumPy*, *Pandas*, dan *Matplotlib*.

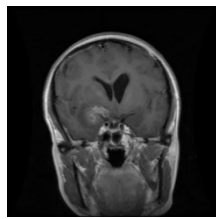
IV. HASIL DAN PEMBAHASAN

Proses klasifikasi citra *MRI* tumor otak dilakukan melalui serangkaian tahapan yang disusun secara sistematis mulai dari akuisisi data, *preprocessing*, pelatihan model, hingga evaluasi hasil klasifikasi. Pada proses akuisisi data, dataset dari *Mendeley Data*, dan *Kaggle* dikumpulkan. Dataset yang diperoleh mencakup empat kelas yaitu *glioma*, *meningioma*, *pituitary*, dan *non-tumor*. Masing-masing kelas data selanjutnya dikelompokkan ke dalam dua bagian utama, yaitu data pelatihan (*training*) dan data pengujian (*testing*) seperti yang tertera pada tabel 3.3.

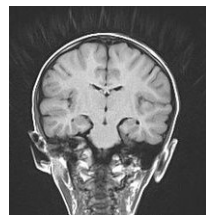
TABEL 3
(JUMLAH DATASET PERSETIAP KELAS)

No	Kelas	Jumah data set
1	<i>Meningioma Images</i>	1645
2	<i>Pituitary Images</i>	1757
3	<i>Glioma Images</i>	1621
4	<i>No-tumor Images</i>	2020

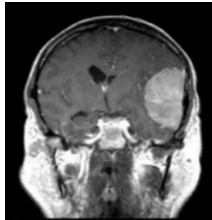
Secara keseluruhan, total dataset yang digunakan dalam penelitian ini berjumlah 7.047 citra MRI. Penyajian visual berupa contoh citra dari masing-masing kelas ditampilkan pada gambar 3.4 hingga gambar 3.7



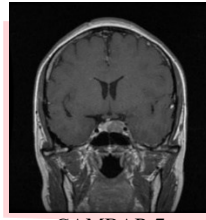
GAMBAR 4
(GLIOMA)



GAMBAR 5
(NO-TUMOR)



GAMBAR 6
(MENINGIOMA)



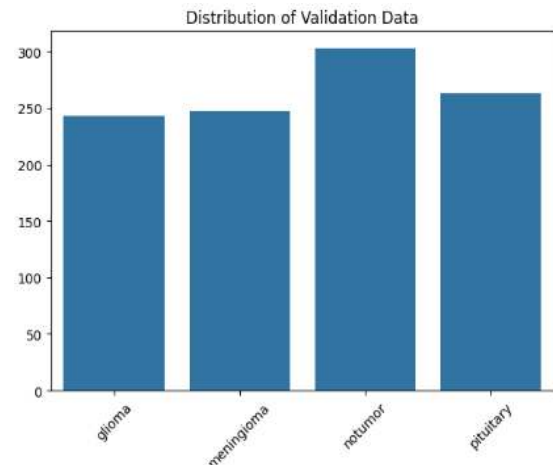
GAMBAR 7
(PITUITARY)

Selanjutnya dilakukan proses *preprocessing* tambahan berupa *resize* dan *padding* terhadap seluruh citra MRI. Pertama, dilakukan *resize* ke dimensi standar 224×224 piksel. *Resize* dilakukan secara otomatis pada saat proses *batch data* menggunakan *ImageDataGenerator*, dengan parameter *target_size*=(224, 224). Setelah itu, *padding* juga ditambahkan dengan *tf.image.resize_with_crop_or_pad* dan *tf.pad* dari *TensorFlow*, yang secara otomatis menambahkan batas hitam (nilai piksel 0) di sisi vertikal atau horizontal hingga citra mencapai ukuran target tanpa mendistorsi bentuk objek.

