ANALISIS KINERJA MODEL *VISION TRANSFORMER* (VIT) MELALUI CITRA *MAGNETIC RESONANCE IMAGING* (MRI) DALAM DETEKSI TUMOR OTAK

Skripsi

Oleh

BENAYA
NPM. 2217031082

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2026

ABSTRACT

PERFORMANCE ANALYSIS OF THE VISION TRANSFORMER (VIT) MODEL VIA MAGNETIC RESONANCE IMAGING (MRI) IN BRAIN TUMOR DETECTION

By

Benaya

Accurate brain tumor detection using Magnetic Resonance Imaging (MRI) poses a significant challenge in medical diagnosis. This study aims to analyze the performance of Vision Transformer (ViT) model in brain tumor classification and identify critical factors affecting its performance. The dataset consists of 7,023 MRI images classified into four classes: glioma, meningioma, pituitary, and no tumor. Research stages include data preprocessing (grayscale, Gaussian blur, thresholding, morphology, contour detection, cropping, and bone color mapping), data augmentation, and ViT model fine-tuning using AdamW optimizer. Experiments were conducted under two scenarios: data balancing through class weighting and augmentation, and comparison between comprehensive versus partial fine-tuning strategies. Results demonstrate that ViT model achieves optimal performance with 97.08% accuracy, 97.01% precision, 97.01% recall, and 97.01% F1-score. Data balancing improves training stability and model generalization with 0.57% accuracy enhancement. Comprehensive fine-tuning proves 7.47% superior compared to partial fine-tuning despite requiring longer computational time. Benchmarking with previous studies confirms ViT performance competitively to the best CNN architectures. This research contributes to the development of reliable artificial intelligence-based diagnostic systems for brain tumor detection.

Keywords: Vision Transformer, brain tumor, MRI, classification, fine-tuning.

ABSTRAK

ANALISIS KINERJA MODEL *VISION TRANSFORMER* (VIT) MELALUI CITRA *MAGNETIC RESONANCE IMAGING* (MRI) DALAM DETEKSI TUMOR OTAK

Oleh

Benaya

Deteksi tumor otak secara akurat menggunakan citra *Magnetic Resonance Imaging* (MRI) merupakan tantangan utama dalam diagnosis medis. Penelitian ini bertujuan menganalisis kinerja model *Vision Transformer* (ViT) dalam klasifikasi tumor otak serta mengidentifikasi faktor-faktor kritis yang memengaruhi performanya. Dataset terdiri atas 7.023 citra MRI yang diklasifikasikan ke dalam empat kelas: glioma, meningioma, pituitary, dan no tumor. Tahapan penelitian mencakup preprocessing data (*grayscale, Gaussian blur, thresholding,* morfologi, deteksi kontur, *cropping*, dan *bone color mapping*), augmentasi data, serta *fine-tuning* model ViT dengan *optimizer* AdamW. Eksperimen dilakukan dengan dua skenario: penyeimbangan data menggunakan pembobotan kelas dan augmentasi, serta perbandingan strategi fine-tuning komprehensif versus parsial. Hasil penelitian menunjukkan bahwa model ViT mencapai kinerja optimal dengan akurasi 97,08%, presisi 97,01%, recall 97,01%, dan F1-score 97,00%. Penyeimbangan data meningkatkan stabilitas pelatihan dan generalisasi model dengan peningkatan akurasi sebesar 0,57%. *Fine-tuning* komprehensif terbukti 7,47% lebih unggul dibandingkan *fine-tuning* parsial meskipun membutuhkan waktu komputasi lebih lama. *Benchmarking* dengan penelitian terdahulu mengonfirmasi performa ViT yang kompetitif dengan arsitektur CNN terbaik. Penelitian ini berkontribusi pada pengembangan sistem diagnosis berbasis kecerdasan buatan yang andal untuk deteksi tumor otak.

Kata kunci: *Vision Transformer*, tumor otak, MRI, klasifikasi, *fine-tuning*

ANALISIS KINERJA MODEL *VISION TRANSFORMER* (VIT) MELALUI CITRA *MAGNETIC RESONANCE IMAGING* (MRI) DALAM DETEKSI TUMOR OTAK

BENAYA

Skripsi

Sebagai Salah Satu Syarat untuk Memperoleh Gelar
SARJANA MATEMATIKA

Pada

Jurusan Matematika

Fakultas Matematika dan Ilmu Pengetahuan Alam

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2026

Judul Skripsi	:	**ANALISIS KINERJA MODEL *VISION TRANSFORMER* (VIT) MELALUI CITRA *MAGNETIC RESONANCE IMAGING* (MRI) DALAM DETEKSI TUMOR OTAK**
Nama Mahasiswa	:	**Benaya**
Nomor Pokok Mahasiswa	:	**2217031082**
Program Studi	:	**Matematika**
Fakultas	:	**Matematika dan Ilmu Pengetahuan Alam**

MENYETUJUI

1. Komisi Pembimbing

Dr. Dian Kurniasari, S.Si., M.Sc.	**Favorisen R. Lumbanraja, S.Kom., M.Si, Ph.D.**
NIP 196903051996032001	NIP 198301102008121002

2. Ketua Jurusan Matematika

Dr. Aang Nuryaman, S.Si., M.Si.
NIP. 19740316200501 1001

Dipindai dengan CamScanner

MENGESAHKAN

1. Tim Penguji

Ketua : **Dr. Dian Kurniasari, S.Si., M.Sc.** .....

Sekretaris : **Favorisen R. Lumbanraja, S.Kom., M.Si, Ph.D.** .....

Penguji
Bukan Pembimbing : **Ir. Warsono, M.S., Ph.D.** .....

2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam

Dr. Eng. Heri Satria, S.Si., M.Si.
NIP. 197110012005011002

Tanggal Lulus Ujian Skripsi: **10 Maret 2026**

Dipindai dengan CamScanner

PERNYATAAN SKRIPSI MAHASISWA

Yang bertanda tangan di bawah ini:

Nama	:	**Benaya**
Nomor Pokok Mahasiswa	:	**2217031082**
Jurusan	:	**Matematika**
Judul Skripsi	:	**Analisis Kinerja Model *Vision Transformer* (ViT) melalui Citra *Magnetic Resonance Imaging* (MRI) dalam Deteksi Tumor Otak**

Dengan ini menyatakan bahwa skripsi ini adalah hasil pekerjaan saya sendiri. Apabila kemudian hari terbukti bahwa skripsi ini merupakan hasil salinan atau dibuat oleh orang lain, maka saya bersedia menerima sanksi sesuai dengan ketentuan akademik yang berlaku.

Bandar Lampung, 10 Maret 2026

Penulis,

Benaya

Dipindai dengan CamScanner

RIWAYAT HIDUP

Penulis memiliki nama lengkap Benaya yang lahir di Jambi pada tanggal 19 Juni 2002. Selama berkuliah, penulis aktif mengikuti perlombaan tingkat nasional dan organisasi pers mahasiswa.

Riwayat pendidikan penulis dimulai di TK Xaverius 2 Jambi pada tahun 2008 sampai dengan 2009 Kemudian penulis menempuh pendidikan dasar di SD Xaverius 2 Jambi pada tahun 2009 sampai dengan 2015. Selanjutnya, penulis melanjutkan pendidikan di SMP Xaverius 1 Jambi pada tahun 2015 sampai dengan 2018, yang dilanjutkan dengan menempuh pendidikan di SMA Xaverius 1 Jambi pada tahun 2018 sampai dengan 2021. Pada tahun 2022, penulis diterima sebagai mahasiswa S1 Jurusan Matematika di Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA) Universitas Lampung melalui jalur SBMPTN.

Penulis mengikuti organisasi Pers Mahasiswa Natural sebagai kepala biro kesekretariatan pada tahun 2023. Penulis juga aktif mengikuti perlombaan sebagai bentuk penyaluran kemampuan penulis dalam bidang akademik. Lomba pertama yang diikuti penulis adalah *Mathematical Analysis Geometry Day* (MaG-D) pada Mei 2023 sebagai semifinalis. Lomba berikutnya adalah *Sanata Dharma Calculus League* pada November 2023 sebagai semifinalis.

Pada tahun 2024, penulis mengembangkan kemampuan kalkulus dan analisis melalui berbagai perlombaan. Penulis mengikuti lomba makalah dan poster ke-2 yang diselenggarakan oleh Komunitas Analisis Matematika (Kamindo) dan Institut Teknologi Bandung (ITB) pada bulan Juli sebagai peraih medali perunggu, lomba Olimpiade Matematika Mahasiswa Tingkat Nasional yang diselenggarakan oleh Universitas Bengkulu pada bulan September sebagai peraih juara 2, dan lomba *Sanata Dharma Calculus League* pada bulan November sebagai peraih juara harapan 1. Selain itu, penulis berpartisipasi pada ONMIPA tingkat regional yang

mewakili Universitas Lampung pada bidang Matematika.

Penulis mengikuti Kuliah Kerja Nyata (KKN) di Desa Priangan Baru, Kecamatan Tanjung Raja, Kabupaten Lampung Utara pada awal 2025 dan melaksanakan kerja praktik di Karya Nusa Tujuh pada pertengahan 2025. Pada tahun 2025, penulis mengeksplor berbagai lomba. Penulis mengikuti lomba penyajian data oleh WikiData pada bulan Mei sebagai peraih juara 4, lomba MaG-D pada bulan Juni sebagai peraih medali perak, dan Lomba Esai Matematika Nasional (LEMNAS) yang diselenggarakan oleh Universitas Pendidikan Indonesia (UPI) pada bulan Desember sebagai peraih juara 1. Penulis juga pernah mengikuti ONMIPA di tingkat nasional sebagai perwakilan Universitas Lampung pada bidang Matematika.

KATA INSPIRASI

Kesuksesan bukanlah akhir, kegagalan bukanlah sesuatu yang fatal: yang terpenting adalah keberanian untuk melanjutkan.
(Winston Churchill)

Yang penting adalah jangan berhenti bertanya. Rasa ingin tahu memiliki alasannya sendiri untuk eksis.
(Albert Einstein)

Gunung tidak pernah mengakui dirinya tinggi
(Anonim)

"Saat dunia terasa sunyi dan tak ada yang menemani, ingatlah: Sang Terang selalu hadir, memberi nafas iman dan menemani setiap langkahmu. Maka, jangan pernah merasa sendiri—karena terang-Nya tak pernah padam, bahkan di lorong paling gelap sekalipun."
(Anonim)

PERSEMBAHAN

Dengan mengucap syukur kepada Tuhan Yang Maha Esa atas berkat dan penyertaan-Nya, skripsi ini dapat terselesaikan dengan baik dan tepat pada waktunya. Dengan rasa syukur dan bahagia, saya persembahkan rasa terimakasih saya kepada:

Ayah dan Ibuku Tercinta

Terimakasih kepada orang tuaku atas segala pengorbanan, motivasi, doa, dan cinta selama saya hidup dalam pelukan hangat. Terimakasih telah memberikan pelajaran berharga yang mendewasakan diriku hingga saat ini. Pengajaranmu membuat diriku tumbuh sebagai garam dan terang dunia.

Dosen Pembimbing dan Pembahas

Terimakasih kepada dosen pembimbing dan pembahas yang sudah sangat membantu, memberikan motivasi, memberikan arahan serta ilmu yang berharga.

Sahabat-sahabatku

Terima kasih kepada sahabat dan teman-teman terkasih, yang telah mewarnai perjalanan ini dengan tawa, air mata, dan kenangan yang tak terlupakan. Setiap semangat, doa, dan dukungan kalian menjadi energi yang menguatkan langkahku.

Almamater Tercinta
Universitas Lampung

SANWACANA

Puji dan syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas berkat dan karunia-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul ”Analisis Kinerja Model *Vision Transformer* (ViT) melalui Citra *Magnetic Resonance Imaging* (MRI) dalam Deteksi Tumor Otak” dengan baik, lancar, dan tepat pada waktu yang telah ditentukan.

Dalam proses penyusunan skripsi ini, banyak pihak yang telah membantu memberikan bimbingan, dukungan, arahan, motivasi serta saran sehingga skripsi ini dapat terselesaikan. Oleh karena itu, dalam kesempatan ini penulis mengucapkan terimakasih kepada:

1. Ibu Dr. Dian Kurniasari, S.Si., M.Sc. selaku Pembimbing I yang telah banyak meluangkan waktunya untuk memberikan arahan, bimbingan, motivasi, saran serta dukungan kepada penulis sehingga dapat menyelesaikan skripsi ini.
2. Bapak Favorisen R. Lumbanraja, S.Kom., M.Si, Ph.D. selaku Pembimbing II yang telah memberikan arahan, bimbingan dan dukungan kepada penulis sehingga dapat menyelesaikan skripsi ini.
3. Bapak Ir. Warsono, M.S., Ph.D. selaku Penguji yang telah bersedia memberikan kritik dan saran serta evaluasi kepada penulis sehingga dapat menjadi lebih baik lagi.
4. Bapak Dr. Aang Nuryaman, S.Si., M.Si. selaku Ketua Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
5. Ibu Dr. Muslim Ansori, M.Si. selaku dosen pembimbing akademik yang telah rendah hati dalam memberikan arahan selama berkuliah.
6. Seluruh dosen, staff dan karyawan Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.

7. Ayah, Ibu, dan Kakak yang senantiasa memberikan doa, harapan, cinta, dan motivasi kepada penulis.
8. Sahabat penulis sejak SMA diantaranya Pattrick Ritter, Fransisko Gilbert, dan Delbert Amadeus Tan yang telah banyak memberikan tawa dan selalu meluangkan waktu untuk mendengarkan ceritaku.
9. Teman seperjuangan dalam satu bimbingan yang telah membantu dan memotivasi selama menyusun skripsi.
10. Semua teman dan kerabat baik di dalam maupun di luar Universitas Lampung yang menemani karier perkuliahan.

Semoga skripsi ini dapat bermanfaat bagi kita semua. Penulis menyadari bahwa skripsi ini masih jauh dari sempurna, sehingga penulis mengharapkan kritik dan saran yang membangun untuk menjadikan skripsi ini lebih baik lagi.

Bandar Lampung, 10 Maret 2026

Benaya

DAFTAR ISI

Halaman

ABSTRACT . . . . . **ii**
ABSTRAK . . . . . **iii**
HALAMAN PENGESAHAN . . . . . **iii**
HALAMAN PERNYATAAN . . . . . **iv**
HALAMAN MOTTO . . . . . **vii**
HALAMAN PERSEMBAHAN . . . . . **viii**
SANWACANA . . . . . **xi**
DAFTAR ISI . . . . . **xiii**
DAFTAR TABEL . . . . . **xiv**
DAFTAR GAMBAR . . . . . **xiv**
I PENDAHULUAN **1**
1.1 Latar Belakang Masalah 1
1.2 Rumusan Masalah 4
1.3 Tujuan Penelitian 4
1.4 Manfaat Penelitian 5
II TINJAUAN PUSTAKA **6**
2.1 Penelitian Terkait 6
2.1.1 Penelitian Pertama (Pratama et al., 2024) 9
2.1.2 Penelitian Kedua (Garcia et al., 2025) 10
2.1.3 Penelitian Ketiga (Kurniasari et al., 2025a) 12
2.1.4 Penelitian Keempat (Sanhaji et al., 2025) 14
2.1.5 Penelitian Kelima (Husen, 2024) 15
2.1.6 Penelitian Keenam (Andre et al., 2021) 16
2.1.7 Penelitian Ketujuh (Kurniasari et al., 2025b) 17
2.2 Tumor Otak 18
2.3 Pencitraan *Magnetic Resonance Imaging* (MRI) 20
2.4 *Artificial Intelligence* (AI) 20
2.4.1 *Machine Learning* 21

2.4.2 *Deep Learning* . . . 22
2.4.3 *Transfer Learning* . . . 22
2.5 *Vision Transformer* (ViT) . . . 23
2.5.1 *Embedding Encoding* . . . 25
2.5.2 *Positional Encoding* . . . 25
2.5.3 *Multi-Head Self-Attention* . . . 26
2.6 *Preprocessing Data* . . . 27
2.6.1 *Graysale* . . . 28
2.6.2 *Gaussian Blur* . . . 29
2.6.3 *Binary Thresholding* . . . 29
2.6.4 *Erosion* dan *Dilation* . . . 30
2.6.5 *Contour Detection* . . . 31
2.6.6 *Cropping* . . . 32
2.6.7 *Bone Color Mapping* . . . 33
2.7 *Split Data* . . . 33
2.7.1 Data Latih (*Training*) . . . 34
2.7.2 Data Validasi (*Validation*) . . . 35
2.7.3 Data Uji (*Testing*) . . . 35
2.8 Augmentasi Data . . . 36
2.8.1 Metode Augmentasi Sampel Tunggal . . . 36
2.8.2 Metode Augmentasi Gabungan Banyak Citra . . . 37
2.8.3 Metode Berbasis GANs (*Generative Adversarial Networks*) 37
2.9 *Fine Tuning* . . . 38
2.10 Teorema Bayes . . . 39
2.11 Evaluasi Performa Model . . . 41
2.11.1 Akurasi . . . 43
2.11.2 Presisi . . . 44
2.11.3 *Sensitivity* (*Recall*) . . . 44
2.11.4 *Specificity* . . . 45
2.11.5 *F1 Score* . . . 46
III METODE PENELITIAN . . . 47
3.1 Waktu dan Tempat Penelitian . . . 47
3.1.1 Tempat Penelitian . . . 47
3.1.2 Waktu Penelitian . . . 47
3.2 Data Penelitian . . . 47

3.3 Alat Penelitian 48
3.4 Langkah Penelitian 50
IV HASIL DAN PEMBAHASAN 53
4.1 Input Data 53
4.2 Distribusi Data 53
4.3 *Cleansing Data* 54
4.4 *Prepocessing Data* 55
4.5 *Splitting Data* 58
4.6 Augmentasi Data 59
4.7 Eksperimen Awal 61
4.7.1 *Fine-Tunning* 61
4.7.2 Evaluasi Model 65
4.8 Analisis Faktor-Faktor Kritis yang Mempengaruhi Kinerja 78
4.8.1 Penyeimbangan Data 78
4.8.2 Strategi *Fine Tunning* 83
4.9 *Benchmarking* Penelitian Terdahulu 86
V KESIMPULAN DAN SARAN 89
5.1 Kesimpulan 89
5.2 Saran 90
DAFTAR PUSTAKA 91