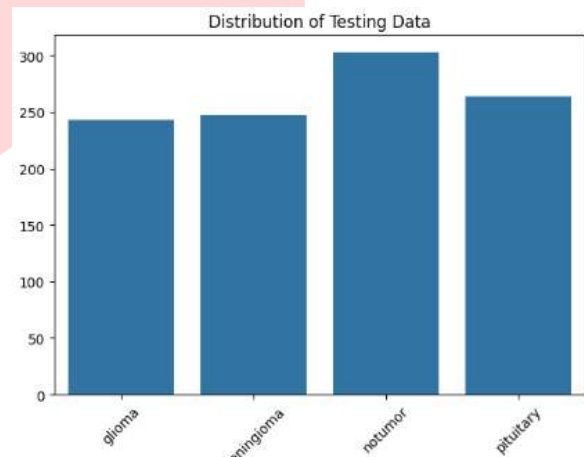
Dataset yang telah dikumpulkan kemudian diproses dengan membaginya menjadi tiga subset, yaitu *training*, *validation*, dan *testing*. Proses pembagian dilakukan secara bertahap menggunakan fungsi *train_test_split* dari pustaka *sklearn*. Sebanyak 70% data dialokasikan untuk pelatihan, sementara 30% sisanya dibagi rata menjadi data validasi dan pengujian, masing-masing 15%. Proses ini mempertahankan distribusi kelas secara proporsional dengan parameter *stratify* sebagaimana ditampilkan pada gambar diagram 3.8 sampai 3.10.



GAMBAR 8
(DISTRIBUSI TRAINING DATA)

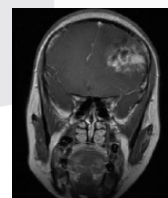


GAMBAR 9
(DISTRIBUSI VALIDATION DATA)

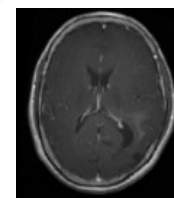


GAMBAR 10
(DISTRIBUSI TESTING DATA)

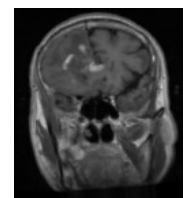
Untuk meningkatkan generalisasi model, dilakukan *augmentasi data* secara *real-time* menggunakan *ImageDataGenerator* dari *TensorFlow Keras*. *Augmentasi* diterapkan langsung saat proses pelatihan tanpa menyimpan file baru di *disk*. *Transformasi* yang digunakan meliputi rotasi, pergeseran, *zoom*, *flip horizontal*, dan penyesuaian kecerahan. Contoh hasil augmentasi ditampilkan pada Gambar 11.



Original image



Rotasi



Perubahan
Skala (Zoom)

GAMBAR 11
(CONTOH HASIL CITRA SEBELUM DAN SESUDAH AUGMENTASI DATA)

Setelah proses pembagian data dan augmentasi dilakukan, tahap selanjutnya adalah pembangunan arsitektur model. Model *VGG16* dan *ResNet50* diimpor tanpa lapisan klasifikasi atas (*include_top=False*) agar dapat disesuaikan dengan kebutuhan klasifikasi multi-kelas. Input citra

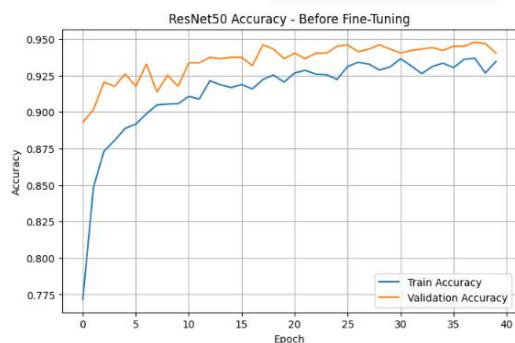
berukuran 224x224 piksel dalam format *RGB* dikirimkan ke model, dilanjutkan proses ekstraksi fitur dengan bobot *pretrained* yang dibekukan. Kemudian, ditambahkan lapisan *costum classification* yang terdiri dari *GlobalAveragePooling2D*, *Dense*, *BatchNormalization*, *Dropout*, dan *output Dense* beraktivasi *softmax*, dengan total parameter masing-masing sekitar 14 juta dan 23 juta.

Pelatihan awal dilakukan selama 40 *epoch* menggunakan *optimizer Adam* dengan *learning rate* 0.001 dan *loss function categorical_crossentropy*. Setelah evaluasi awal dengan data uji, model dilanjutkan ke tahap *fine-tuning* dengan membuka seluruh lapisan (*trainable=True*) dan menurunkan *learning rate* menjadi 0.00001. Proses ini dilengkapi dengan *EarlyStopping* dan *ReduceLROnPlateau* untuk menjaga stabilitas pelatihan dan mencegah *overfitting*. Konfigurasi hyperparameter disajikan pada tabel 4.4.