DAFTAR TABEL

Tabel Halaman

1 Penelitian Terkait ViT dan Klasifikasi Tumor Otak 6

2 Distribusi Dataset Tumor Otak 48
3 Spesifikasi Alat Penelitian 50

4 Pembagian Data untuk Klasifikasi Tumor Otak 58
5 Durasi *Fine Tuning* dengan Model ViT 63
6 Hasil Evaluasi Klasifikasi 77
7 Perbandingan Kinerja Model Berdasarkan Data 82
8 Perbandingan Durasi Fine-Tuning Model ViT 83
9 Perbandingan Kinerja Model Berdasasrkan Stategi *Fine Tunning* . . 85
10 *Benchmarking* Hasil Penelitian 88

DAFTAR GAMBAR

Gambar Halaman

1 Arsitektur Model ViT (Dosovitskiy et al., 2021). 24
2 *Transformer Encoder* (Dosovitskiy et al., 2021). 25
3 Struktur *Multi-Head Attention* (Zhou et al., 2025). 27
4 *Confussion Matrix*: (a) Dua kelas (b) *Multi class* (Markoulidakis et al., 2021). 42
5 Alur Tahapan Penelitian. 52
6 Contoh Data pada Setiap Kelas. 53
7 Visualisasi Distribusi Data pada Setiap Kelas. 54
8 Hasil *Preprocessing* Data dari *Grayscale* hingga Dilatasi. 56
9 Hasil Deteksi Kontur. 57
10 Hasil *Cropping* dan *Resize*. 57
11 Hasil *Preprocessing* Data *Bone Color Mapping*. 58
12 Hasil Augmentasi Data. 59
13 Grafik Durasi *Fine Tunning* dengan Model ViT. 63
14 Grafik *Loss* pada Model ViT. 64
15 Grafik Akurasi Model ViT. 65
16 *Confussion Matrix* pada Model ViT. 66
17 Sampel Gambar yang Misklasifikasi. 68
18 Gambar yang Salah Diprediksi Sebagai *Meningioma*. 68
19 Hasil Pembobotan Kelas dari Setiap Kelas. 79
20 Hasil Penyeimbangan Data. 80
21 Grafik *Loss* pada Model ViT dengan Data yang Diseimbangkan. . . 81
22 Grafik Akurasi pada Model ViT dengan Data yang Diseimbangkan. 82

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Tumor otak dikenal sebagai penyakit pada otak yang ditandai dengan adanya pertumbuhan sel-sel jaringan otak yang abnormal di dalam otak ataupun di sekitar otak. Tumor otak dapat bersifat jinak atau ganas yang dapat mendesak ruang pada otak dan desakan tersebut dapat meningkatkan tekanan intracranial. Tumor otak dapat dibagi menjadi primer dan sekunder. Tumor otak primer merupakan pertumbuhan sel abnormal yang terjadi di dalam otak (Kepakisan et al., 2022), sedangkan tumor otak sekunder adalah pertumbuhan sel abnormal yang terjadi di bagian tubuh lain yang terkena kanker (Andre et al., 2021).

Tumor otak merupakan salah satu penyakit mematikan di dunia dan gangguan neurologis serius yang ditandai dengan tingkat mortalitas tinggi, terutama untuk kasus glioblastoma yang sangat agresif dan sulit diobati. Jika melihat kasus tumor otak secara global, angka kejadian tumor otak terus meningkat setiap tahunnya, yaitu mencapai sekitar 308.102 kasus di seluruh dunia pada tahun 2020 dengan insidensi yang berkisar sekitar 10,82 per 100.000 penduduk setiap tahun (Pratama et al., 2024). Benua Amerika dan Eropa memperlihatkan insidensi tumor otak yang meningkat dari 17.6/100.000 menjadi 22.0/100.000 populasi, di mana sekitar 18.500 kasus baru tumor otak primer didiagnosis setiap tahun. Keseriusan ini diperparah di Indonesia, di mana angka mortalitas pada pasien tumor otak primer mencapai 4,25 per 100.000 populasi per tahun, dengan angka kejadian sebesar 7 per 100.000 populasi (Andre et al., 2021).

Diagnosis tumor otak diperlukan proses yang kompleks dan membutuhkan kecepatan serta akurasi tinggi karena penyakit ini dapat mengancam nyawa jika tidak ditangani dengan baik (Husen, 2024). Diagnosis ini masih didominasi oleh teknik citra medis seperti *Magnetic Resonance Imaging* (MRI) dan *Computed*

Tomography (CT) *scan* yang memiliki keterbatasan seperti fasilitas diagnosis yang modern, waktu interpretasi yang relatif lama, dan dibutuhkan operator yang berpengalaman. Jika dipandang dalam sisi pengembangan model, masalah yang sering ditemui adalah ketersediaan data medis yang sedikit, resiko *overfitting* dalam membangun model tanpa augmentasi data, dan keterbatasan dalam mengenali data (Sanhaji et al., 2025).

Convolutional Neural Network (CNN) merupakan arsitektur yang mendominasi pengolahan citra medis karena memiliki bias induktif spasial yang sangat kuat (Kawadkar, 2025). Model ini bekerja dengan menerapkan filter atau kernel secara berlapis untuk mengekstraksi fitur visual dari data input (Salehi et al., 2023). Lapisan konvolusi memungkinkan identifikasi pola lokal yang efisien melalui deteksi fitur sederhana hingga fitur yang lebih kompleks. Implementasi model seperti *ResNet* dan *EfficientNet* telah membuktikan efektivitasnya dalam berbagai tugas klasifikasi gambar yang bersifat spesifik. Fokus utama arsitektur konvensional ini adalah pada hubungan piksel yang berdekatan untuk memahami struktur spasial dalam citra secara mendalam (Chibuike & Yang, 2024).

Integrasi antara CNN dan *transfer learning* muncul sebagai strategi untuk menggabungkan ekstraksi fitur lokal dengan pemahaman konteks global. Struktur hibrida ini memanfaatkan lapisan konvolusi untuk menangkap fitur spasial dasar sebelum diproses oleh mekanisme atensi. Pendekatan gabungan tersebut bertujuan mengatasi keterbatasan unit konvolusi dalam memodelkan ketergantungan jarak jauh antar wilayah citra. Model seperti RMT-Net mengintegrasikan kedua arsitektur tersebut untuk menghasilkan representasi fitur yang lebih komprehensif bagi sistem diagnosis klinis. Sinergi ini memberikan fleksibilitas dalam menangani variasi data medis yang memerlukan detail fitur lokal sekaligus pemahaman citra secara menyeluruh (Chibuike & Yang, 2024).

Vision Transformer (ViT) merevolusi visi komputer dengan memperlakukan citra sebagai urutan patch yang diproses melalui mekanisme perhatian mandiri murni. Model ini mampu menangkap representasi global secara langsung sejak lapisan pertama sehingga sangat efektif untuk tugas diagnostik yang kompleks (Rodrigo et al., 2024). Data penelitian menunjukkan bahwa ViT sering kali menghasilkan akurasi yang lebih tinggi dengan variansi performa yang lebih rendah pada berbagai modalitas pencitraan. Keunggulan ViT terlihat sangat nyata dalam klasifikasi

tumor otak di mana model ini mampu mencapai performa yang jauh lebih superior dibandingkan arsitektur tradisional (Kawadkar, 2025). Stabilitas pelatihan serta ketahanan tinggi terhadap distorsi visual menjadikan penggunaan ViT sebagai pilihan yang sangat baik untuk memajukan standar akurasi dalam penelitian masa depan (Rodrigo et al., 2024).

Penelitian dalam klasifikasi tumor otak telah didominasi oleh arsitektur berbasis CNN dengan berbagai tingkat keberhasilan. Studi-studi terdahulu menunjukkan pencapaian akurasi yang terus meningkat melalui optimasi arsitektur dan metode hibrida. Beberapa contoh representatif antara lain:

- Model ringan seperti *MobileNetV2* berhasil mencapai akurasi sebesar 88.64% untuk klasifikasi empat kelas tumor otak (Winarto et al.,2022).
- Studi oleh Archana et al. (2023) melaporkan pencapaian akurasi 94.76% menggunakan arsitektur *AlexNet*, sementara penelitian komparatif lain oleh Wani et al. (2023) yang melibatkan *AlexNet*, *GoogleNet*, dan *VGG-19* mencapai akurasi maksimal 90.625%.
- Pendekatan hibrida yang menggabungkan fitur dari *VGG16* dengan *Decision Tree (DT)* sebagai *classifier* menunjukkan peningkatan akurasi menjadi 96.08% (Rajinikanth et al.,2022).
- Penelitian yang mengimplementasikan *Vision Transformer (ViT)* pada penyakit Alzheimer menunjukkan hasil yang sangat kompetitif dengan akurasi hingga 98.19% (Kurniasari et al., 2025).

Data di atas menunjukkan bahwa meskipun arsitektur CNN mampu memberikan performa yang baik, masih terdapat potensi yang signifikan untuk peningkatan akurasi dan robustisitas model.

Penelitian ViT sangat penting karena arsitektur ViT, terutama ketika diintegrasikan dengan teknik *deep learning* lainnya, menawarkan jalur menuju tingkat akurasi dan efisiensi dalam diagnosis tumor yang belum pernah ada sebelumnya. ViT memiliki keunggulan unik dalam kemampuannya mengekstrak asosiasi jarak jauh dari data citra melalui mekanisme *self-attention*, yang sangat penting untuk interpretasi citra yang lebih komprehensif (Jraba et al., 2023). Dengan meningkatkan akurasi diagnostik secara signifikan, sistem berbasis ViT dapat membantu mengurangi

misdiagnosis dan kasus positif palsu. Pada akhirnya, peningkatan deteksi dini dan intervensi yang tepat waktu melalui alat diagnostik yang andal dan sangat akurat berpotensi meningkatkan perawatan pasien dan kualitas hidup secara keseluruhan (Halder et al., 2024).

Kontribusi utama dari penelitian ini adalah pengembangan dan validasi sebuah model klasifikasi tumor otak berbasis Vision Transformer yang terbukti lebih akurat dan robust dibandingkan arsitektur konvensional. Dengan menunjukkan secara empiris superioritas ViT dalam menangkap konteks global pada citra MRI, penelitian ini bertujuan untuk menghasilkan sebuah alat bantu yang mampu menyederhanakan proses pengambilan keputusan klinis, menawarkan jalur diagnosis yang lebih andal dan efisien bagi para profesional kesehatan.

1.2 Rumusan Masalah

Rumusan masalah pada penelitian adalah sebagai berikut:

1. Bagaimana efektivitas dan kinerja model Vision Transformer (ViT) dalam deteksi tumor otak berdasarkan citra MRI, jika dilihat dari metrik akurasi, presisi, sensitivitas, spesifisitas, dan *F1-Score*?
2. Bagaimana pengaruh penerapan strategi penyeimbangan data, seperti pembobotan kelas dan augmentasi, terhadap stabilitas pelatihan, risiko *overfitting*, dan peningkatan kinerja model?
3. Bagaimana perbandingan efektivitas antara strategi *fine-tuning* komprehensif (pada seluruh lapisan) dengan *fine-tuning* parsial (hanya pada *head classifier*) terhadap optimalisasi kemampuan representasi fitur dan akurasi model ViT pada data medis?

1.3 Tujuan Penelitian

Adapun tujuan dari penelitian ini yaitu:

1. Mengkaji model ViT pada dataset citra MRI tumor otak untuk mengklasifikasikan jenis tumor dan memahami hal yang dibutuhkan dalam membangun model ViT.

2. Menganalisis tingkat akurasi, presisi, *recall*, dan *F1-score* yang dihasilkan oleh model ViT dalam proses klasifikasi untuk mengetahui performa model pada data MRI tumor otak.
3. Mengidentifikasi dan menganalisis faktor-faktor kritis yang memengaruhi performa ViT dalam konteks klasifikasi citra medis, termasuk aspek konfigurasi model (seperti ukuran patch, *learning rate*, dan *weight decay*), strategi tuning, serta teknik pra-pemrosesan data yang spesifik untuk domain MRI.

1.4 Manfaat Penelitian

Adapun manfaat dari penelitian ini yaitu:

1. Memberikan kontribusi dalam pengembangan teknologi *deep learning* untuk klasifikasi citra medis.
2. Menjadi referensi bagi penelitian selanjutnya yang ingin mengembangkan sistem berbasis kecerdasan buatan dalam deteksi penyakit.
3. Membantu dokter dan tenaga medis dalam melakukan diagnosis tumor otak secara lebih cepat dan akurat.
4. Mengurangi risiko kesalahan manusia dalam menginterpretasikan hasil pencitraan MRI.

BAB II

TINJAUAN PUSTAKA

2.1 Penelitian Terkait

Penelitian terkait menjadi acuan dalam penulisan dan perbandingan hasil klasifikasi. Topik penelitian yang menjadi acuan adalah metode *Vision Transformer* (ViT) dan klasifikasi tumor otak. Tabel 1 menjelaskan penelitian terkait penerapan ViT dan klasifikasi tumor otak.

Tabel 1 Penelitian Terkait ViT dan Klasifikasi Tumor Otak

No.	Penelitian	Data	Metode	Hasil
1.	Perbandingan Model Klasifikasi *Transfer Learning Convolutional Neural Network* Tumor Otak Menggunakan Citra *Magnetic Resonance Imaging* (Pratama et al., 2024)	3.264 citra MRI *(meningioma, glioma, pituitary*, dan non-tumor otak). Data *Training* 90%, *Testing* 10%	• *Preprocessing Data*: *resize* • Augmentasi Data: tidak ada • Model: DenseNet169 dan VGG16	DenseNet169: • Akurasi: 98% • *Precision*: 0.97 • *Recall*: 0.98 • *F1-Score*: 0.98 VGG16: • Akurasi: 75%

No.	Penelitian	Data	Metode	Hasil
2.	*Clinical Application of Vision Transformers for Melanoma Classification: A Multi-Dataset Evaluation Study* (Garcia et al., 2025)	Data *Training*: • *Nevus*: 12.875 • *Melanoma*: 4.522 Data *Testing*: • *Nevus*: 2.495 • *Melanoma*: 1.327	• *Preprocessing Data*: normalisasi, *resize*, dan *padding* dengan piksel hitam • Augmentasi Data: *random resized crop* (skala 0.7–1.0), *random horizontal flipping*, RandAugment, dan StyleGAN2-ADA • Model: Vision Transformer (ViT)	ROC-AUC: • Tanpa augmentasi GAN: 0.902 • Dengan augmentasi GAN: 0.915 • Setelah *fine-tuning*: 0.926
3.	*Transforming Alzheimer's Disease Diagnosis: Implementing Vision Transformer (ViT) for MRI Images Classification* (Kurniasari et al., 2025a)	Setiap kelas terdiri dari 2.000 citra MRI: • *Non-Dementia*: 2.000 • *Moderate*: 2.000 • *Mild*: 2.000 • *Very Mild Dementia*: 2.000	• *Preprocessing Data*: normalisasi dan *resize* • Augmentasi Data: *resize, cropping, flipping, rotation, color adjustment*, dan *grayscale* • Model: Vision Transformer (ViT)	Akurasi: 98.19% *Sensitivity*: 96.34% *Specificity*: 98.80% *F1-score*: 96.37%

No.	Penelitian	Data	Metode	Hasil
4.	Implementasi Algoritma *Convolutional Neural Network* Untuk Klasifikasi dan Deteksi Tumor Otak (Sanhaji et.al., 2025)	Data *Training*: • *Glioma*: 1321 gambar • *Meningioma*: 1339 gambar • *No Tumor*: 1595 gambar • *Pituitary*: 1457 gambar Data *Testing*: • *Glioma*: 300 gambar • *Meningioma*: 306 gambar • *No Tumor*: 405 gambar • *Pituitary*: 300 gambar	• *Preprocessing Data*: *resize*, konversi format, dan *one-hot encoding* • Model: CNN	• Akurasi: 88% • *Sensitivity*: 88% • *Specificity*: 88% • *F1-score*: 88%
5.	Klasifikasi Citra MRI Tumor Otak Menggunakan Metode *Convolutional Neural Network* (Husen, 2024)	Berasal dari Kaggle sebanyak 610 data (*glioma, meningioma, no tumor*, dan *pituitary*). Data *training* 80%, data *testing* 10%, dan data *validation* 10%	• *Preprocessing Data*: *resize* • Augmentasi Data: *flipping, scaling, random brightness*, dan *rotation* • Model: CNN	Akurasi: 92.97% Presisi: 93.5% *Sensitivity*: 92.97% *F1-score*: 92.97%

No.	Penelitian	Data	Metode	Hasil
6.	Klasifikasi Tumor Otak Menggunakan *Convolutional Neural Network* dengan Arsitektur Efficientnet-B3 (Andre et.al., 2021)	Terdiri dari 4091 data yang dibagi menjadi 3 kelas dan berasal dari Kaggle. • *Glioma*: 394 data • *Meningioma*: 827 data • *Pituitary*: 2870 data.	• Model: Efficientnet-B3 • *Learning rate* 0.02 dan *dense layer* 256 neuron	Akurasi: 99.7% *F1-score*: 99.6%
7.	*Leveraging Transfer Learning Using Convolutional Neural Networks to Enhance Brain Tumor Images Classification* (Kurniasari et.al., 2025b)	Data terdiri dari 7020 data yang berasal dari Kaggle. • *Glioma*: 1620 gambar • *Meningioma*: 1644 gambar • *No Tumor*: 2000 gambar • *Pituitary*: 1756 gambar	• *Preprocessing Data*: *Grayscale, Gaussian blur, binary thresholding, erotion, dilation, contour extraction, cropping, bone color map*, dan *resize* • Augmentasi Data: rotasi, translasi, *shear*, *horizontal flip*, dan *fill mode* • Model terbaik: ResNet50V2	Akurasi: 97.7% Presisi: 97.8% *Sensitivity*: 97.8% *Specificity*: 99.3%

2.1.1 Penelitian Pertama (Pratama et al., 2024)

Pratama et al. (2024) melakukan penelitian untuk membandingkan dan mengevaluasi kinerja delapan model arsitektur jaringan *pre-built* dalam mengklasifikasikan

tumor otak melalui citra *MRI*, dengan memanfaatkan metodologi *transfer learning*. Analisis tumor otak ini dianggap sebagai subjek penting dalam kedokteran karena deteksi yang cepat dan akurat dapat menghasilkan perawatan yang lebih baik. Studi ini berupaya memberikan wawasan tentang cara kerja masing-masing arsitektur dengan menggunakan *confusion matrix* dan pengukuran performa seperti *recall*, akurasi, *precision*, dan *F1-score*.

Data yang digunakan dalam penelitian ini adalah kumpulan citra *Magnetic Resonance Imaging (MRI)* sebanyak 3.264 citra. Dataset citra ini diklasifikasikan ke dalam empat kategori: *meningioma*, *glioma*, *pituitary*, dan yang tidak menderita tumor otak (*No Tumor*). Metode penelitian ini berpusat pada *Transfer Learning*, yang memungkinkan pemanfaatan kekuatan *Convolutional Neural Network (CNN)* tanpa harus melatih model dari awal, yang mana membutuhkan banyak data dan sumber daya komputasi. Dalam tahapan pra-pemrosesan, dataset yang telah diambil selanjutnya dilabeli, dan kemudian di-*resize* ukurannya menjadi 150×150 piksel. Setelah pra-pemrosesan, data dibagi menjadi data pelatihan (*training*) sebesar 90% dan data pengujian (*testing*) sebesar 10% dari keseluruhan dataset.

Dua arsitektur *pre-built* utama yang diteliti secara mendalam dan dibandingkan kinerjanya adalah *VGG16* dan *DenseNet169*. Setelah model *pre-built* diunduh, parameter *include_top* diatur menjadi *False* untuk menghilangkan lapisan output aslinya, dan ditambahkan lapisan output baru yang sesuai dengan klasifikasi empat kelas tumor otak. Hasil eksperimen menunjukkan bahwa pendekatan *transfer learning* berhasil mengklasifikasikan jenis tumor otak dengan memuaskan, bahkan dengan dataset yang terbatas. Namun, *DenseNet169* terbukti unggul secara signifikan dibandingkan *VGG16* dalam semua metrik evaluasi. *DenseNet169* mencapai akurasi 98%, sementara *VGG16* hanya mencapai 75%. Selain itu, *DenseNet169* memiliki nilai *Precision* 0.97, *Recall* 0.98, dan *F1-Score* 0.98, yang secara konsisten lebih tinggi dibandingkan *VGG16*.

2.1.2 Penelitian Kedua (Garcia et al., 2025)

Penelitian oleh Garcia et al. (2025) bertujuan untuk mengevaluasi penerapan klinis *Vision Transformers* (ViT) untuk klasifikasi melanoma menggunakan gambar dermosko- pik. Inti dari penelitian ini adalah untuk mengatasi kesulitan diagnosis

melanoma yang sering kali terlihat mirip dengan tahi lalat jinak (*nevus*) dengan memanfaatkan model kecerdasan buatan (AI). Para peneliti menggunakan model ViT-L/16, yang telah dilatih sebelumnya, dan melakukan *fine-tuning* pada dataset gambar lesi kulit. Salah satu metode inovatif yang digunakan adalah augmentasi data menggunakan gambar sintetis yang dihasilkan oleh algoritma *Generative Adversarial Networks* (GANs) untuk meningkatkan kapasitas generalisasi model. Kinerja ViT kemudian dibandingkan dengan model CNN *Deep Learning* lainnya dan sistem diagnostik komersial, MoleAnalyzer Pro, untuk mendukung pengambilan keputusan klinis yang lebih baik.

Dataset utama yang digunakan untuk melatih dan menyetel halus model adalah ISIC 2019 *Challenge dataset*, yang merupakan koleksi besar gambar dermoskopik lesi kulit dengan anotasi ahli. Set pelatihan asli ISIC 2019 mencakup total lebih dari 25000 gambar, dengan rincian 12875 gambar *nevus* dan 4522 gambar melanoma. Untuk menguji kemampuan generalisasi model pada data klinis yang belum pernah dilihat, digunakan dataset eksternal MN187 sebagai set uji, yang terdiri dari 187 gambar dermoskopik yang telah dikonfirmasi biopsi (96 *nevus* jinak dan 91 melanoma ganas). Semua gambar diubah ukurannya dan di-*padding* dengan piksel hitam untuk mencapai ukuran input yang konsisten yaitu 224×224 piksel pada tahap prapemrosesan. Standarisasi ini diterapkan pada setiap model untuk memungkinkan perbandingan kinerja yang adil, meskipun model ViT secara arsitektur dapat menanga- ni ukuran input yang bervariasi.

Metode penelitian ini berpusat pada ViT, khususnya arsitektur ViT-L/16, yang merupakan model besar yang telah dilatih sebelumnya pada dataset ImageNet-1k. Model ViT memproses gambar dengan membaginya menjadi *patch* berukuran tetap, memproyeksikannya secara linear, dan memprosesnya melalui 24 lapisan *transformer encoder*. Kemampuan utama ViT adalah mekanisme *multi-head self-attention* yang memungkinkannya menangkap dependensi jarak jauh dan hubungan kontekstual global antara wilayah gambar, sebuah keuntungan dibandingkan CNN yang terbatas pada fitur lokal. Selain ViT-L/16, model CNN lainnya seperti ResNet- 152, DenseNet- 201, EfficientNet-B7, ConvNeXt-XL, dan ViT-B/16 yang lebih kecil juga dievaluasi sebagai pembanding. Semua model *Deep Learning* ini berbagi arsitektur *classification head* yang sama, yang terdiri dari empat lapisan *fully connected* dengan aktivasi ReLU.

ViT-L/16 yang digunakan memiliki total sekitar 307 juta parameter pada *backbone* dan secara arsitektur memiliki 16 *attention heads* dan *hidden size* sebesar 1024 per lapisan *transformer*. Untuk konfigurasi pelatihan, fungsi kerugian yang digunakan adalah `BCEWithLogitsLoss` untuk klasifikasi biner. *Optimizer* yang digunakan adalah AdamW dengan *learning rate* sebesar 2×10^{-4} dan *weight decay* sebesar 0.05. Sebuah *cosine annealing learning rate scheduler* dengan parameter $T_{\text{max}} = 30$ digunakan untuk secara progresif mengurangi *learning rate* seiring dengan bertambahnya *epoch*. Augmentasi data tambahan dilakukan dengan menghasilkan 50000 gambar sintetis menggunakan StyleGAN2-ADA, kemudian 46231 gambar sintetis dipilih dan ditambahkan ke set pelatihan untuk meningkatkan kinerja.

Hasil penelitian menunjukkan bahwa model ViT-L/16 memberikan kinerja yang superior dalam klasifikasi melanoma. Dalam evaluasi awal tanpa augmentasi data GAN, ViT-L/16 mencapai skor ROC-AUC *baseline* sebesar 0.902 pada set uji MN187, melampaui semua model CNN lainnya dan sistem komersial MoleAnalyzer Pro (AUC 0.856). Namun, perbedaan kinerja awal ini tidak signifikan secara statistik ($p = 0.07$) dibandingkan dengan MoleAnalyzer Pro. Setelah menambahkan 46231 gambar sintetis yang telah disaring ke data pelatihan, ROC-AUC ViT-L/16 meningkat menjadi 0.915 pada set uji eksternal MN187. Peningkatan kinerja yang didukung augmentasi data ini secara statistik signifikan terhadap MoleAnalyzer Pro ($p = 0.032$). Selain itu, versi ViT-L/16 yang lebih halus, yang didapatkan dengan melatih ulang enam lapisan *transformer* terakhir, mencapai ROC-AUC tertinggi 0.926 pada MN187, yang terbukti signifikan secara statistik ($p = 0.006$) dibandingkan MoleAnalyzer Pro.

2.1.3 Penelitian Ketiga (Kurniasari et al., 2025a)

Penelitian Kurniasari et al. (2025) memiliki tujuan untuk menilai efikasi arsitektur *Vision Transformer* (ViT) dalam klasifikasi multikelas citra MRI untuk diagnosis penyakit Alzheimer. Penyakit Alzheimer sendiri merupakan kelainan neurologis progresif yang bertanggung jawab atas 60-80% kasus demensia, sehingga deteksi dini sangat penting. Penelitian ini secara spesifik berfokus pada peningkatan akurasi dan keandalan model melalui penggunaan teknik augmentasi data selama fase pra-pemrosesan. Dengan demikian, studi ini bertujuan untuk meminimalkan kesalahan diagnosis dan mendukung intervensi klinis yang tepat waktu.

Data yang digunakan dalam penelitian ini terdiri dari 8.000 citra MRI otak manusia yang bersumber secara daring dari *Alzheimer's Disease Neuroimaging Initiative* (ADNI) dan diakses melalui platform Kaggle. Data tersebut dikategorikan ke dalam empat klasifikasi berbeda: *non-demented* (tidak demensia), *very mild demented* (demensia sangat ringan), *mild demented* (demensia ringan), dan *moderately demented* (demensia sedang). Penting untuk dicatat bahwa setiap kelas berisi 2.000 citra, yang memastikan distribusi yang seimbang di seluruh kategori yang diteliti. Metode yang diusulkan melibatkan pendekatan pembelajaran mendalam (*deep learning*) menggunakan model ViT untuk tugas klasifikasi citra. Dalam tahap pra-pemrosesan, citra dengan resolusi asli 200 × 190 piksel disesuaikan ukurannya menjadi 224 × 224 piksel agar sesuai dengan dimensi input model ViT yang telah dilatih sebelumnya.

Metodologi penelitian melibatkan fine-tuning model ViT yang telah dilatih sebelumnya. Data dibagi menjadi 90% (7.200 sampel) untuk pengembangan model (pelatihan dan validasi) dan 10% (800 sampel) untuk pengujian akhir. Pra-pemrosesan juga mencakup teknik augmentasi data seperti *resizing, cropping, flipping, rotation*, dan *color adjustment*, yang bertujuan untuk meningkatkan keberagaman dataset. Peneliti menguji berbagai kombinasi learning rates (0.01, 0.001, 0.0001) dan batch sizes (32, 64) selama 50 epoch untuk mencapai kinerja optimal.

Berdasarkan analisis hasil, kombinasi parameter yang paling efektif ditemukan adalah learning rate 0.0001 dan batch size 64. Konfigurasi optimal ini menghasilkan metrik evaluasi yang sangat kuat, termasuk akurasi sebesar 98.19%, sensitivitas 96.34%, spesifisitas 98.80%, dan F1-Score 96.37%. Kinerja model ini menunjukkan bahwa kategori *Moderate Demented* mencapai performa sempurna dengan semua metrik (akurasi, sensitivitas, spesifisitas, dan F1-score) bernilai 100%. Meskipun demikian, penelitian ini mencatat bahwa masih ada tantangan, terutama dalam membedakan fitur antara kasus *Non-Demented* dan *Very Mild Demented*, yang sering tumpang tindih dan dapat menyebabkan kesalahan klasifikasi.

2.1.4 Penelitian Keempat (Sanhaji et al., 2025)

Penelitian yang dilakukan oleh Sanhaji et al. mengimplementasikan algoritma *Convolutional Neural Network* (CNN) untuk mendeteksi keberadaan tumor otak melalui analisis citra digital. Pemilihan metode ini didasarkan pada keunggulan arsitektur jaringan saraf dalam mengekstraksi fitur kompleks secara otomatis tanpa intervensi manual. Peneliti mengadopsi teknik *transfer learning* guna mengoptimalkan efisiensi waktu pelatihan serta meningkatkan ketepatan prediksi pada dataset yang tersedia. Arsitektur model ini dirancang khusus untuk mengenali pola spasial pada citra medis yang sulit diidentifikasi oleh mata manusia secara konvensional.

Dataset utama dalam studi ini bersumber dari citra *Magnetic Resonance Imaging* (MRI) yang mencakup empat kategori klasifikasi berbeda. Kelompok data tersebut terdiri atas glioma, meningioma, tumor pituitari, serta kondisi non-tumor. Secara keseluruhan, terdapat 7.023 sampel data yang dialokasikan ke dalam beberapa kelompok besar untuk kebutuhan eksperimen. Data pelatihan berjumlah 5.712 citra, sedangkan data pengujian berjumlah 1.311 citra guna memvalidasi performa model.

Proses pembangunan sistem diawali dengan tahap pra-pemrosesan yang melibatkan penyesuaian dimensi seluruh gambar menjadi 256x256 piksel agar seragam,. Label kategori kemudian dikonversi menggunakan format *one-hot encoding* guna memudahkan pengenalan numerik oleh model *deep learning*. Arsitektur model disusun melalui beberapa lapisan konvolusi (*Conv2D*), penyusutan spasial (*MaxPooling2D*), serta mekanisme *dropout* untuk meminimalkan risiko *overfitting*. Pelatihan dilakukan selama sepuluh iterasi penuh atau *epoch* yang memakan waktu total sekitar 733 detik.

Evaluasi terhadap model akhir menunjukkan tingkat akurasi keseluruhan sebesar 88% dalam mengklasifikasikan jenis lesi otak. Analisis kinerja dilakukan secara mendalam menggunakan instrumen *confusion matrix* serta metrik presisi, *recall*, dan *F1-score*. Hasil pengujian menunjukkan bahwa kategori non-tumor memperoleh nilai *F1-score* tertinggi yakni 0.94. Di sisi lain, kategori meningioma menunjukkan performa yang sedikit lebih rendah karena terdapat sampel yang tidak dikenali dengan benar oleh sistem. Secara umum, model ini dianggap stabil dan mampu melakukan generalisasi pada data baru tanpa menunjukkan tanda *overfitting*.

2.1.5 Penelitian Kelima (Husen, 2024)

Penelitian yang dilakukan oleh Husen (2024) berfokus pada peningkatan efisiensi diagnosis medis melalui klasifikasi otomatis lesi intrakranial pada citra *Magnetic Resonance Imaging* (MRI). Arsitektur yang diimplementasikan dalam studi ini adalah *Convolutional Neural Network* (CNN) dengan desain kategori klasik guna mengevaluasi pengaruh struktur data terhadap performa sistem. Pemilihan model tersebut bertujuan menyajikan alat bantu diagnostik yang presisi bagi tenaga medis dalam mendeteksi pertumbuhan sel abnormal di area otak.

Dataset penelitian menggunakan repositori publik Kaggle yang terdiri dari empat kategori utama, yaitu tumor jenis glioma, meningioma, pituitari, dan kelompok tanpa tumor. Koleksi data tersebut mencakup citra medis asli dengan resolusi tinggi sebesar 512x512 piksel sebelum memasuki tahap pemrosesan teknis. Keberadaan empat label kelas ini memungkinkan model melakukan klasifikasi multilabel yang komprehensif terhadap berbagai variasi neoplasma otak.

Pengembangan sistem dimulai dengan tahap pra-pemrosesan yang meliputi penyesuaian dimensi gambar menjadi 256x256 piksel untuk mengoptimalkan kecepatan komputasi. Peneliti menerapkan beragam teknik augmentasi data, mencakup *flipping*, *scaling*, serta rotasi acak, demi memperkaya variasi sampel tanpa mengubah label asli. Optimasi dilakukan melalui uji coba parameter *epoch* dalam rentang 5 sampai 20 siklus pelatihan guna mencapai titik konvergensi yang paling stabil.

Hasil eksperimen menunjukkan bahwa integrasi teknik augmentasi meningkatkan kemampuan generalisasi model secara signifikan. Skema *Vertical & Horizontal Flip* pada 20 *epoch* berhasil mencatatkan performa maksimal dengan nilai akurasi, presisi, serta *F1-score* menyentuh angka 93%. Temuan ini menegaskan bahwa pemilihan ukuran citra yang sesuai beserta pengayaan data sintetis menjadi faktor penentu dalam menghasilkan sistem deteksi tumor otak yang andal.

2.1.6 Penelitian Keenam (Andre et al., 2021)

Penelitian yang dipaparkan oleh Andre et al. (2021) berfokus pada pengembangan sistem otomatisasi untuk identifikasi penyakit kritis pada organ vital manusia melalui pemanfaatan teknologi kecerdasan buatan. Tujuan utama dari studi ini adalah membangun sebuah model klasifikasi citra tumor otak yang mampu meminimalkan risiko kesalahan diagnosis manual serta mempercepat proses pengambilan keputusan medis dibandingkan metode biopsi konvensional. Melalui pendekatan *computer vision*, peneliti berusaha menghadirkan solusi teknologi yang dapat membedakan jaringan lunak abnormal secara objektif dan efisien guna membantu tenaga medis dalam menangani peningkatan kasus tumor otak.

Arsitektur model yang diimplementasikan dalam riset ini adalah *Convolutional Neural Network* (CNN) dengan spesifikasi *EfficientNet-B3*. Pemilihan arsitektur tersebut didasarkan pada keunggulannya dalam menyeimbangkan dimensi lebar, kedalaman, dan resolusi jaringan sehingga mampu menghasilkan akurasi tinggi dengan parameter operasi yang tetap efisien. Model ini bekerja dengan melakukan ekstraksi fitur secara otomatis dari citra dua dimensi tanpa memerlukan proses pembelajaran fitur manual seperti pada algoritma *machine learning* tradisional.

Data yang menjadi landasan pembelajaran model berasal dari repositori Kaggle yang terdiri dari 4.091 gambar *Magnetic Resonance Imaging* (MRI) dalam format *grayscale*. Kumpulan data tersebut mencakup berbagai kategori medis, antara lain jenis tumor glioma, meningioma, dan pituitari. Penggunaan citra MRI dipilih karena kemampuannya dalam menyajikan visualisasi jaringan lunak otak yang jauh lebih detail dan jelas dibandingkan dengan pemindaian CT-Scan.