TABEL 4
(COSTUM CLASSIFIER PADA VGG16 DAN RESNET50)

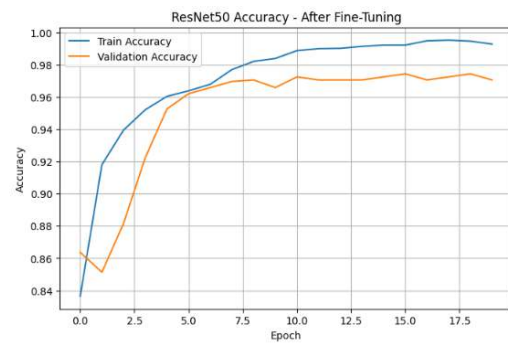
No	Parameter	Nilai
1	Input Shape	(224, 224, 3)
2	Jumlah Epoch	40 (<i>pre-training</i>), 60 (<i>fine tuning</i>)
3	Batch Size	32
4	Learning Rate	0.001 (<i>Pre-training</i>), 0.00001 (<i>fine-tuning</i>)

Setelah pembangunan arsitektur, model *ResNet50* dilatih menggunakan pendekatan *transfer learning* dengan bobot *pretrained* dari *ImageNet* dan struktur sebanyak 185 layer. Proses pelatihan ini dibantu oleh *callback* seperti *EarlyStopping* dan *ReduceLROnPlateau*.



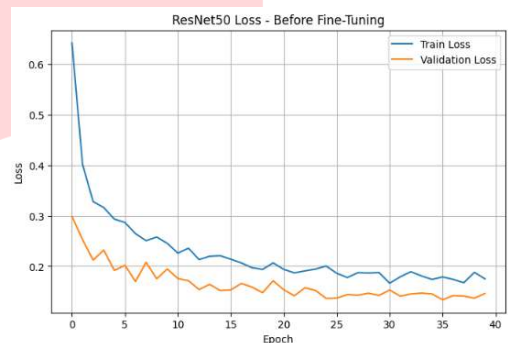
GAMBAR 12
(GRAFIK AKURASI RESNET50 SEBELUM FINE-TUNING)

Gambar 12 menunjukkan akurasi pelatihan awal sebelum *fine-tuning*. Terlihat kurva *training* meningkat dari 77% hingga 93%, sementara kurva validasi sejak awal stabil tinggi di kisaran 89–95%, menandakan model memiliki generalisasi yang baik tanpa *overfitting*.



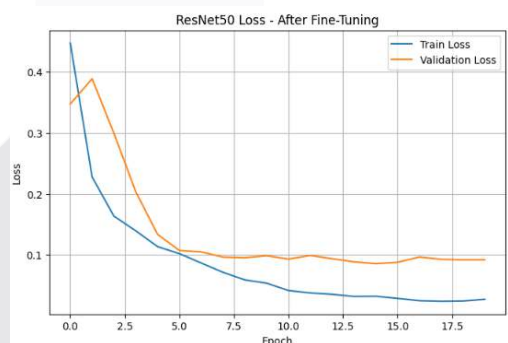
GAMBAR 13
(AKURASI RESNET50 SETELAH FINE-TUNING)

Gambar 13 menunjukkan peningkatan akurasi setelah *fine-tuning*, dengan akurasi *training* mencapai 99% dan validasi stabil di 96–97%. Selisih antar kurva kecil, menandakan kestabilan model.



GAMBAR 14
(GRAFIK LOSS RESNET50 SEBELUM FINE-TUNING)

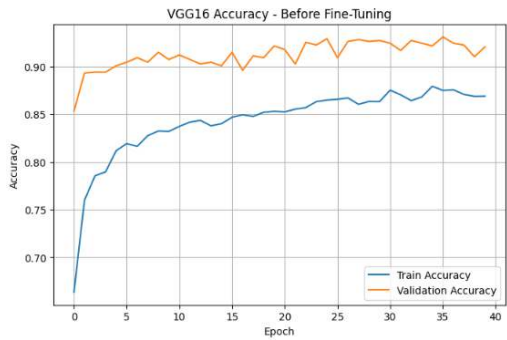
Gambar 4.15 memperlihatkan penurunan *train loss* dari 0.65 ke 0.18, sedangkan *validation loss* menurun stabil dari 0.30 ke kisaran 0.13, tanpa lonjakan signifikan.