Proses pembangunan model dijalankan melalui serangkaian tahapan sistematis yang mengadopsi metode *Waterfall*, mulai dari analisis kebutuhan perangkat lunak hingga desain sistem menggunakan *Data Flow Diagram* (DFD). Peneliti melakukan *hyper-parameter optimization* dengan menguji berbagai skenario *learning rate* serta variasi jumlah neuron pada *dense layer* untuk menemukan konfigurasi yang paling optimal. Hasil akhir pengujian menunjukkan performa yang sangat impresif, di mana kombinasi *learning rate* 0,02 dan 256 neuron menghasilkan akurasi mencapai 99,7% dengan nilai *F1-Score* sebesar 99,6%. Kesuksesan eksperimen ini kemudian diintegrasikan ke dalam sebuah aplikasi berbasis web menggunakan

bahasa pemrograman Python dan *framework* Flask sebagai antarmuka pengguna.

2.1.7 Penelitian Ketujuh (Kurniasari et al., 2025b)

Penelitian yang dilakukan oleh Kurniasari et al. (2025b) bertujuan untuk meningkatkan akurasi klasifikasi citra tumor otak dengan memanfaatkan pendekatan *transfer learning* pada arsitektur *Convolutional Neural Network*. Latar belakang penelitian ini didasari oleh tantangan dalam diagnosis tumor otak yang memerlukan ketepatan tinggi, sementara metode konvensional rentan terhadap kesalahan interpretasi. *Transfer learning* dipilih sebagai strategi utama untuk mengatasi keterbatasan data pelatihan yang besar dengan memanfaatkan pengetahuan yang telah diperoleh model dari dataset ImageNet. Penelitian ini menggunakan lima arsitektur *pre-trained*, yaitu DenseNet121, InceptionResNetV2, MobileNetV2, NasNetMobile, dan ResNet50V2, yang kemudian dibandingkan performanya dalam mengklasifikasikan empat kategori citra tumor otak.

Data yang digunakan dalam penelitian ini bersumber dari Kaggle dengan total 7.020 citra MRI yang terbagi menjadi empat kelas, yakni glioma, meningioma, *pituitary tumor*, dan tanpa tumor. Setiap citra memiliki resolusi 512×512 piksel dengan kedalaman warna 24-bit dalam format JPG. Sebelum memasuki tahap pelatihan, data melalui serangkaian proses *pre-processing* yang meliputi konversi *grayscale*, penerapan *Gaussian blur* untuk mengurangi *noise*, *binary thresholding* berbasis Otsu, operasi morfologi erosi dan dilasi, ekstraksi kontur, pemotongan citra, serta *resizing*. Proses augmentasi data juga diterapkan menggunakan *Image Data Generator* dengan parameter seperti rotasi, pergeseran lebar dan tinggi, *shear*, serta *flipping horizontal* untuk memperkaya variasi data pelatihan.

Pembangunan model dilakukan melalui dua skenario utama, yaitu pengujian tanpa *transfer learning* menggunakan arsitektur CNN yang dirancang sendiri dan pengujian dengan *transfer learning* pada kelima model *pre-trained*. Tahap *hyperparameter tuning* dilakukan dengan metode *grid search* untuk menentukan kombinasi optimal antara *dropout rate* dan *dense units*. Hasil *tuning* menunjukkan bahwa ResNet50V2 memiliki konfigurasi terbaik dengan *dropout rate* 0,15 dan *dense units* 64. Pelatihan model dilakukan dengan *learning rate* 0,0001, *batch size* 128, dan fungsi *loss categorical cross-entropy* selama 15 *epoch*. Evaluasi performa

model menggunakan metrik akurasi, presisi, sensitivitas, dan spesifisitas yang dihitung berdasarkan *confusion matrix*.

Hasil penelitian menunjukkan bahwa model ResNet50V2 unggul dibandingkan keempat model lainnya dengan akurasi rata-rata 97.70 persen, presisi 97.80 persen, sensitivitas 97.70 persen, dan spesifisitas 99.30 persen. Model ini juga menunjukkan kemampuan generalisasi yang baik dengan kurva *loss* yang konvergen antara data *training* dan *validation* tanpa indikasi *underfitting* maupun *overfitting* yang signifikan. Performa ResNet50V2 melampaui beberapa penelitian sebelumnya dalam konteks klasifikasi multi-kelas, meskipun tetap kompetitif dengan studi yang menggunakan pendekatan biner. Kontribusi utama penelitian ini terletak pada demonstrasi bahwa *transfer learning* pada arsitektur CNN, khususnya ResNet50V2, dapat menghasilkan performa yang andal dalam klasifikasi tumor otak.

2.2 Tumor Otak

Tumor otak adalah istilah yang merujuk pada sekelompok besar neoplasma yang berasal dari jaringan intrakranial, termasuk neuron, sel glia, astrosit, dan meningen atau selaputnya. Secara definisi, tumor otak merupakan pertumbuhan sel yang abnormal pada jaringan otak atau di dalam dan di sekitar otak secara tidak wajar dan tidak terkendali (Essianda et al., 2023). Kondisi ini juga sering disebut sebagai neoplasma atau proses desak ruang (*Space Occupying Lesion* atau SOL) di dalam rongga tengkorak (Kasiatun & Sukraeny, 2024).

Dalam konteks klasifikasi citra medis, istilah SOL atau neoplasma ini menjadi target utama identifikasi melalui modalitas *Magnetic Resonance Imaging* (MRI) (Yunisafitri, 2025). MRI dipilih karena kemampuannya dalam memberikan informasi citra yang sangat detail, terutama dalam membedakan antara jaringan lunak yang sehat dengan jaringan keras atau lesi tumor yang terdapat pada otak (Andre et al., 2021). Hal ini menjadi dasar penting bagi model *Vision Transformer* (ViT) untuk mengenali pola visual yang kompleks dari massa tumor tersebut (Jraba et al., 2023).

Secara klinis, tumor otak dapat dibagi berdasarkan sifat keganasannya menjadi jinak (*benign*) dan ganas (*malignant*) (Kepakisan et al., 2022). Tumor ganas biasanya

disebut kanker dan memiliki progresivitas tinggi, serta dapat menyerang atau menekan jaringan di sekitarnya dan mengancam nyawa. Contoh tumor ganas meliputi *Glioblastoma Multiforme* (GBM), jenis kanker otak ini merupakan kanker otak paling agresif dan paling umum, serta glioma. Sementara itu, tumor jinak tumbuh relatif lambat dan jarang bermetastasis, namun tetap memerlukan penanganan karena dapat memengaruhi fungsi otak. Contoh tumor jinak yang paling sering terdiagnosis secara histologi adalah meningioma dan adenoma pituitari (Essianda et al., 2023).

Klasifikasi klinis antara sifat jinak dan ganas ini sering kali diimplementasikan ke dalam kategori kelas (*labels*) dalam penelitian *deep learning*, seperti kategori glioma, meningioma, *no tumor* dan *pituitary* (Sanhaji et al., 2025). Dalam tugas klasifikasi multikelas, model akan dilatih untuk mengenali ciri khas histopatologis dari masing-masing jenis tersebut, misalnya Meningioma yang berasal dari sel selaput otak (meninges) atau adenoma pituitari yang tumbuh pada kelenjar pituitari (Astuti et al., 2023). Perbedaan karakteristik morfologi, warna, dan tekstur pada citra MRI inilah yang akan dianalisis kinerjanya oleh arsitektur *Vision Transformer* melalui mekanisme *self-attention* untuk menangkap hubungan spasial global antar bagian citra (Sanhaji et al., 2025).

Berdasarkan asal mulanya, tumor otak diklasifikasikan menjadi tumor otak primer dan sekunder (metastasis). Tumor otak primer adalah tumor yang pertama kali tumbuh dan berkembang di sel otak itu sendiri. Meskipun tumor primer dapat menyebar ke bagian lain dari sistem saraf, jarang sekali menyebar ke organ lain di luar otak dan sumsum tulang belakang (Sumaiyah et al., 2024). Sebaliknya, tumor otak sekunder adalah tumor yang menyebar (metastasis) ke otak dari kanker yang berasal dari bagian tubuh lain, seperti kanker paru, payudara, melanoma, ginjal, dan usus (Andre et al., 2021).

Klasifikasi antara kategori Glioma, Meningioma, Pituitary, dan kondisi No Tumor merupakan tantangan krusial dalam diagnosis otomatis karena jenis-jenis tumor tersebut sering kali menunjukkan pola radiologis atau karakteristik morfologi yang serupa pada citra MRI (Sanhaji et al., 2025). Penggunaan model Vision Transformer (ViT) bertujuan untuk meminimalkan risiko kesalahan diagnosis tersebut dengan memanfaatkan mekanisme self-attention untuk mengekstraksi fitur penting secara otomatis dan menangkap hubungan spasial global tanpa memerlukan rekayasa fitur manual (Sanhaji et al., 2025). Melalui proses fine-tuning pada dataset citra MRI

yang telah dilabeli secara akurat ke dalam empat kategori tersebut, sistem klasifikasi ini diharapkan mampu memberikan dukungan keputusan klinis yang lebih cepat dan objektif bagi tenaga medis dalam mendeteksi keberadaan serta membedakan jenis tumor otak secara spesifik (Jraba et al., 2023).

2.3 Pencitraan *Magnetic Resonance Imaging* (MRI)

Magnetic Resonance Imaging (MRI) merupakan sebuah perangkat mutakhir dalam ranah medis yang menggabungkan teknologi komputer, medan magnet kuat, dan gelombang radio untuk mengemulasi representasi potongan tubuh manusia dan menghasilkan gambaran yang lebih terperinci dan jelas. Pemeriksaan ini memiliki keunggulan dibandingkan modalitas pencitraan diagnostik lainnya karena menggunakan radiasi non-ionisasi dan menghasilkan resolusi tinggi pada jaringan lunak. Karena keandalannya, MRI telah menjadi modalitas pilihan utama (*primary modality*) dan standar baku emas (*gold standard*) dalam mendiagnosis tumor otak (Astuti et al., 2023).

Sebagai domain masalah, MRI beroperasi dalam bidang diagnostik medis yang sangat penting, khususnya untuk mendeteksi kelainan intrakranial seperti tumor otak (Yunisafitri, 2025) dan mengevaluasi perubahan struktural otak terkait penyakit Alzheimer (Kurniasari et al., 2025). Meskipun MRI menyediakan gambaran anatomi yang sangat detail, domain ini menghadapi tantangan signifikan yang mendesak intervensi kecerdasan buatan (AI). Pertama, interpretasi citra MRI secara manual oleh tenaga medis menuntut keahlian yang sangat tinggi dan bersifat subjektif, sehingga rentan terhadap kesalahan yang diperburuk oleh faktor manusia seperti kelelahan atau keterbatasan pengalaman. Kedua, metode diagnostik konvensional ini dibatasi oleh biaya tinggi, kebutuhan akan operator dan fasilitas terlatih, serta waktu yang relatif lama untuk mendapatkan interpretasi hasil yang akurat (Sanhaji et al., 2025).

2.4 *Artificial Intelligence* (AI)

Artificial Intelligence (AI) merupakan domain ilmu komputer yang membekali mesin dengan kemampuan untuk melakukan tugas-tugas tertentu melalui proses

pembelajaran berbasis pengalaman yang diperoleh dari pemrosesan data. Kapasitas pembelajaran tersebut sangat ditentukan oleh kuantitas dan kualitas data yang diolah melalui perangkat keras maupun perangkat lunak pendukung. Secara komprehensif, AI berfungsi sebagai bidang induk yang memayungi berbagai teknik komputasi modern, termasuk di dalamnya adalah *Machine Learning* dan jaringan saraf tiruan (Diponegoro et al., 2021).

Dalam struktur taksonominya, AI merupakan payung besar bagi *Machine Learning*, yang kemudian memiliki sub-cabang lebih spesifik seperti *Deep Learning* dan *Transfer Learning*. Teknologi ini memiliki kapabilitas untuk menganalisis pola yang kompleks, memahami konteks, serta melakukan penyesuaian proses secara otomatis berdasarkan kebutuhan data yang diproses. Implementasi AI saat ini telah merambah ke berbagai sektor strategis, termasuk pengembangan sistem pembelajaran adaptif dan tutor virtual yang terbukti mampu meningkatkan retensi serta efektivitas pendidikan (Turmuzi, 2025).

2.4.1 *Machine Learning*

Machine Learning (ML) adalah cabang dari Kecerdasan Buatan (*Artificial Intelligence*, AI) yang berfokus pada seni pemrograman komputer sehingga ia dapat belajar dari data (Fatmah et al., 2023). Menurut Arthur Samuel (1959, seperti yang dikutip oleh (Andre et al., 2021)), ML merupakan pembelajaran yang memberikan perintah kepada komputer untuk belajar dari data tanpa harus mengikuti instruksi yang diprogram secara terperinci. Tujuannya adalah melatih mesin untuk mengenali pola-pola rumit atau hubungan yang kompleks dalam data dan membuat prediksi atau keputusan berdasarkan data dan algoritma yang diberikan (Dzakwan et al., 2025).

ML bersifat dinamis, artinya memiliki kemampuan untuk mengubah dirinya sendiri ketika diterapkan pada data yang lebih banyak. Dalam ML tradisional, seperti algoritma *K-Nearest Neighbors* (K-NN), *Support Vector Machine* (SVM), dan *Random Forest*, seringkali memerlukan ekstraksi fitur khusus dalam proses *feature learning* untuk memproses data gambar (Harianto et al., 2025). Namun, seiring berkembangnya data berskala besar dan kompleks, ML tradisional sering menghadapi keterbatasan kinerja, khususnya pada data berdimensi tinggi (Ujianto et

al., 2025).

2.4.2 *Deep Learning*

Deep Learning (DL) adalah sub-bidang dari Machine Learning yang menggunakan arsitektur jaringan saraf tiruan (*Artificial Neural Networks*, ANN) yang dicirikan oleh penggunaan lapisan ganda yang lebih dalam dan kompleks (*multi-layer neural networks*) (Siregar & Husein, 2024). DL, yang merupakan implementasi dari Neural Network, mengatasi kelemahan ML tradisional dengan kemampuannya untuk secara otomatis mengekstraksi fitur yang kompleks dan representatif dari data (terutama citra) tanpa memerlukan rekayasa fitur manual (Kurniasari et al., 2025).

Arsitektur DL yang populer, seperti *Convolutional Neural Network* (CNN), dirancang khusus untuk memproses data visual dengan meniru cara kerja sistem visual manusia (Fadilah et al., 2023). CNN sendiri diklasifikasikan sebagai *Deep Neural Network* karena kedalaman jaringannya yang tinggi dan banyak diterapkan pada data citra (Auliaddina & Arifin, 2024). Meskipun model DL seperti CNN unggul dalam akurasi dan efisiensi analisis citra, teknologi ini umumnya membutuhkan jumlah data berlabel yang sangat besar (*data hungry*) untuk pelatihan optimal agar menghindari *overfitting* (Pratama et al., 2024).

2.4.3 *Transfer Learning*

Transfer Learning adalah sebuah metode pada *Machine Learning* dan *Deep Learning* yang dirancang untuk mengatasi masalah keterbatasan data pelatihan yang dihadapi oleh model DL yang kompleks. Metode ini didefinisikan sebagai teknik di mana model diberikan masalah baru tetapi menerapkan pengalaman dari pelatihan sebelumnya yang telah dilakukan pada tugas terkait (Pradana et al., 2024).

Secara lebih spesifik, *transfer learning* memanfaatkan pengetahuan (bobot dan bias) yang telah diperoleh dari model pra-terlatih (*pre-trained model*) yang sebelumnya dilatih pada dataset yang sangat besar (seperti ImageNet, yang memiliki lebih dari satu juta citra dari 1.000 kelas) (Song et al., 2024). Pengetahuan yang telah dipelajari pada domain asal (*source domain*) kemudian digunakan untuk meningkatkan kualitas dan performa model pada domain target (*target domain*) yang seringkali memiliki

dataset lebih kecil atau terbatas (Mufidatuzzainiya & Faisal, 2025).

Transfer learning terbukti dapat meningkatkan efisiensi pelatihan, mempercepat proses, mengurangi kebutuhan sumber daya komputasi yang besar, serta meningkatkan akurasi dan generalisasi model DL pada tugas-tugas spesifik seperti klasifikasi citra medis tumor otak (Sanhaji et al., 2025) atau identifikasi tanaman herbal (Sanjaya et al., 2025).

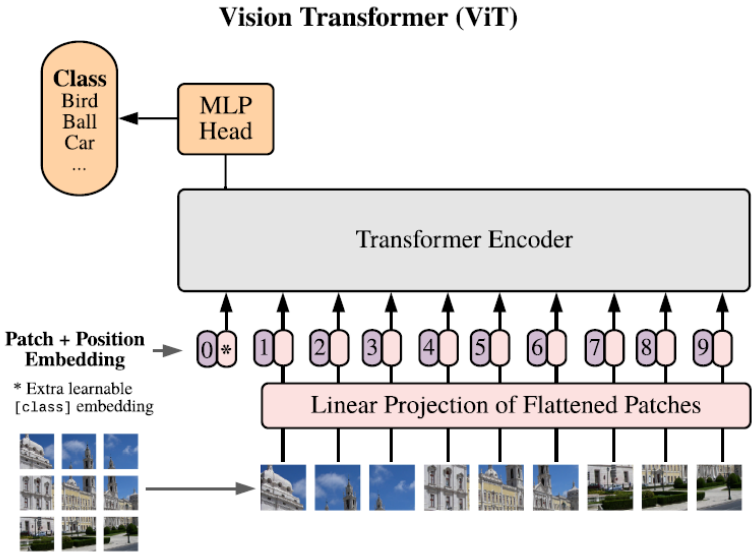
2.5 *Vision Transformer* (ViT)

Vision Transformer (ViT) adalah arsitektur *deep learning* inovatif yang telah terbukti sangat efektif dalam klasifikasi citra. ViT pertama kali diperkenalkan pada tahun 2021 oleh Dosovitskiy et al., mengadopsi arsitektur *transformer* yang awalnya dirancang untuk tugas Pemrosesan Bahasa Alami (*Natural Language Processing, NLP*). Konsep dasar ViT memproses gambar dengan membaginya menjadi urutan *patch* kecil yang ukurannya tetap, dan memperlakukan setiap *patch* ini sebagai *token input*, serupa dengan cara teks atau kata diproses dalam NLP (Dosovitskiy et al., 2021).

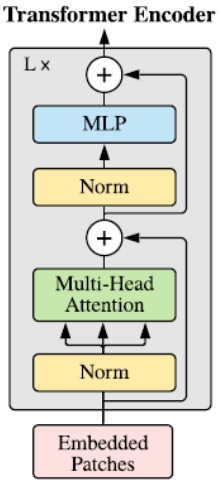
Pendekatan ini memungkinkan ViT untuk memanfaatkan mekanisme *self-attention* yang efisien, yang memungkinkannya menangkap *long-range dependencies* dan hubungan kontekstual global dalam data visual secara lebih efektif dibandingkan *Convolutional Neural Networks (CNN)* tradisional yang cenderung fokus pada bidang reseptif lokal (Rodrigo et al., 2024).

ViT mengadopsi struktur *encoder* dari model *Transformer* asli yang dirancang untuk tugas pemrosesan bahasa. Alur lengkap proses klasifikasi citra dengan ViT dapat dilihat pada Gambar 1, yang menunjukkan bahwa arsitektur ini bermula dari pembagian citra input menjadi *patch* berukuran tetap. *Patch* tersebut kemudian diratakan dan diproyeksikan secara linear (*Linear Projection*) untuk membentuk *embedded patches*. Sebuah token khusus yang dapat dipelajari ditambahkan pada urutan *patch* ini sebagai representasi global untuk klasifikasi, sementara informasi posisi spasial tiap *patch* diinjeksikan melalui *positional embedding*.

Urutan vektor yang telah dilengkapi informasi posisi ini kemudian diproses oleh inti model, yaitu *transformer encoder*. Struktur internal *encoder* ini diuraikan secara lebih rinci pada Gambar 2 yang menunjukkan susunan lapisan identik yang diulang sebanyak L kali. Setiap lapisan *encoder* mengandung dua blok utama: *Multi-Head Self-Attention* dan *Multi-Layer Perceptron* (MLP), yang masing-masing didahului oleh layer normalization (Norm) dan dihubungkan oleh sambungan *residual*. Output dari token khusus pada lapisan *encoder* terakhir kemudian diteruskan ke *MLP Head* untuk menghasilkan prediksi kelas akhir. Berdasarkan bagan ini, pembahasan akan difokuskan pada tiga komponen kunci dalam tahap awal transformasi citra, yaitu pembentukan *embedded patches* atau *embedding encoding*, penambahan informasi lokasi melalui *positional encoding*, dan mekanisme inti *multi-head self-attention*.



Gambar 1 Arsitektur Model ViT (Dosovitskiy et al., 2021).



Gambar 2 *Transformer Encoder* (Dosovitskiy et al., 2021).

2.5.1 *Embedding Encoding*

Embedding encoding atau *patch encoding* merupakan arsitektur pada gambar yang memiliki dimensi tinggi, lebar, dan jumlah saluran dibagi menjadi serangkaian *patch* yang tidak tumpang tindih (Sanhaji et al., 2025). Setiap *patch* berukuran tetap, misalnya 16 x 16 piksel (Kurniasari et al., 2025), dan kemudian diratakan menjadi vektor satu dimensi, yang disebut *flattened patches*.

Vektor-vektor ini selanjutnya diproyeksikan secara linier menggunakan matriks *embedding* yang dapat dilatih ke dimensi yang lebih tinggi untuk menghasilkan *patch embeddings* (Pangestu et al., 2024). *Layer* yang melakukan transformasi ini sering diidentifikasi sebagai *layer conv2d* yang mengubah citra menjadi vektor *embedding* (Sanjaya et al., 2025).

2.5.2 *Positional Encoding*

Selanjutnya, *positional Encoding* ditambahkan ke *patch embeddings* untuk menyediakan informasi mengenai posisi relatif dari setiap *patch* dalam urutan. Tahapan ini sangat penting karena ViT memproses *patch* secara paralel, sehingga tidak memiliki mekanisme bawaan seperti model rekurensi (*Recurrent Neural Network, RNN*) untuk mengidentifikasi urutan spasial alami dalam gambar.

Fungsi sinus dan kosinus digunakan untuk menciptakan pola posisi yang berbeda pada setiap posisi, dan *position embedding* ini kemudian digabungkan dengan *class embedding* untuk menghasilkan *patch embeddings* yang lengkap dalam arsitektur *transformer* (Bahari & Dewi, 2024).

2.5.3 *Multi-Head Self-Attention*

Multi-Head Self-Attention (MHSA) merupakan mekanisme inti yang sangat penting dalam arsitektur Transformer, yang dirancang untuk mengatasi keterbatasan model sekuensial tradisional seperti *Recurrent Neural Network (RNN)* dan *Long Short-Term Memory (LSTM)* (Firmanto et al., 2024). Mekanisme MHSA memungkinkan model untuk secara efisien memperhitungkan seluruh konteks input alih-alih memproses data secara berurutan. Hal ini memungkinkan model untuk menangkap dependensi jangka panjang dan informasi kontekstual yang krusial untuk pemrosesan data yang akurat.

Secara struktural, MHSA adalah implementasi dari mekanisme *self-attention* yang memungkinkan model untuk fokus pada berbagai bagian input secara simultan, sehingga MHSA mampu menangkap berbagai aspek dari hubungan antar elemen dalam data (Lin et al., 2022).

Menurut Lin et al. (2022), *self-attention* dihitung menggunakan *query (Q)*, *key (K)*, dan *value (V)* dengan persamaan (1):

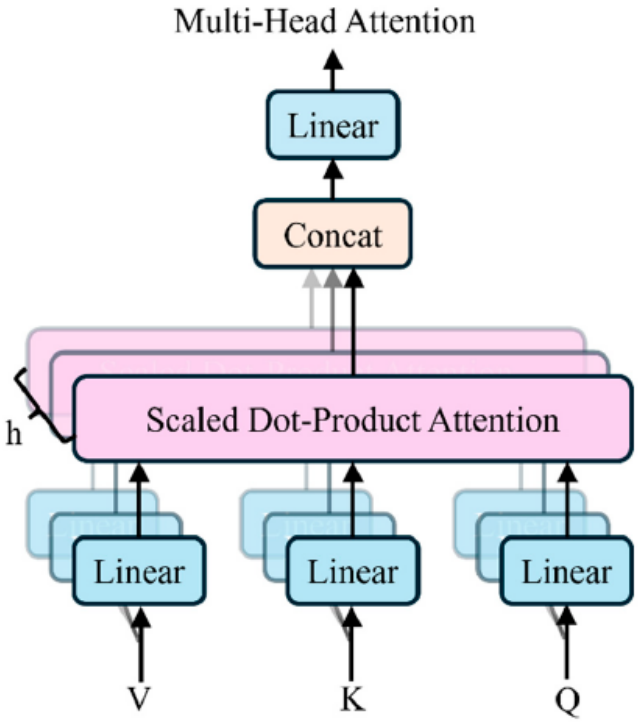
$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{D_K}}\right)V \tag{1}$$

di mana Q, K, dan V merupakan representasi input yang diperoleh dari input X dikalikan dengan bobot masing-masing:

$$Q = XW_Q, \quad K = XW_K, \quad V = XW_V \tag{2}$$

MHSA mengintegrasikan beberapa *head attention* individu, yang hasilnya kemudian digabungkan (*concat*) untuk menghasilkan output akhir. MHSA adalah komponen utama, diikuti oleh *dropout*, *layer normalization*, dan *feed-forward network* dalam

blok *encoder transformer* (Priatna et al., 2025). Visualisasi dari penjelasan tersebut ditampilkan pada Gambar 3.



Gambar 3 Struktur *Multi-Head Attention* (Zhou et al., 2025).

2.6 *Preprocessing Data*

Pra-pemrosesan citra berperan sebagai tahap esensial dan menentukan dalam sistem diagnosis berbantuan komputer *(Computer-Aided Diagnostic*/CAD) untuk analisis citra medis, khususnya pada *Magnetic Resonance Imaging* (MRI) tumor otak. Tahapan ini bertujuan meningkatkan kualitas visual citra dengan menekan gangguan seperti *noise* dan artefak, sekaligus menstandarkan format data agar proses ekstraksi fitur dan klasifikasi oleh model pembelajaran mendalam dapat berlangsung secara optimal. *Vision Transformer* (ViT) memiliki kemampuan alami dalam mengenali pola dan fitur kompleks, namun model ini tetap memerlukan masukan yang telah dipersiapkan dengan cermat guna mencapai tingkat efisiensi, akurasi, serta kecepatan pemrosesan yang tinggi (Ganesh et al., 2024).

Penerapan ViT dalam bidang pengolahan citra medis menuntut tahapan

pra-pemrosesan yang mampu menekan variasi tidak relevan pada data masukan. Proses ini menjamin keseragaman ukuran dan rentang nilai piksel sebelum citra diubah menjadi representasi vektor dan diproses oleh lapisan ViT. Normalisasi intensitas, penskalaan, dan teknik pra-pemrosesan lain juga berfungsi menyederhanakan beban komputasi model pembelajaran mendalam. Mengingat kompleksitas struktur ViT serta karakteristik data medis yang terbatas dan tidak seimbang, pra-pemrosesan yang terencana dengan baik menjadi dasar penting bagi peningkatan kemampuan generalisasi model dalam mendeteksi pola dan anomali secara andal (Kaifi, 2023).

2.6.1 *Graysale*

Konversi citra menjadi grayscale merupakan salah satu tahapan awal yang krusial dalam proses pre-processing citra medis, khususnya pada citra *Magnetic Resonance Imaging* (MRI). Proses ini bertujuan untuk menyederhanakan representasi data dengan menghilangkan informasi warna yang umumnya tidak memberikan kontribusi signifikan terhadap analisis struktur anatomis pada MRI. Dengan merepresentasikan citra dalam satu kanal intensitas, setiap piksel secara langsung mencerminkan tingkat kecerahan yang berkaitan dengan perbedaan karakteristik jaringan, sehingga pola dan batas antar jaringan dapat diamati dengan lebih jelas dan konsisten.

Selain mengurangi kompleksitas data, penggunaan grayscale juga memberikan keuntungan dari sisi komputasi dan stabilitas analisis. Representasi satu dimensi pada intensitas piksel memungkinkan algoritma ekstraksi fitur dan model pembelajaran mesin untuk bekerja lebih efisien, serta mengurangi potensi gangguan yang berasal dari variasi warna yang tidak relevan. Dalam konteks analisis MRI, pendekatan ini membantu menonjolkan kontras alami antar jaringan, seperti perbedaan antara jaringan sehat dan area abnormal, sehingga meningkatkan kualitas informasi yang digunakan pada tahap pemrosesan selanjutnya (Kaifi, 2023).

2.6.2 *Gaussian Blur*

Gaussian blur yaitu filter yang digunakan untuk mereduksi noise dengan cara menghaluskan citra. Penggunaan *gaussian blur* membantu mengurangi artefak yang dapat mengganggu proses segmentasi dan meningkatkan kualitas ekstraksi fitur dari citra MRI. Pada aplikasinya, *gaussian blur* menjadi tahap dasar dalam berbagai penelitian pengolahan citra MRI tumor otak (Kaifi, 2023). Perhitungan *gaussian blur* mengikuti persamaan (3).

$$G(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \tag{3}$$

Keterangan Variabel *Gaussian Blur*:

- $G(x, y)$: Nilai kernel Gaussian pada posisi (x, y).
- σ: Standar deviasi (mengontrol tingkat blur).
- x, y: Koordinat relatif terhadap pusat kernel.

2.6.3 *Binary Thresholding*

Binary thresholding merupakan tahapan penting dalam pre-processing citra medis yang bertujuan untuk mengonversi citra grayscale menjadi citra biner dengan dua kelas utama, yaitu objek dan latar belakang. Pada konteks citra MRI, teknik ini digunakan untuk memisahkan area tumor dari jaringan di sekitarnya berdasarkan perbedaan intensitas piksel. Dengan menetapkan suatu nilai ambang (threshold), piksel dengan intensitas di atas ambang diklasifikasikan sebagai area objek, sedangkan piksel di bawahnya dianggap sebagai latar belakang. Proses ini secara signifikan menyederhanakan representasi citra dan membantu menonjolkan struktur yang memiliki karakteristik intensitas berbeda, seperti area tumor.

Pemilihan nilai threshold yang tepat menjadi faktor krusial dalam keberhasilan segmentasi menggunakan binary thresholding. Metode otomatis seperti Otsu banyak digunakan karena mampu menentukan nilai ambang optimal dengan

memaksimalkan variansi antar kelas berdasarkan distribusi histogram intensitas piksel. Pendekatan ini mengurangi ketergantungan pada penentuan ambang secara manual serta meningkatkan konsistensi hasil segmentasi. Dengan menghasilkan citra biner yang lebih tegas, binary thresholding mempermudah proses deteksi lanjutan, pengukuran ukuran tumor, serta analisis morfologi secara otomatis, sehingga berperan penting dalam sistem analisis citra MRI berbasis komputer (Kaifi, 2023).

2.6.4 *Erosion* dan *Dilation*

Erosion dan *dilation* diterapkan pada citra biner untuk memperbaiki bentuk objek hasil segmentasi dengan menghilangkan noise kecil (*erosion*) dan memperbesar area objek (*dilation*). Kedua operasi ini sangat berguna untuk memperkuat kontur tumor yang menjadi fokus analisis dan mengurangi kesalahan segmentasi akibat fragmentasi citra (Kaifi,2023; Peng et al., 2024).

Erosi merupakan operasi morfologi yang menyebabkan penyusutan ukuran objek pada citra biner berdasarkan persamaan (4), yang berarti suatu titik z dipertahankan hanya jika seluruh elemen *structuring element* yang dipusatkan pada z termuat dalam objek. Operasi ini bekerja dengan mengikis tepi objek sehingga menghilangkan tonjolan kecil dan memisahkan objek yang terhubung oleh jembatan sempit. Dalam konteks segmentasi tumor, erosi bermanfaat untuk menghilangkan noise dan artefak kecil yang masih tersisa setelah proses *thresholding*.

$$A \ominus B = \{z \in E | B_z \subseteq A\} \tag{4}$$

Dilasi merupakan kebalikan dari erosi yang memperluas area objek berdasarkan persamaan (5), yang berarti suatu titik z menjadi bagian objek jika terdapat irisan antara refleksi *structuring element* yang dipusatkan pada z dengan objek asli. Operasi ini melebarkan tepi objek sehingga mengisi celah-celah kecil dan menyatukan fragmentasi objek yang terputus. Dalam analisis citra tumor, dilasi berguna untuk mengembalikan ukuran objek setelah erosi dan memastikan kontur tumor terbentuk secara utuh sebelum dilakukan deteksi kontur lebih lanjut.

$$A \oplus B = \{z \in E | (\hat{B})_z \cap A \neq \varnothing\} \tag{5}$$

Keterangan Variabel:

- A: Citra biner input.
- B: *Structuring element* (kernel).
- B_z: *Structuring element* yang dipusatkan pada titik z.
- $\hat{B}$: Refleksi dari B.
- E: Ruang Euclidean (domain citra).

2.6.5 *Contour Detection*

Contour detection merupakan teknik yang digunakan untuk mendeteksi batas objek tumor setelah tahap *binary thresholding* dan morfologi. Teknik ini menghasilkan garis kontur yang memudahkan visualisasi dan pengukuran geometris tumor. Proses segmentasi menjadi lebih akurat dan jelas dalam membedakan tumor dari jaringan di sekitarnya dengan adanya *contour detection* (Kaifi, 2023).

Persamaan (6) dan (7) menunjukkan kernel Sobel untuk menghitung gradien citra pada sumbu X (G_x) dan sumbu Y (G_y). Kernel G_x sensitif terhadap perubahan intensitas horizontal, sedangkan G_y sensitif terhadap perubahan intensitas vertikal. Simbol "*" merepresentasikan operasi konvolusi antara kernel dengan citra I, yang menghasilkan peta gradien pada masing-masing arah.

$$G_x = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix} * I \tag{6}$$

$$G_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ +1 & +2 & +1 \end{bmatrix} * I \tag{7}$$

Persamaan (8) menghitung magnitudo gradien (G) yang merepresentasikan kekuatan tepi pada setiap piksel, dihitung sebagai akar kuadrat dari jumlah kuadrat G_x dan

G_y. Nilai magnitudo yang tinggi menunjukkan adanya perubahan intensitas yang tajam, yang menandakan tepi objek. Persamaan (9) menghitung arah gradien (θ) yang menunjukkan orientasi tepi, diperoleh dari arcustangen rasio G_y terhadap G_x.

$$G = \sqrt{G_x^2 + G_y^2} \tag{8}$$

$$\theta = \arctan\left(\frac{G_y}{G_x}\right) \tag{9}$$

Keterangan variabel:

- I: Gambar input.
- G_x, G_y: Gradien arah horizontal dan vertikal.
- $*$: Operasi konvolusi.
- G: Magnitudo gradien.
- θ: Arah gradien.

2.6.6 *Cropping*

Cropping citra digunakan untuk memfokuskan area pemrosesan pada region of interest (tumor) sehingga mengurangi beban komputasi dan meningkatkan efisiensi metode klasifikasi dan segmentasi selanjutnya. Tahap cropping didasarkan pada bounding box dari tumor yang sebelumnya sudah terdeteksi sehingga hanya area relevan yang dianalisis lebih lanjut (Kaifi, 2023; Zhang et al.,2024). Perhitungan *cropping* ditunjukkan pada persamaan (10).

$$\text{ROI} = I[y_{\min} : y_{\max}, x_{\min} : x_{\max}] \tag{10}$$

Keterangan variabel:

- I: Citra input.
- $x_{\min}, x_{\max}$: Koordinat x minimum dan maksimum *contour*.
- $y_{\min}, y_{\max}$: Koordinat y minimum dan maksimum *contour*.

2.6.7 *Bone Color Mapping*

Bone color mapping umumnya diterapkan pada tahap *post-processing* sebagai metode visualisasi tambahan untuk meningkatkan interpretabilitas citra medis. Teknik ini bekerja dengan memetakan rentang intensitas tertentu ke dalam skema warna khusus yang dirancang untuk menonjolkan struktur tulang di sekitar area tumor. Dengan memberikan representasi warna yang lebih kontras, *bone color mapping* membantu memperjelas perbedaan antara jaringan keras dan jaringan lunak, sehingga detail anatomi yang sebelumnya kurang terlihat dalam citra *grayscale* dapat diamati dengan lebih jelas.