GAMBAR 16
(GRAFIK LOSS RESNET50 SESUDAH FINE-TUNING)

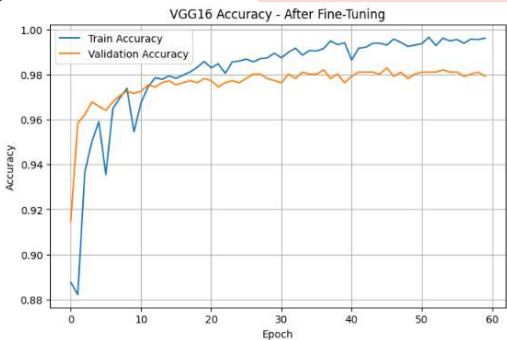
Gambar 16 menunjukkan hasil setelah *fine-tuning*, di mana *train loss* menurun hingga di bawah 0.02, dan *validation loss* turun stabil ke kisaran 0.09.

Selanjutnya, model *VGG16* dilatih menggunakan bobot *pretrained* dari *ImageNet* dan struktur sebanyak 26 layer. Proses pelatihan ini dibantu oleh *callback* seperti *EarlyStopping* dan *ReduceLROnPlateau*.



GAMBAR 17
(GRAFIK AKURASI VGG16 SEBELUM FINE-TUNING)

Gambar 17 menunjukkan akurasi sebelum *fine-tuning*, di mana akurasi *training* meningkat dari 67% ke 87%, dan validasi stabil tinggi dari awal di atas 90%, menunjukkan model tidak mengalami *overfitting* dan menghasilkan akurasi *testing* sebesar 91,39%.



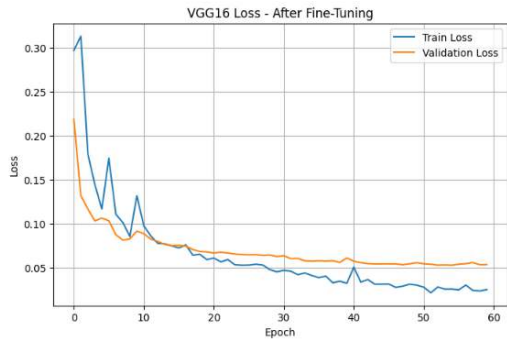
GAMBAR 18
(GRAFIK AKURASI VGG16 SETELAH FINE-TUNING)

Gambar 18 menunjukkan peningkatan lebih signifikan setelah *fine-tuning*, dengan akurasi *training* mencapai hampir 100% dan validasi stabil di kisaran 98%. Akurasi *testing* pun meningkat menjadi 98,68%, menandakan keberhasilan *fine-tuning* dalam meningkatkan performa.



GAMBAR 19
(GRAFIK LOSS VGG16 SEBELUM FINE-TUNING)

Gambar 19 memperlihatkan penurunan *train loss* dari 0.91 ke 0.30, sedangkan *validation loss* turun stabil dari 0.44 ke 0.22, menunjukkan model memiliki kestabilan yang baik sebelum *fine-tuning*.



GAMBAR 20
(GRAFIK LOSS VGG16 SESUDAH FINE-TUNING)

Gambar 20 menunjukkan *loss training* menurun dari 0.31 ke bawah 0.03, dan *validation loss* stabil rendah di kisaran 0.07 hingga akhir pelatihan, memperkuat indikasi tidak terjadinya *overfitting*.

Setelah pelatihan model selesai, tahap evaluasi dilakukan untuk mengukur kemampuan klasifikasi menggunakan data uji sebanyak 1.057 gambar. Evaluasi dilakukan melalui laporan klasifikasi dan confusion matrix. Sebelum *fine-tuning*, model *ResNet50* menunjukkan performa tinggi dengan akurasi keseluruhan sebesar 94,80%. Precision tertinggi dicapai oleh kelas *notumor* sebesar 1.00, diikuti *pituitary* 0.99, *glioma* 0.98, dan *meningioma* 0.84. Recall tertinggi juga pada kelas *notumor* dan *pituitary* (0.99 dan 0.97), sedangkan *glioma* 0.85 dan *meningioma* 0.97. F1-score tertinggi tetap pada *notumor* (0.99), diikuti *pituitary* (0.98), *glioma* (0.91), dan *meningioma* (0.90) seperti yang terlihat pada tabel 4.5.

TABEL 5
(CLASSIFICATION RESNET50 SEBELUM FINE-TUNING)

ResNet50 Classification Report (Before Fine-tuning)				
Precision		Recall	F-1 Score	Support
Glioma	0.98	0.85	0.91	243
Meningioma	0.84	0.97	0.90	247
Notumor	1.00	0.99	0.99	303
Pituitary	0.99	0.97	0.98	264
Accuracy			0.95	1057
Macro avg	0.95	0.94	0.95	1057
Weighted avg	0.95	0.95	0.95	1057
Akurasi keseluruhan sebelum <i>fine-tuning</i> : 94.80%				

Setelah *fine-tuning*, performa meningkat signifikan. Akurasi keseluruhan menjadi 97,54%. Semua metrik precision, recall, dan F1-score meningkat hampir merata pada keempat kelas, dengan nilai F1 tertinggi tetap pada *notumor* (1.00), menunjukkan model sangat andal pasca *fine-tuning* seperti yang terlihat pada tabel 4.6.

TABEL 6
(CLASSIFICATION RESNET50 SETELAH FINE-TUNING)

ResNet50 Classification Report (After Fine-tuning)				
Precision		Recall	F-1 Score	Support
Glioma	0.96	0.96	0.96	243
Meningioma	0.95	0.96	0.95	247
Notumor	1.00	0.99	1.00	303
Pituitary	0.99	0.99	0.99	264

<i>Accuracy</i>			0.98	1057
<i>Macro avg</i>	0.97	0.97	0.97	1057
<i>Weighted avg</i>	0.98	0.98	0.98	1057
Akurasi keseluruhan setelah <i>fine-tuning</i> : 97.54%				

Sebelum *fine-tuning*, akurasi model *VGG16* mencapai 91,39%. Precision tertinggi dicapai kelas *notumor* (0.98) dan *pituitary* (0.93), sedangkan *glioma* dan *meningioma* sama-sama 0.86. Dari sisi *recall*, *pituitary* menonjol dengan nilai 0.98, *notumor* 0.97, *glioma* 0.90, dan *meningioma* terendah di 0.79. *F1-score* tertinggi lagi-lagi pada *notumor* (0.98), diikuti *pituitary* (0.95), *glioma* (0.88), dan *meningioma* (0.82) seperti terlihat pada tabel 4.7.

TABEL 7
(CLASSIFICATION VGG16 SEBELUM FINE-TUNING)

VGG16 Classification Report (Before Fine-tuning)				
Precision		Recall	F-1 Score	Support
Glioma	0.86	0.90	0.88	243
Meningioma	0.86	0.79	0.82	247
Notumor	0.98	0.97	0.98	303
Pituitary	0.93	0.98	0.95	264
Accuracy			0.91	1057
Macro avg	0.91	0.91	0.91	1057
Weighted avg	0.91	0.91	0.91	1057
Akurasi keseluruhan sebelum <i>fine-tuning</i> : 91.39%				

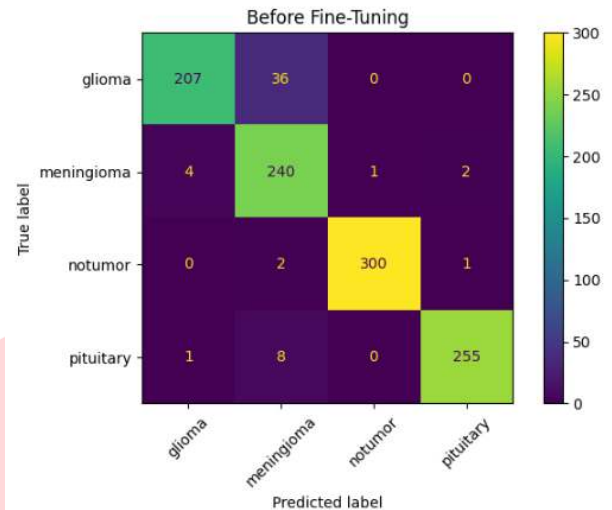
Setelah dilakukan *fine-tuning*, terjadi peningkatan tajam dengan akurasi keseluruhan sebesar 98,68%. Kelas *notumor* dan *pituitary* mencapai precision dan *recall* sempurna (1.00), sedangkan *glioma* dan *meningioma* juga mengalami peningkatan signifikan. *F1-score* untuk semua kelas mendekati nilai sempurna seperti yang terlihat ada tabel 4.8.