Selain berfungsi sebagai alat bantu visual bagi radiolog, *bone color mapping* juga berperan penting dalam mendukung sistem analisis berbasis kecerdasan buatan. Representasi warna yang konsisten pada struktur tulang memungkinkan algoritma *computer vision* dan *deep learning* untuk membedakan area tumor dari jaringan tulang secara lebih efektif, khususnya pada citra MRI atau *CT scan*. Pendekatan ini tidak hanya meningkatkan kejelasan visual, tetapi juga berkontribusi pada peningkatan akurasi interpretasi dan validasi hasil segmentasi tumor, terutama ketika struktur tulang berada dekat atau beririsan dengan area yang dianalisis (Zhang et al., 2024).

2.7 *Split Data*

Pembagian data adalah langkah penting dalam pembangunan model *machine learning*. Data dibagi melalui metode *hold-out*. Tujuannya adalah untuk melihat bagaimana model berfungsi pada data yang belum pernah dilihat sebelumnya. Pilihan rasio data pelatihan dan pengujian tidak memiliki aturan baku. Rasio 80:20

atau 90:10 dapat digunakan untuk *dataset* besar, sedangkan rasio 70:30 dapat digunakan untuk *dataset* kecil guna menghindari kekurangan data untuk validasi dan pemeriksaan (Sivakumar et al., 2024).

Strategi pembagian data selain metode acak, seperti *stratified split*, sangat krusial, terutama untuk *dataset* dengan kelas tidak seimbang. *Stratified sampling* menjamin proporsi setiap kelas target tetap terjaga dan konsisten pada data *training* serta *testing*, sehingga mengurangi bias serta meningkatkan generalisasi. Untuk data berurutan, seperti *time series*, pembagian berbasis waktu (*temporal split*) lebih sesuai dibandingkan *random split*. Hal ini bertujuan mencegah model secara tidak sengaja melihat data masa depan selama proses pelatihan (Catania et al., 2022).

2.7.1 Data Latih (*Training*)

Data latih merupakan sub-kumpulan data terbesar dan menjadi fondasi utama dalam proses pembelajaran model. Sub-kumpulan ini digunakan secara langsung oleh algoritma untuk mempelajari pola, menemukan hubungan antarvariabel, serta menyesuaikan parameter internalnya, seperti bobot (*weights*) pada jaringan saraf tiruan. Pada tahap ini, model secara iteratif dibangun dan disesuaikan untuk meminimalkan kesalahan (*error*) atau *loss* pada data yang disajikan kepadanya. Kualitas dan kuantitas data pelatihan memiliki pengaruh signifikan terhadap kemampuan akhir model (Kaifi, 2023).

Dalam konteks model *Vision Transformer* (ViT) untuk klasifikasi tumor otak, data latih berfungsi sebagai dasar pembelajaran representasi fitur kompleks dari citra MRI. Melalui mekanisme *attention* yang menjadi ciri khas arsitektur transformer, model memproses data latih untuk mengidentifikasi pola-pola visual yang relevan secara diagnostik, termasuk tekstur, intensitas, dan morfologi lesi tumor. Representasi ini kemudian dioptimalkan melalui proses *backpropagation* untuk menghasilkan pemetaan yang akurat antara fitur visual dengan label kelas tumor. Proporsi data latih yang cukup besar memastikan bahwa model memiliki eksposur yang memadai terhadap variasi alami dalam data medis, termasuk perbedaan teknik akuisisi, resolusi citra, dan karakteristik pasien, sehingga membentuk fondasi yang kokoh untuk generalisasi pada data baru.

2.7.2 Data Validasi (*Validation*)

Data validasi berfungsi sebagai proksi yang tidak bias untuk data pengujian selama proses pengembangan dan penyetelan model. Fungsinya adalah mengevaluasi kinerja model ketika melakukan penyetelan seperti *learning rate*, jumlah *hidden layer*, atau tingkat *regularization*, serta membantu dalam menentukan arsitektur model yang optimal (Ganesh et al.,2024). Penggunaan data validasi sangat penting untuk mencegah kebocoran informasi dari data pengujian. Jika data pengujian digunakan untuk penyetelan *hyperparameter*, maka hasil evaluasi akhir tidak lagi mencerminkan kinerja yang benar-benar tidak bias.

Pada implementasi ViT untuk klasifikasi tumor otak, data validasi memainkan peran kritis dalam mengontrol proses *early stopping* dan pemilihan model terbaik dari berbagai iterasi pelatihan. Metrik performa yang dihitung pada data validasi, seperti akurasi dan *F1-score*, memberikan indikasi objektif mengenai kemampuan generalisasi model sebelum evaluasi akhir pada data uji. Alokasi data validasi memungkinkan estimasi performa yang stabil secara statistik tanpa mengurangi secara berlebihan jumlah data yang tersedia untuk pelatihan. Proses validasi berulang selama pelatihan juga membantu mengidentifikasi fenomena *overfitting*, di mana model mulai menghafal pola spesifik pada data latih namun kehilangan kemampuan generalisasi, sehingga memungkinkan intervensi tepat waktu melalui penyesuaian strategi *regularization* atau penghentian pelatihan pada titik optimal.

2.7.3 Data Uji (*Testing*)

Data uji merupakan subkumpulan data yang sepenuhnya "ditahan" (*held-out*) dan tidak pernah digunakan selama proses pelatihan atau penyetelan *hyperparameter*. Tujuannya adalah memberikan satu kali evaluasi akhir yang benar-benar tidak bias terhadap kinerja dan kemampuan generalisasi model yang telah final (Chen et al.2024; Ganesh et al., 2024). Kinerja pada set data ini dianggap sebagai estimasi paling akurat terhadap performa model pada data baru di dunia nyata.

Dalam praktik terbaik, data pengujian hanya digunakan sekali setelah model terbaik dipilih dan seluruh proses penyetelan selesai dilakukan. Sebagai contoh, Rahman et al. (2023) mengalokasikan 20% dari total *dataset* secara eksklusif sebagai data

pengujian untuk evaluasi akhir setelah model terbaik dipilih menggunakan data validasi.

2.8 Augmentasi Data

Augmentasi data digunakan untuk mengatasi keterbatasan data dan meningkatkan kinerja, generalisasi, serta ketahanan model. Hal yang dilakukan dalam augmentasi data ialah meningkatkan keragaman dan ukuran dataset melalui berbagai transformasi serta menghasilkan sintesis data berdasarkan data yang sudah ada. Secara umum, metode augmentasi data diklasfikasikan menjadi 2 kategori besar, yaitu metode yang menghasilkan sampel citra baru berdasarkan data yang sudah ada (melalui transformasi langsung) dan metode yang memanfaatkan Jaringan Adversarial Generatif (*Generative Adversarial Networks*, disingkat GANs) untuk menghasilkan sampel baru. Metode yang menggunakan data yang sudah ada dapat dibagi lagi menjadi pendekatan yang mengaugmentasi sampel tunggal secara independen dan pendekatan yang melibatkan kombinasi banyak citra (Zeng, 2024).

2.8.1 Metode Augmentasi Sampel Tunggal

Metode ini melibatkan peningkatan citra yang sudah ada. Teknik yang paling umum adalah transformasi geometri yang meliputi rotasi, pemotongan (*cropping*), pembalikan (*flipping*), dan pengubahan skala. Selain transformasi geometri, ada juga teknik perubahan tingkat piksel seperti penambahan noise seperti *Gaussian noise*, pengaburan (*blurring*), dan perubahan warna (*color jittering*). Teknik-teknik ini bertujuan untuk memperkenalkan ketahanan model terhadap kondisi yang bervariasi. Metode yang lebih canggih pada kategori ini yaitu *cutout* untuk menghilangkan beberapa piksel sehingga model dapat memperhatikan wilayah yang tersisa, *random erasing* untuk menutupi daerah yang tidak berarti, dan *grid mask* dengan menyebarkan blok yang dihilangkan dalam pola kisi dan mengurangi kemungkinan hilangnya wilayah penting secara keseluruhan (Kumar et al., 2025).

2.8.2 Metode Augmentasi Gabungan Banyak Citra

Metode ini bertujuan memanfaatkan citra yang tersedia secara lebih efisien dengan menggabungkan banyak sampel untuk menghasilkan sampel baru yang lebih tinggi. Contoh metode pada kategori ini adalah *CutMix, Mixup,* dan *Mosaic*. Berikut adalah rincian dari masing-masing metode.

- *CutMix* mengatasi masalah piksel tak berarti dalam Cutout dengan mengganti wilayah yang dipotong dari satu citra dengan potongan (*patch*) dari citra lain (Yun et al., 2019).
- *Mixup* melakukan pencampuran tingkat piksel di seluruh wilayah citra untuk menciptakan sampel yang beragam dan benar-benar baru (Zeng, 2024).
- *Mosaic* adalah teknik yang populer dalam deteksi objek, yang menggabungkan dan mengompresi beberapa citra sebelum menjahitnya bersama untuk membuat sampel baru (Bochkovskiy et al., 2020).

2.8.3 Metode Berbasis GANs (*Generative Adversarial Networks*)

Metode augmentasi data berbasis *Generative Adversarial Networks* (GANs) merepresentasikan pendekatan generatif yang memanfaatkan kompetisi antara dua arsitektur jaringan saraf, yaitu generator dan diskriminator. Generator bertugas menyintesis sampel data baru dari vektor acak, sementara diskriminator berupaya membedakan secara akurat antara sampel riil dari dataset asli dan sampel buatan yang dihasilkan oleh generator. Berbeda dengan teknik augmentasi tradisional yang hanya melakukan transformasi geometris atau manipulasi warna pada citra yang sudah ada, GANs mampu menghasilkan sampel yang sepenuhnya baru dengan konten piksel yang berbeda, sehingga secara signifikan memperluas diversitas dan cakupan dataset. Proses pembelajaran iteratif ini memungkinkan model untuk menangkap distribusi fitur yang kompleks dan menghasilkan citra sintetis yang realistis, yang pada akhirnya meningkatkan performa dan kemampuan generalisasi model pembelajaran mendalam (Zeng et al., 2023).

Keunggulan strategis dari metode berbasis GANs terletak pada efektivitasnya dalam menangani fenomena ketidakseimbangan kelas (*class imbalance*) dan skenario data terbatas melalui variasi arsitektur seperti *Conditional* GAN (CGAN), *Imbalance*

CGAN (imbCGAN), dan *Data Augmentation* GAN (DAGAN). GANs dapat menyintesis sampel tambahan untuk kategori minoritas yang presisi sesuai dengan distribusi data asli melalui integrasi kondisi kategori tertentu dalam ruang laten, sehingga memitigasi risiko *overfitting*. Meskipun menawarkan potensi ekspansi data yang masif tanpa ketergantungan penuh pada sampel yang sudah ada, implementasi GANs menghadapi tantangan teknis berupa kebutuhan sumber daya komputasi yang tinggi serta sensitivitas yang ekstrem terhadap desain arsitektur, fungsi kerugian (*loss function*), dan strategi pelatihan (Zeng et al., 2023).

2.9 *Fine Tuning*

Fine-tuning merupakan teknik yang sangat penting dalam pembelajaran mendalam (*deep learning*) dan *machine learning* yang bertujuan menyempurnakan kinerja model dengan menyesuaikan kembali bobot dan bias model yang telah dilatih sebelumnya (*pre-trained model*) untuk tugas baru yang lebih spesifik (Pradana et al., 2024). Dalam konteks model Transformer, termasuk *Vision Transformers* (ViTs) dan model bahasa berbasis Transformer, *fine-tuning* merupakan metode dominan untuk mengaplikasikan model-model besar yang telah dilatih pada korpus umum atau *dataset* berskala masif seperti ImageNet ke tugas hilir (*downstream tasks*) (Ji et al., 2025).

Proses kerjanya dimulai dengan mengambil model yang telah dilatih pada *dataset* besar sebagai model dasar, kemudian lapisan terakhir yang disebut lapisan kelas atau lapisan keluaran biasanya diganti untuk menyesuaikan dengan tugas baru, dan lapisan ini dilatih ulang sementara lapisan lainnya tetap menggunakan bobot yang sudah ada (Pradana et al., 2024).

Namun, seiring meningkatnya ukuran model Transformer, pembaruan seluruh parameter (*full-model fine-tuning*) menjadi sangat mahal dan tidak efisien dari segi sumber daya komputasi. Oleh karena itu, muncul pendekatan *Parameter-Efficient Fine-Tuning* (PEFT) yang hanya memperbarui sebagian kecil parameter, yaitu sekitar 0,1% hingga 5% dari parameter spesifik tugas, namun tetap menghasilkan kinerja kompetitif. Dalam konteks *Vision Transformer Foundation Models* (ViTFs), salah satu pendekatan efisien adalah *LayerNorm fine-tuning*, yaitu hanya menyempurnakan parameter *Layer Normalization* (LayerNorm) yang dianggap

penting dalam ViTs, sehingga dapat memberikan hasil kompetitif atau bahkan superior dengan pembaruan bobot model yang minimal (Tan et al., 2024).

2.10 Teorema Bayes

Teorema Bayes merupakan konsep fundamental dalam teori probabilitas dan statistika yang dikembangkan oleh matematikawan Inggris Thomas Bayes pada abad ke-18. Teorema ini menggunakan kerangka statistik fundamental yang digunakan untuk melakukan inferensi dengan menggabungkan informasi awal (*prior*) dan data observasi (*likelihood*) guna menghasilkan distribusi posterior sebagai dasar pengambilan keputusan. Berbeda dengan pendekatan statistik konvensional atau frekuentis yang umumnya bersifat deterministik dan cenderung menghasilkan satu nilai estimasi tunggal, pendekatan Bayesian memungkinkan prediksi direpresentasikan dalam bentuk distribusi probabilitas yang eksplisit terhadap ketidakpastian (Ramadani et al., 2026). Secara filosofis, teori ini menjelaskan bagaimana derajat kepercayaan subjektif seseorang terhadap suatu hipotesis harus diperbarui secara rasional ketika terdapat bukti atau informasi baru yang diobservasi (Ahzuri et al., 2024).

Secara operasional, teori ini menyediakan struktur matematis formal untuk memperbarui probabilitas suatu peristiwa melalui komponen probabilitas prior yang mencerminkan pengetahuan awal, *likelihood* yang merepresentasikan peluang data terhadap parameter model, serta probabilitas posterior yang mencerminkan estimasi terbaru beserta tingkat ketidakpastiannya. Kekuatan utama dari pendekatan Bayesian dalam konteks penelitian ilmiah terletak pada fleksibilitasnya untuk mengintegrasikan pengetahuan subjektif pakar ke dalam proses pemodelan, sehingga sangat efektif diterapkan pada kondisi keterbatasan data historis dan fluktuasi yang tinggi. Melalui penyediaan interval kepercayaan (*credible interval*), teori ini menghasilkan informasi probabilistik yang lebih komprehensif dibandingkan metode titik tunggal, yang memungkinkan analisis risiko dilakukan secara lebih terukur dan rasional (Ramadani et al., 2026).

Pembentukan Teorema Bayes secara teoretis didasarkan pada prinsip-prinsip dasar probabilitas kondisional dan probabilitas konjungtif yang diturunkan melalui

beberapa langkah aljabar yang sistematis. Berikut adalah tahapan pembentukan teorema Bayes menurut Downey (2012):

1. Landasan utama dari Teorema Bayes adalah probabilitas kondisional, yang disimbolkan sebagai $p(A \mid B)$. Notasi ini merepresentasikan peluang terjadinya peristiwa A dengan syarat bahwa peristiwa B telah terbukti benar atau terjadi. Nilai probabilitas ini bersifat dinamis karena sangat bergantung pada latar belakang informasi atau kondisi tertentu yang melingkupinya.

2. Langkah selanjutnya melibatkan pemahaman mengenai probabilitas konjungtif, yaitu peluang di mana dua peristiwa, A dan B, terjadi secara bersamaan, yang ditulis sebagai $p(A \cap B)$. Secara matematis, probabilitas dari sebuah konjungsi dinyatakan dengan rumus:

$$p(A \cap B) = p(A) \cdot p(B \mid A) \tag{11}$$

   Persamaan ini menyatakan bahwa peluang terjadinya A dan B secara bersamaan adalah hasil perkalian antara peluang A dengan peluang kondisional B setelah A terjadi.

3. Tahap krusial dalam derivasi ini adalah pengaplikasian sifat komutatif pada operasi konjungsi. Dalam teori probabilitas, urutan antara dua peristiwa dalam sebuah konjungsi tidak mengubah nilai probabilitas keseluruhannya, sehingga berlaku persamaan:

$$p(A \cap B) = p(B \cap A) \tag{12}$$

4. Berdasarkan sifat komutatif tersebut, terdapat dua cara untuk menghitung nilai konjungsi yang sama. Jika telah mengetahui $p(A)$, perhitungan dapat mengalikannya dengan $p(B \mid A)$. Sebaliknya, jika telah mengetahui $p(B)$, kita dapat mengalikannya dengan $p(A \mid B)$. Hal ini menghasilkan kesamaan matematis sebagai berikut:

$$p(B) \cdot p(A \mid B) = p(A) \cdot p(B \mid A) \tag{13}$$

5. Untuk mendapatkan formula Teorema Bayes yang umum digunakan, persamaan 13 dibagi dengan variabel $p(B)$ guna mengisolasi nilai probabilitas kondisional $p(A \mid B)$. Hasil akhirnya adalah:

$$p(A \mid B) = \frac{p(A) \cdot p(B \mid A)}{p(B)} \tag{14}$$

Berdasarkan persamaan 14, terdapat empat komponen utama yang menyusun teori ini:

- *Prior Probability* ($P(A)$): Merupakan pengetahuan awal atau asumsi mengenai pola yang diamati sebelum data baru diperhitungkan.
- *Likelihood* ($P(B \mid A)$): Merepresentasikan probabilitas munculnya data yang diamati dengan asumsi bahwa hipotesis atau parameter model tertentu adalah benar.
- *Posterior Probability* ($P(A \mid B)$): Hasil pembaruan pengetahuan yang mencerminkan estimasi terbaru beserta tingkat ketidakpastiannya setelah bukti data diperhitungkan.
- *Marginal Likelihood/Evidence* ($P(B)$): Probabilitas total dari data yang diamati di bawah semua kemungkinan hipotesis, berfungsi sebagai faktor normalisasi agar total probabilitas bernilai satu.