TABEL 8
(CLASSIFICATION VGG16 SETELAH FINE-TUNING)

VGG16 Classification Report (After Fine-tuning)				
Precision		Recall	F-1 Score	Support
Glioma	0.99	0.97	0.98	243
Meningioma	0.96	0.98	0.97	247
Notumor	1.00	1.00	1.00	303
Pituitary	0.99	1.00	0.99	264
Accuracy			0.99	1057
Macro avg	0.99	0.99	0.99	1057
Weighted avg	0.99	0.99	0.99	1057
Akurasi keseluruhan sesudah <i>fine-tuning</i> : 98.68%				

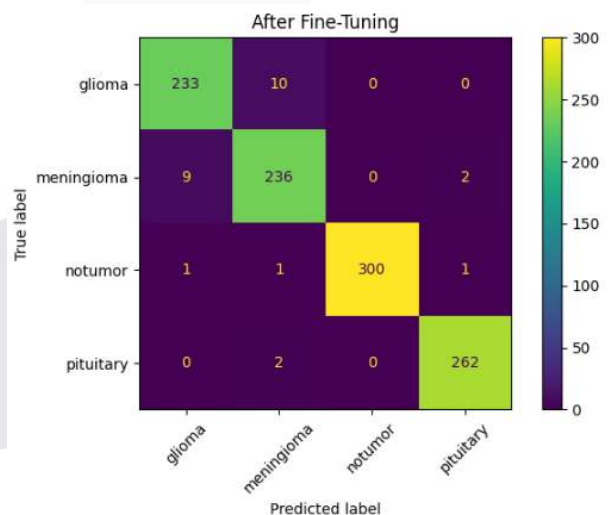
Selanjutnya evaluasi menggunakan *confusion matrix*. Pada model *ResNet50* sebelum *fine-tuning* pada gambar 4.21 terlihat bahwa performa klasifikasi awal sudah cukup baik. Kelas *glioma* memiliki 207 TP, namun masih menyisakan 36 FN dan 5 FP, menandakan adanya kesalahan dalam membedakan *glioma* dengan *meningioma*. Sementara itu, kelas *meningioma* menunjukkan jumlah FP yang cukup tinggi (46), menandakan model cenderung *overpredict* ke

kelas ini. Kelas *notumor* memiliki performa paling stabil dengan 300 TP, 3 FN, dan hanya 1 FP, menandakan model hampir selalu akurat dalam mengenali citra tanpa tumor. Kelas *pituitary* juga menunjukkan performa baik dengan 255 TP, hanya 9 FN dan 3 FP.



GAMBAR 21
(HASIL CONFUSION MATRIX RESNET50 SEBELUM FINE-TUNING)

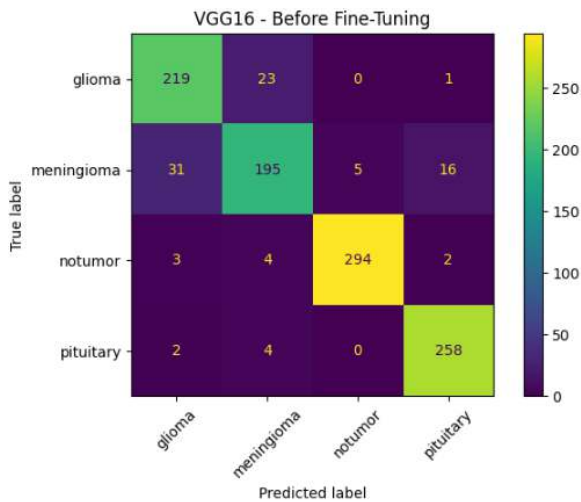
Setelah *fine-tuning*, performa *ResNet50* meningkat pada semua kelas. TP meningkat secara signifikan pada kelas *glioma* (233), *meningioma* (236), dan *pituitary* (262). Nilai FN dan FP menurun pada hampir semua kelas, terutama kelas *meningioma* yang semula memiliki 46 FP berkurang menjadi 13. Kelas *notumor* berhasil diprediksi hampir sempurna dengan 300 TP dan tanpa FP seperti yang terlihat pada gambar 4.22.



GAMBAR 22
(HASIL CONFUSION MATRIX RESNET50 SESUDAH FINE-TUNING)

Pada model *VGG16* sebelum *fine-tuning* terlihat bahwa meskipun performa awal cukup kompeten, masih terdapat cukup banyak kesalahan klasifikasi. Kelas *meningioma* menjadi yang paling banyak disalahklasifikasikan, dengan 52 FN dan 31 FP. Kelas *glioma* juga menunjukkan 36 FP, yang kebanyakan berasal dari kelas *meningioma*, mengindikasikan adanya kemiripan fitur antara kedua kelas ini. Sementara itu, kelas *notumor* dan *pituitary* menunjukkan hasil lebih stabil,