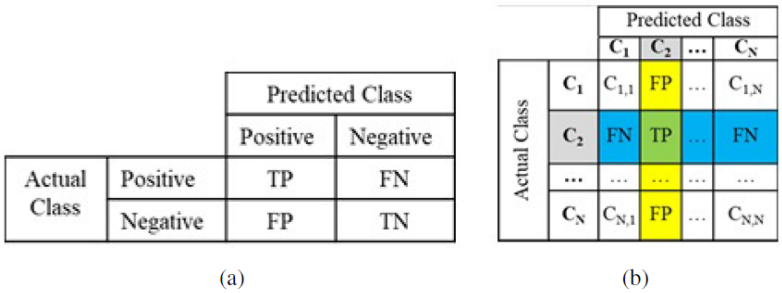
2.11 Evaluasi Performa Model

Menurut Zhao et al. (2024), terdapat berbagai metrik evaluasi yang dapat digunakan untuk evaluasi model. Elemen-elemen dari evaluasi model adalah sebagai berikut.

- *True Negative* (TN) menunjukkan jumlah data yang sebenarnya negatif dan diprediksi negatif oleh model.
- *True Positive* (TP) menunjukkan jumlah data yang sebenarnya positif dan diprediksi positif oleh model.
- *False Negative* (FN) menunjukkan jumlah data yang sebenarnya positif, tetapi diprediksi negatif oleh model.
- *False Positive* (FP) menunjukkan jumlah data yang sebenarnya negatif, tetapi diprediksi positif oleh model.

Evaluasi performa model klasifikasi pada skenario *multitarget* atau *multi-class* memerlukan dekomposisi matriks kebingungan (*confusion matrix*) ke dalam elemen-elemen fundamental melalui pendekatan *one-versus-all*. Dalam kerangka kerja statistik Bayesian, proses evaluasi ini didasarkan pada prinsip pembaruan

probabilitas guna menghasilkan estimasi kelas yang paling representatif terhadap data observasi (Bishop, 2006). Pada struktur *confusion matrix* berukuran $N \times N$, di mana N merepresentasikan jumlah kategori, setiap baris menunjukkan akumulasi instansi pada kelas prediksi sementara kolom merepresentasikan instansi pada kelas aktual (Sathyanarayanan & Tantri, 2024). Gambar 4 menunjukkan penempatan TN, TP, FN, dan FP dalam *confussion matrix* pada banyak kelas yang berbeda.



Gambar 4 *Confussion Matrix*: (a) Dua kelas (b) *Multi class* (Markoulidakis et al., 2021).

Secara deskriptif, elemen *True Positive* (TP) untuk kelas C_i merepresentasikan kondisi di mana model, melalui kalkulasi probabilitas posterior, secara akurat memprediksi instansi yang secara aktual memang termasuk dalam kategori C_i. Sebaliknya, *False Positive* (FP), yang sering diidentifikasi sebagai galat Tipe I, terjadi ketika instansi dari kelas lain (C_j di mana $j \neq i$) diprediksi secara keliru sebagai kelas C_i karena nilai *likelihood* dan prior kelas tersebut menghasilkan probabilitas posterior yang dominan pada wilayah keputusan yang salah. Elemen *False Negative* (FN) atau galat Tipe II muncul apabila instansi aktual kelas C_i gagal diidentifikasi oleh model dan justru dikategorikan ke dalam kelas lain, yang mengindikasikan bahwa model memiliki tingkat sensitivitas yang rendah terhadap fitur unik kelas tersebut. Terakhir, *True Negative* (TN) mencakup seluruh instansi aktual selain kelas C_i yang berhasil diprediksi secara tepat oleh model ke dalam kategori-kategori selain C_i, yang secara kolektif mencerminkan kemampuan model dalam melakukan penolakan yang benar (*correct rejection*) terhadap data yang tidak relevan dengan target kelas terkait (Sathyanarayanan & Tantri, 2024).

Notasi matematis untuk elemen-elemen *confusion matrix* (TP, TN, FP, FN) bagi suatu kelas target C_i dalam sistem *multi-class* dengan label aktual Y dan label prediksi $\hat{Y}$ dapat didefinisikan sebagai berikut.

- *True Positive* (TP): $Y = C_i \cap \hat{Y} = C_i$

- *True Negative* (TN): $Y \neq C_i \cap \hat{Y} \neq C_i$
- *False Positive* (FP): $Y \neq C_i \cap \hat{Y} = C_i$
- *False Negative* (FN): $Y = C_i \cap \hat{Y} \neq C_i$

Setelah memahami definisi dan notasi matematis dari setiap elemen *confusion matrix*, langkah selanjutnya adalah memanfaatkan nilai-nilai tersebut untuk mengevaluasi kinerja model secara kuantitatif. Berbagai metrik evaluasi dapat dihitung berdasarkan kombinasi TP, TN, FP, dan FN, yang masing-masing menyoroti aspek berbeda dari kemampuan klasifikasi model. Berikut ini adalah penjelasan lengkap mengenai metrik-metrik tersebut beserta formulasi matematisnya.

2.11.1 Akurasi

Akurasi merupakan metrik evaluasi paling sederhana yang mengukur proporsi prediksi benar yang dilakukan model dari keseluruhan jumlah prediksi. Metrik ini memberikan gambaran umum tentang sejauh mana model mampu mengklasifikasikan data dengan tepat, baik untuk kelas positif maupun negatif. Dalam eksperimen klasifikasi gambar menggunakan Transfer Learning, akurasi sering digunakan sebagai indikator utama untuk membandingkan kinerja berbagai arsitektur model pada dataset tertentu.

Dari sisi matematika probabilistik, akurasi dapat dipandang sebagai estimasi dari peluang bahwa label yang diprediksi oleh model ($\hat{Y}$) sama dengan label sebenarnya (Y), atau secara simbolis ditulis pada persamaan 15. Pendekatan ini dihitung dengan membagi jumlah prediksi positif yang benar (*True Positives*) dan prediksi negatif yang benar (*True Negatives*) dengan total populasi data. Persamaan dari akurasi ditunjukkan pada persamaan 16

$$\text{Akurasi} = P(\hat{Y} = Y) \tag{15}$$

$$\text{Akurasi} = \frac{TN + TP}{TN + FN + TP + FP} \tag{16}$$

2.11.2 Presisi

Presisi adalah metrik kinerja yang secara khusus berfokus pada kemampuan model dalam memprediksi kelas positif secara akurat. Metrik ini menunjukkan tingkat kepercayaan terhadap prediksi positif model, dengan tujuan meminimalkan kesalahan prediksi positif yang salah (False Positives). Presisi sangat penting dalam skenario di mana biaya atau risiko dari kesalahan deteksi positif palsu sangat tinggi, seperti dalam sistem rekomendasi atau deteksi objek tertentu.

Secara probabilistik, presisi merepresentasikan peluang bersyarat bahwa suatu instance benar-benar positif diberikan model telah memprediksinya sebagai positif yang ditunjukkan pada persamaan 17. Notasi $Y = k$ menunjukkan bahwa kelas sebenarnya adalah kelas k, sedangkan notasi $\hat{Y} = k$ menunjukkan bahwa kelas prediksi adalah kelas k. Nilai ini dihitung dengan membagi True Positives dengan total jumlah prediksi positif (*True Positive* + *False Positive*). Persamaan dari presisi ditunjukkan pada persamaan 18.

$$\text{Presisi} = P(Y = k \mid \hat{Y} = k) = \frac{P(Y = k \cap \hat{Y} = k)}{P(\hat{Y} = k)} \tag{17}$$

$$\text{Presisi} = \frac{TP}{TP + FP} \tag{18}$$

2.11.3 *Sensitivity* (*Recall*)

Sensitivity, yang juga sering disebut sebagai *recall*, mengukur persentase dari seluruh *instance* positif aktual yang berhasil diidentifikasi dengan benar oleh model. Metrik ini sangat krusial dalam bidang medis, seperti deteksi COVID-19 atau penyakit langka, di mana kegagalan dalam menemukan kasus positif (*False Negative*) dapat berakibat fatal bagi pasien. *Sensitivity* memberikan informasi tentang seberapa luas cakupan model dalam menangkap semua kasus relevan dalam sebuah dataset.

Dalam pendekatan probabilistik, *sensitivity* adalah peluang bersyarat bahwa model memprediksi positif diberikan instance tersebut memang positif secara aktual, atau $P(\hat{Y} = k \mid Y = k)$. Notasi $\hat{Y} \neq k$ menunjukkan bahwa kelas prediksi adalah

selain kelas k. Metrik ini juga dikenal sebagai *True Positive Rate* (TPR) dalam analisis kurva ROC dan berbanding terbalik dengan *False Negative Rate* (FNR) yang dinotasikan dengan $P(\hat{Y} \neq k \mid Y = k)$. Persamaan dari *sensitivity* ditunjukkan pada persamaan 19 hingga 21.

$$\text{Sensitivity} = P(\hat{Y} = k \mid Y = k) = \frac{P(\hat{Y} = k \cap Y = k)}{P(Y = k)} \tag{19}$$

$$\text{Sensitivity} = 1 - FNR = 1 - P(\hat{Y} \neq k \mid Y = k) \tag{20}$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{21}$$

2.11.4 *Specificity*

Specificity mengukur kemampuan model dalam mengidentifikasi *instance* negatif secara benar. Meskipun sumber tidak mendefinisikannya secara terpisah seperti presisi, metrik ini berkaitan erat dengan deteksi *True Negative* yang disebut dalam penghitungan akurasi. Metrik ini penting untuk memastikan bahwa model tidak salah mengidentifikasi data normal sebagai sesuatu yang anomali, terutama dalam tugas-tugas seperti deteksi *Out-of-Distribution* (OOD) atau diagnosis medis untuk menghindari prediksi yang keliru.

Secara probabilistik, *specificity* mewakili peluang bersyarat bahwa model memprediksi negatif diberikan bahwa *instance* tersebut memang negatif secara aktual, atau $P(\hat{Y} \neq k \mid Y \neq k)$. Notasi $Y \neq k$ menunjukkan bahwa kelas sebenarnya adalah selain kelas k. Metrik ini berbanding terbalik dengan *False Positive Rate* (FPR), di mana nilai *Specificity* $= 1 -$ FPR dan FPR dinotasikan sebagai $P(\hat{Y} = k \mid Y \neq k)$. Hal ini mencerminkan sejauh mana model mampu menolak kategori yang salah. Persamaan dari *specificity* ditunjukkan pada persamaan 22 hingga 24.

$$\text{Specificity} = P(\hat{Y} \neq k \mid Y \neq k) = \frac{P(\hat{Y} \neq k \cap Y \neq k)}{P(Y \neq k)} \tag{22}$$

$$\text{Specificity} = 1 - \text{FPR} = 1 - P(\hat{Y} = k \mid Y \neq k) \tag{23}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{24}$$

2.11.5 *F1 Score*

F1-Score adalah metrik yang memberikan evaluasi seimbang dengan menggabungkan presisi dan *recall* melalui rata-rata harmonik. Metrik ini sangat berguna ketika terdapat ketidakseimbangan kelas dalam data, atau ketika peneliti perlu mempertimbangkan kedua jenis kesalahan (*False Positive* dan *False Negative*) secara bersamaan untuk mendapatkan skor tunggal yang adekuat. Dalam pengujian arsitektur CNN, *F1-Score* sering digunakan untuk memvalidasi apakah model tetap tangguh meskipun performa presisi dan *recall*-nya bervariasi.

Dari sisi matematika probabilistik, *F1-Score* mencoba mengoptimalkan titik temu antara keyakinan prediksi ($P(Y = k \mid \hat{Y} = k)$) dan cakupan aktual ($P(\hat{Y} = k \mid Y = k)$). Dengan menggunakan rata-rata harmonik daripada rata-rata aritmatika, metrik ini memastikan bahwa jika salah satu dari presisi atau *recall* memiliki nilai yang sangat rendah, maka nilai *F1-Score* akan menurun secara signifikan. Persamaan dari F1 Score ditunjukkan pada persamaan 25 dan 26.

$$F1 = 2 \times \frac{\text{Presisi} \cdot \text{Recall}}{\text{Presisi} + \text{Recall}} \tag{25}$$

$$F1 = 2 \times \frac{P(\hat{Y} = k \mid Y = k) \cdot P(\hat{Y} = k \mid Y = k)}{P(\hat{Y} = k \mid Y = k) + P(\hat{Y} = k \mid Y = k)} \tag{26}$$

BAB III

METODE PENELITIAN

3.1 Waktu dan Tempat Penelitian

3.1.1 Tempat Penelitian

Penelitian ini dilakukan pada semester ganjil tahun ajaran 2025/2026 di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung yang beralamatkan di Jalan Prof. Dr. Ir. Soemantri Brojonegoro, Gedong Meneng, Kecamatan Rajabasa, Kota Bandar Lampung, Lampung.

3.1.2 Waktu Penelitian

Penelitian dimulai dari bulan Agustus 2025. Penelitian ini dilakukan ke dalam tiga tahap. Tahap pertama yaitu tahap awal penelitian yang terdiri dari studi literatur, penentuan tema, pengumpulan data, penyusunan bab pertama hingga ketiga, dan seminar proposal. Tahap kedua merupakan tahap lanjutan penelitian setelah dilaksanakan seminar proposal. Tahap ketiga merupakan tahap terakhir yang terdiri dari penyusunan bab keempat hingga kelima, seminar hasil penelitian, dan seminar komprehensif.

3.2 Data Penelitian

Data yang digunakan adalah dataset sekunder dari Kaggle yang berupa gambar MRI tumor otak. Dataset terdiri dari 7023 gambar yang dibagi menjadi 4 kelas, yaitu *glioma*, *meningioma*, *no tumor*, dan *pituitary*. Berikut penjelasan dari masing-masing kelas.

- *Glioma* adalah tumor ganas yang berasal dari sel glial di otak atau sumsum tulang belakang, dikenal agresif dan berpotensi menyebar ke jaringan otak sekitar.
- *Meningioma* adalah tumor yang biasanya jinak dan tumbuh lambat, berasal dari selaput pelindung otak dan sumsum tulang belakang (meningen).
- *No tumor* mengacu pada kondisi normal tanpa adanya pertumbuhan massa atau lesi abnormal pada citra MRI otak.
- *Pituitary* (atau adenoma hipofisis) adalah tumor yang umumnya jinak yang tumbuh di kelenjar pituitari, seringkali memengaruhi sistem hormonal tubuh.

Dataset ini merupakan gabungan dari 3 dataset yang berasal dari Figshare, SARTAJ Dataset, dan Br35H. Setiap gambar MRI memiliki ukuran 512×512 piksel dan berformat JPG dengan *bit depth* sebesar 24. Bit depth menunjukkan jumlah bit yang digunakan untuk menyimpan informasi warna setiap piksel atau setiap komponen warna dalam suatu gambar. Tautan data penelitian dapat diakses melalui `https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri- dataset`. Tabel 2 memperlihatkan rincian dataset yang digunakan dalam penelitian.

Tabel 2 Distribusi Dataset Tumor Otak

Kelas	**Total**	**Proporsi**
Glioma	1621	23,08%
Meningioma	1645	23,42%
Pituitary	1757	25,02%
No tumour	2000	28,48%
Total	**7023**	**100%**

3.3 Alat Penelitian

Adapun spesifikasi perangkat keras (*hardware*) dan perangkat lunak (*software*) yang digunakan dalam penelitian ini.

1. Perangkat Keras (*Hardware*)
   Perangkat keras yang digunakan dalam penelitian adalah laptop dengan merek HP 14 inch Laptop 14-ep0002TU. Perangkat ini memiliki memori (RAM)

sebesar 8 GB DDR4 dan penyimpanan SSD sebesar 512 GB. Prosesor pada perangkat ini adalah Intel® Core™ i5-1335U 13th Generation.

2. Perangkat Lunak (*Software*)
   Perangkat lunak digunakan untuk menunjang penelitian ini. Berikut rinciannya.

   (a) Sistem operasi Windows 11 *Home* 64-bit

   (b) Python versi 3.12.12
   Peneliti menggunaka website Google Colaboratory sebagai *notebook* Python. Adapun *library* yang digunakan dalam penelitian ini:

   - TensorFlow versi 2.19.0
     Library deep learning *end-to-end* untuk membuat dan melatih model neural network dengan skalabilitas produksi. TensorFlow mendukung *deployment multi-platform* dan komputasi GPU/TPU terakselerasi.
   - PyTorch versi 2.8.0+cu126
     Framework deep learning dengan komputasi tensor dinamis dan autograd yang intuitif untuk penelitian. Versi CU126 mendukung akselerasi GPU NVIDIA dengan kompatibilitas CUDA terbaru.
   - NumPy versi 2.0.2
     Library fundamental untuk komputasi saintifik dengan array multidimensi dan fungsi matematika efisien. Library ini menyediakan struktur data dasar untuk hampir semua ekosistem *data science* Python.
   - Pandas versi 2.2.2
     Tool untuk manipulasi dan analisis data terstruktur dengan struktur DataFrame dan Series. Menyediakan operasi pembersihan, transformasi, dan agregasi data.
   - Matplotlib versi 3.10.0
     Library visualisasi data 2D yang komprehensif untuk membuat plot statis, interaktif, dan animasi. Merupakan fondasi plotting Python dengan kustomisasi tingkat rendah yang fleksibel.
   - Scikit-learn versi 1.6.1
     Library machine learning untuk algoritma klasik dan tool preprocessing data yang sederhana. Menyediakan API konsisten untuk training, evaluasi, dan validasi model.

- Pillow (PIL) versi 11.3.0
  Library ini digunakan untuk membuka, memanipulasi, dan menyimpan berbagai format gambar (seperti JPEG, PNG, BMP, GIF) di Python.