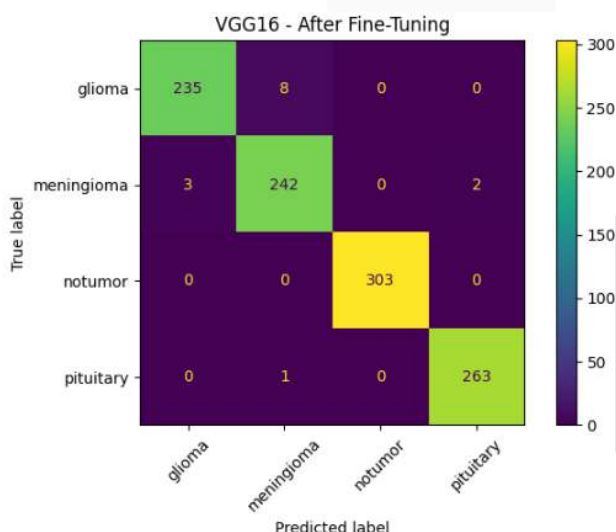
masing-masing dengan TP sebesar 294 dan 258, serta FN dan FP yang relatif kecil.



GAMBAR 23

(HASIL CONFUSION MATRIX VGG16 SEBELUM FINE-TUNING)

Setelah proses *fine-tuning* pada gambar 4.23, *VGG16* mengalami peningkatan signifikan dalam hal akurasi dan keandalan prediksi. Kelas *notumor* berhasil diklasifikasikan dengan sempurna (TP = 303, FN = 0, FP = 0), yang menunjukkan sensitivitas dan presisi yang sangat tinggi. Kelas *glioma* dan *meningioma* mengalami penurunan jumlah FN dan FP secara drastis. *Glioma* memiliki TP sebesar 235 dan hanya 8 FN serta 3 FP, sedangkan *meningioma* memiliki 242 TP dan hanya 5 FN dan 9 FP. Kelas *pituitary* juga menunjukkan hasil yang hampir sempurna dengan 262 TP, 1 FN, dan hanya 2 FP.



GAMBAR 24

(HASIL CONFUSION MATRIX VGG16 SETELAH FINE-TUNING)

V. KESIMPULAN

Berdasarkan hasil penelitian terkait perbandingan kinerja model transfer learning *VGG16* dan *ResNet50* dalam klasifikasi tumor otak pada citra MRI, dapat disimpulkan bahwa penerapan metode transfer learning terbukti efektif dalam tugas klasifikasi medis ini. Pada tahap awal pelatihan

tanpa *fine-tuning*, model *ResNet50* menunjukkan akurasi yang lebih tinggi dibandingkan *VGG16*, masing-masing sebesar 94,80% dan 91,39%. Namun, setelah dilakukan *fine-tuning*, performa kedua model mengalami peningkatan signifikan. *VGG16* mencatatkan akurasi tertinggi sebesar 98,68%, mengungguli *ResNet50* yang mencapai 97,54%. Hasil ini menunjukkan bahwa *VGG16* lebih unggul secara keseluruhan, baik dalam hal akurasi maupun stabilitas klasifikasi, pada dataset yang digunakan. Selain itu, proses *fine-tuning* terbukti mampu meningkatkan kemampuan generalisasi kedua model, memperkuat temuan bahwa penyesuaian lanjutan terhadap model pralatih memberikan dampak positif dalam meningkatkan performa klasifikasi citra medis.

REFERENSI

- [1] C. Watson, M. Kirkcaldie, and G. Paxinos, *The Brain: An Introduction to Functional Neuroanatomy*. CiNii Books, Oct. 31, 2023. [Online]. Available: <https://ci.nii.ac.jp/ncid/BB04049625> [Accessed: Oct. 30, 2024].
- [2] N. Shettigar, C. L. Yang, K. C. Tu, and C. S. Suh, "On the Biophysical Complexity of Brain Dynamics: An Outlook," *Dynamics*, vol. 2, no. 2, pp. 114–148, 2022.
- [3] J. Ferlay *et al.*, "Global Cancer Observatory: Cancer Today," *International Agency for Research on Cancer*, Lyon, France, 2020. [Online]. Available: <https://gco.iarc.fr/today> [Accessed: Oct. 29, 2024].
- [4] D. N. Louis *et al.*, "The 2021 WHO Classification of Tumors of the Central Nervous System: A Summary," *Neuro-Oncology*, vol. 23, no. 8, pp. 1231–1251, 2021.
- [5] M. Matia *et al.*, "Magnetic Resonance Image of Primary Adult Brain Tumors: State of the Art and Future Perspectives," *Biomedicines*, vol. 11, no. 2, pp. 364, 2023.
- [6] A. B. Abdulalomov, M. Mukhiddinov, and T. K. Whangbo, "Brain Tumor Detection Based on Deep Learning Approaches and Magnetic Resonance Imaging," *Cancers (Basel)*, vol. 15, no. 16, pp. 4172, 2023.
- [7] A. Kascenas, N. Pugeault, and A. Q. O'Neil, "Denoising Autoencoders for Unsupervised Anomaly Detection in Brain MRI," in *Proc. Medical Imaging with Deep Learning*, Edinburgh, Scotland, vol. 172, pp. 653–664, 2022.
- [8] A. J. Rozaqi, A. Sunyoto, and M. R. Arief, "Implementasi Transfer Learning pada Algoritme Convolutional Neural Network untuk Identifikasi Penyakit Daun Kentang," *Procedia Engineering and Life Sciences*, vol. 1, no. 1, 2021.
- [9] K. R. Reddy and R. Dhuli, "A Novel Lightweight CNN Architecture for the Diagnosis of Brain Tumors Using MR Images," *Diagnostics*, vol. 13, no. 2, pp. 312, 2023.
- [10] American Association of Neurological Surgeons, "Brain Tumors." [Online]. Available: <https://www.aans.org/patients/conditions-treatments/brain-tumors/> [Accessed: Nov. 2, 2024].
- [11] National Institute of Biomedical and Bioengineering, "Magnetic Resonance Imaging (MRI)," Apr. 2022. [Online]. Available: <https://www.nibib.nih.gov/sites/default/files/2022->

- 05/Fact-Sheet-Magnetic-Resonance-Imaging-MRI.pdf [Accessed: Nov. 2, 2024].
- [12] D. Bhavani, A. Vasari, and P. T. Keshava, "Machine Learning: A Critical Review of Classification Techniques," *International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE)*, vol. 3, no. 11, pp. 17–23, 2016.
- [13] K. Sharifani and M. Amini, "Machine Learning and Deep Learning: A Review of Methods and Applications," *World Information Technology and Engineering Journal*, vol. 10, no. 7, pp. 3897–3904, 2023.
- [14] R. Geetha and T. Thilagam, "A Review on the Effectiveness of Machine Learning and Deep Learning Algorithms for Cyber Security," *Archives of Computational Methods in Engineering*, vol. 28, no. 4, pp. 2861–2879, 2021.
- [15] A. Hosna, E. Merry, J. Gyalmo, Z. Alom, Z. Aung, and M. Abdul Azim, "Transfer Learning: A Friendly Introduction," *Journal of Big Data*, vol. 9, no. 1, pp. 102, 2022.
- [16] Amanatulla, "Fine-Tuning the Model: What, Why, and How," *Medium*, Sep. 21, 2023. [Online]. Available: <https://medium.com/@amanatulla1606/fine-tuning-the-model-what-why-and-how-e7fa52bc8ddf> [Accessed: May 30, 2025].
- [17] A. Garg, "Image Classification Using ResNet-50 Deep Learning Model," *Data Science Blogathon*, Sep. 20, 2022. [Online]. Available: <https://www.analyticsvidhya.com/blog/2022/09/image-classification-in-stl-10-dataset-usingresnet-50-deep-learning-model/> [Accessed: Nov. 9, 2024].
- [18] V. Kurama, "A Review of Popular Deep Learning Architectures: AlexNet, VGG16, and GoogleNet," *DigitalOcean*, Sep. 25, 2024. [Online]. Available: <https://www.digitalocean.com/community/tutorials/popular-deep-learning-architectures-alexnet-vgg-googlenet> [Accessed: Nov. 9, 2024].
- [19] H. Naviwala, "What is a Confusion Matrix? Understand the 4 Key Metric of its Interpretation," *Data Science Dojo*, Sep. 23, 2024. [Online]. Available: <https://datasciencedojo.com/blog/confusion-matrix/> [Accessed: Nov. 9, 2024].
- [20] Deepanshi, "In-depth Understanding of Confusion Matrix," *Data Science Blogathon*, Nov. 21, 2023. [Online]. Available: <https://www.analyticsvidhya.com/blog/2021/05/in-depth-understanding-of-confusion-matrix/> [Accessed: Nov. 9, 2024].
- [21] V. Brusica and J. Zeleznikow, "Knowledge Discovery and Data Mining in Biological Databases," *The Knowledge Engineering Review*, vol. 14, no. 3, pp. 257–277, 1999.