Tabel 3 menunjukkan ringkasan singkat mengenai alat penelitian.

Tabel 3 Spesifikasi Alat Penelitian

Komponen	**Spesifikasi**
Merek dan Tipe	HP 14 inch Laptop 14-ep0002TU
Prosesor	Intel® Core™ i5-1335U 13th Generation
Memori (RAM)	8 GB DDR4 RAM
Penyimpanan	512 GB SSD Hard drive
Kartu Grafis (GPU)	Intel Iris X Graphics
Sistem Operasi	Windows 11 Home 64-bit
Bahasa Pemrograman	Python 3.12.12
Library Pendukung	TensorFlow 2.19.0, PyTorch 2.8.0+cu126, NumPy 2.0.2, Pandas 2.2.2, Matplotlib 3.10.0, Scikit-learn 1.6.1, dan Pillow 11.3.0

3.4 Langkah Penelitian

Metode penelitian yang digunakan yaitu:

1. Mengumpulkan dataset citra MRI tumor otak dari Kaggle.
2. Pre-processing data meliputi:
   - Konversi dan Normalisasi: Mengkonversi citra ke skala abu-abu (grayscale) dan menormalisasi nilai pixel ke rentang [0,1] atau [-1,1] untuk stabilitas numerik.
   - Peningkatan Kualitas Citra: Mereduksi noise menggunakan filter Gaussian.
   - Segmentasi dan Ekstraksi ROI (*Region of Interest*): Memotong (cropping) citra untuk berfokus pada area yang mengandung tumor guna mengurangi komputasi dan noise dari latar belakang.

3. Dataset dibagi menjadi data latih (*training*), validasi (*validation*), dan uji (*testing*). Data latih akan dilakukan augmentasi data untuk mengatasi keterbatasan data dan mencegah overfitting. Data validasi dan testing tidak dilakukan augmentasi.

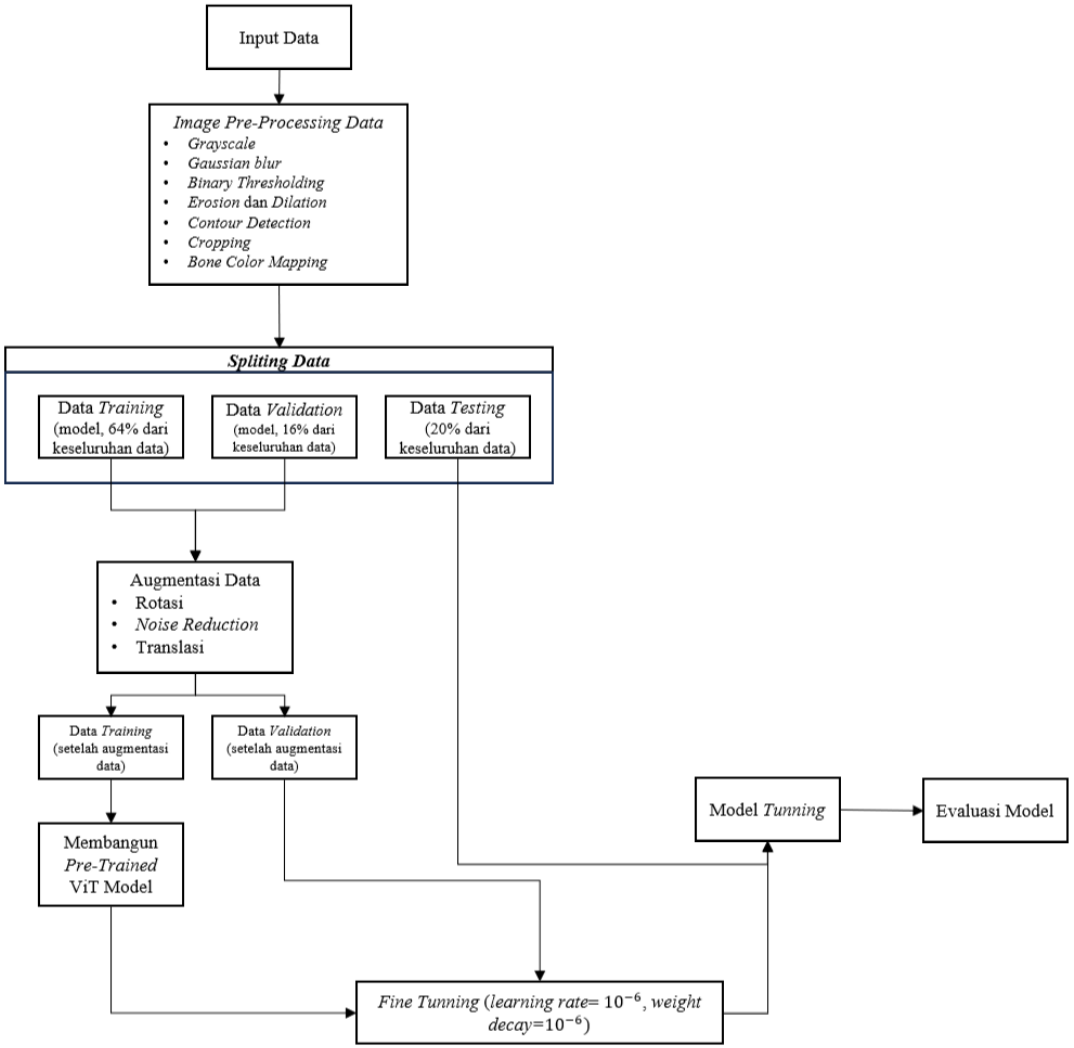
4. Pembangunan model menggunakan arsitektur *Vision Transformer* (ViT) *pre-train ed* dengan konfigurasi *patch size* 16×16 dan *input resolution* 224×224 piksel. Implementasi model mengadopsi pendekatan *transfer learning* dengan memanfaatkan *weights* pre-trained dari *timm library*. Adaptasi model untuk tugas klasifikasi tumor otak dilakukan dengan memodifikasi layer klasifikasi akhir (*head*) untuk menyesuaikan dengan jumlah kelas dalam dataset. Model di-deploy pada komputasi GPU untuk mengoptimasi proses training dan inferensi.

5. Fine-tuning dilakukan selama 10 *epoch* menggunakan *optimizer* AdamW dengan *learning rate* 10^{-6} dan *weight decay* 10^{-6}. Proses training memanfaatkan *gradient clipping* untuk stabilitas dan dievaluasi menggunakan *loss function* CrossEntropyLoss, dengan monitoring *real-time* akurasi pada data training dan validasi.

6. Proses pelatihan akan dipantau dengan melihat kurva *loss* dan akurasi (atau metrik lain seperti F1-Score) baik pada data latih maupun validasi untuk setiap *epoch*.

7. Model terbaik yang diperoleh dari proses pelatihan akan diuji secara komprehensif menggunakan data uji (*testing set*) yang belum pernah dilihat sama sekali selama proses pelatihan dan fine-tuning. Evaluasi model dilakukan untuk mengukur performa model secara objektif.

Gambar 5 menunjukkan alur penelitian yang akan dilakukan.



Gambar 5 Alur Tahapan Penelitian.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil eksperimen dan pembahasan yang telah dilakukan, penelitian ini menyimpulkan hal-hal berikut:

1. Model Vision Transformer (ViT) membuktikan efektivitasnya dalam klasifikasi tumor otak berbasis citra MRI. Model mencapai kinerja yang sangat kompetitif dengan akurasi akhir sebesar 97.08%, presisi 97.01%, *recall* 97.01%, spesifisitas 99.03% dan *F1-Score* 97.00%. Hasil ini menunjukkan model ViT dalam menangkap dependensi global dan konteks spasial pada citra medis, sekaligus menjawab tujuan pertama dan kedua penelitian.

2. Penyeimbangan data melalui pembobotan kelas dan augmentasi terbukti menjadi faktor kritis. Strategi ini secara signifikan meningkatkan stabilitas pelatihan, mengurangi risiko *overfitting*, dan mendorong peningkatan kinerja model pada semua kelas, terutama kelas minoritas. Peningkatan akurasi sebesar 0.57% dan *F1-Score* sebesar 0.19% mengonfirmasi pengaruh positifnya terhadap generalisasi model.

3. Strategi *fine-tuning* yang komprehensif menghasilkan kinerja lebih unggul dibanding *fine-tuning* parsial. Meskipun *fine-tuning* hanya pada *head classifier* lebih efisien waktu di setiap *epoch*, pendekatan ini mengorbankan akurasi secara signifikan. *Fine-tuning* seluruh lapisan model diperlukan untuk mengoptimalkan kemampuan representasi fitur ViT dalam konteks spesifik data medis.

5.2 Saran

Untuk pengembangan penelitian selanjutnya, dapat dipertimbangkan beberapa saran berikut:

1. Melakukan eksplorasi *hyperparameter tuning* yang lebih mendalam. Penelitian ini menggunakan konfigurasi *baseline*. Eksperimen lebih lanjut terhadap *learning rate*, *batch size*, *optimizer*, atau skema *weight decay* yang berbeda berpotensi meningkatkan konvergensi dan akurasi model secara lebih optimal.
2. Mengintegrasikan teknik interpretabilitas model seperti *attention map*. Visualisasi daerah mana dalam citra MRI yang menjadi fokus perhatian model ViT akan meningkatkan transparansi dan kepercayaan terhadap prediksi. Hal ini sangat berharga untuk validasi klinis dan pemahaman pola deteksi model.
3. Menguji model pada dataset eksternal yang lebih beragam. Evaluasi terhadap data dari sumber berbeda dengan variasi perangkat MRI akan memberikan gambaran lebih nyata mengenai *robustness* dan kemampuan generalisasi model dalam skenario klinis yang sesungguhnya.
4. Memperkaya variasi teknik augmentasi data yang tetap mempertimbangkan aspek keragaman secara anatomis agar model ViT terpapar pada spektrum variasi yang lebih luas. Pendekatan ini diharapkan dapat meningkatkan kemampuan generalisasi model dalam mendeteksi tumor otak pada citra MRI dengan kondisi akuisisi yang beragam di lingkungan klinis sebenarnya.

DAFTAR PUSTAKA

Ahzuri, A., Zulfah, Astuti, & Ediputra, K. (2024). Teorema Bayes: Statistika matematika kecerdasan buatan dan pembelajaran mesin. *Cendekia: Jurnal Ilmiah Multidisiplin*, 1(2), 9-14.

Andre, R. R., Wahyu, B. P., & Purbaningtyas, R. (2021). Klasifikasi tumor otak menggunakan convolutional neural network dengan arsitektur EfficientNet-B3. *Jurnal Sistem dan Teknologi Informasi, 11*(3), 55–59.

Archana, B., Karthigha, M., & Suresh Lavanya, D. (2023). Comparative analysis of optimisers used in CNN for brain tumor detection. *Institute of Electrical and Electronics Engineers (IEEE).* (Proceedings of 2023 International Conference on Recent Trends in Electronics and Communication, pp. 1-6).

Astuti, A. W., Prasetya, I. M. L., & Hariyanto, T. (2023). Teknik Pemeriksaan Magnetic Resonance Imaging (MRI) Brain Dengan Penambahan Sequence Axial 3D-Fiesta Pada Kasus Facial Tic. *Compromise Journal, 1*(4), 18–28.

Auliaddina, S., & Arifin, T. (2024). Penggunaan Data Augmentasi dan Hyperparameter Tuning dalam Klasifikasi Jenis Batik menggunakan Model CNN. *SISTEMASI: Jurnal Sistem Informasi, 13*(1), 114–128.

Bahari, A., & Dewi, K. E. (2024). Peringkasan teks otomatis abstraktif menggunakan transformer pada teks bahasa Indonesia. *KOMPUTA: Jurnal Ilmiah Komputer dan Informatika, 13*(1), 45–56.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer Science+Business Media.

Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection. DOI: 2004.10934.

Catania, C., Guerra, J., Romero, J. M., Caffaratti, G., & Marchetta, M. (2022). Beyond Random Split for Assessing Statistical Model Performance. `https://doi.org/10.48550/arXiv.2209.03346`

Chen, W., Tan, X., Zhang, J., Du, G., Fu, Q., & Jiang, H. (2024). A robust approach for multi-type classification of brain tumor using deep feature fusion. *Frontiers in Neuroscience, 18*, 1288274. https://doi.org/10.3389/fnins.2024.1288274

Chibuike, O., & Yang, X. (2024). Convolutional Neural Network–Vision Transformer Architecture with Gated Control Mechanism and Multi-Scale Fusion for Enhanced Pulmonary Disease Classification. *Diagnostics, 14*(14), 2790. https://doi.org/10.3390/diagnostics14142790

Diponegoro, M. H., Kusumawardani, S. S., & Hidayah, I. (2021). Tinjauan pustaka sistematis: Implementasi metode deep learning pada prediksi kinerja murid. *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*, 10(2), 1-8.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). An image is worth 16x16 words: Transformers for image recognition at scale. *Proceedings of the 9th International Conference on Learning Representations (ICLR)*.

Downey, A. B. (2012). *Think Bayes: Bayesian statistics made simple* (Versi 1.0.9). Green Tea Press.

Dzakwan, M. N. A., Prawoto, F. D., & Purwanto, A. N. I. (2025). Tren dan Perkembangan Supervised versus Unsupervised Learning. *Jurnal Teknik Informatika dan Teknologi Informasi, 5*(2), 619–625.

Essianda, V., Indrasari, A. D., Widyastuti, P., Syahla, T., & Rohadi. (2023). Brain Tumor: Molecular Biology, Pathophysiology, and Clinical Symptoms. *Jurnal Biologi Tropis, 23*(4), 260–269.

Fadilah, S. N., Rini, D. C., & Hakim, L. (2023). Pengaruh Reduksi Fitur Pada Klasifikasi Kanker Paru Menggunakan CNN Dengan Arsitektur GoogLeNet. *Jurnal Fourier, 12*(1), 20–32.

Fatmah, C. R., Fintri, I., & Indra, R. R. (2023). Klasifikasi Tumor Otak Berbasis Magnetic Resonance Imaging Menggunakan Algoritma Convolutional Neural Network. *Digital Transformation Technology (Digitech), 3*(2), 918–924.

Firmanto, B., Aziz, A. S., & Sesoca, J. (2024). Tinjauan Perkembangan Kecerdasan Buatan Berbasis Arsitektur Transformer. *SMARTICS Journal, 10*(1), 33–38.

Ganesh, S., Kannadhasan, S., & Jayachandran, A. (2024). Multi class robust brain tumor with hybrid classification using DTA algorithm. *Heliyon, 10*(2), e23610. https://doi.org/10.1016/j.heliyon.2023.e23610

Garcia, A., Zhou, J., Pinero-Crespo, G., Beachkofsky, T., & Huang, X. (2025). *Clinical application of Vision Transformers for melanoma classification: A multi-dataset evaluation study*. *Cancers*, 17(21), 3447. DOI:10.3390/cancers17213447

Halder, A., Gharami, S., Sadhu, P., Singh, P. K., Woźniak, M., & Ijaz, M. F. (2024). Implementing vision transformer for classifying 2D biomedical images. *Scientific Reports, 14*(1), 12567. https://doi.org/10.1038/s41598-024-62843-0

Harianto, F. R. A., Alawi, Z., & Sa'ida, I. A. (2025). Pengaruh Komposisi Split Data pada Akurasi Klasifikasi Penderita Diabetes Menggunakan Algoritma Machine Learning. *Jurnal Sistem Informasi dan Informatika (Simika), 8*(1), 36–43.

Husen, D. (2024). Klasifikasi citra MRI tumor otak menggunakan metode convolutional neural network. *Bit-Tech (Binary Digital – Technology), 7*(1), 143–152.

Ji, X., Zhao, Z., Chen, X., Zhao, X., & Liu, Z. (2025). Progtuning: Progressive fine-tuning framework for transformer-based language models. In *International Conference on Neural Information Processing* (pp. 304-317). Singapore: Springer Nature Singapore. https://doi.org/10.48550/arXiv.2506.21119.

Jraba, S., Elleuch, M., Ltifi, H., & Kherallah, M. (2023). Comparative analysis of CNNs and vision transformer models for brain tumor detection. *Proceedings of the 17th International Conference on Agents and Artificial Intelligence (ICAART): Vol. 3* (pp. 1432–1439). SCITEPRESS. https://doi.org/10.5220/0013381900003890

Kaifi, R. (2023). A review of recent advances in brain tumor diagnosis based on AI-based classification. *Diagnostics, 13*(18), 3007. https://doi.org/10.3390/diagnostics13183007

Kasiatun, S., & Sukraeny, N. (2024). Reduction of pain intensity in space occupying lesion (SOL) patients using murrotal therapy: A case study. *Menara Journal of Health Science, 3*(4), 586–597.

Kawadkar, K. (2025). Comparative Analysis of Vision Transformers and Convolutional Neural Networks for Medical Image Classification. DOI:2507.21156.

Kepakisan, I. K. S., Kesanda, I. M. P., & Putra, I. M. A. A. (2022). SPACE OCCUPYING LESION (SOL) CEREBRI. *Ganesha Medicina, 2*(1), 16–21. https://doi.org/10.23887/GM.V2I1.47295

Kumar, S., Esiowu, U., Asiamah, P., & Jolaoso, O. (2025). Enhancing image classification with augmentation: Data augmentation techniques for improved image classification. https://arxiv.org/abs/2502.18691

Kurniasari, D., Dwi Pratama, M., Junaidi, A., & Faisol, A. (2025). Transforming Alzheimer's disease diagnosis: Implementing Vision Transformer (ViT) for MRI images classification. *Journal of Information and Communication Technology, 24*(1), 130–152. https://doi.org/10.32890/jict2025.24.1.6

Kurniasari, D., Maulana, M. A., Lumbanraja, F. R., & Warsono. (2025). Leveraging transfer learning using convolutional neural networks to enhance brain tumor images classification. *International Journal of Computer Science*, 52(11), 4150-4163.

Lin, T., Wang, Y., Liu, X., & Qiu, X. (2022). A survey of transformers. *AI Open, 3*, 111–132. https://doi.org/10.1016/j.aiopen.2022.10.001

Markoulidakis, I., Rallis, I., Georgoulas, I., Kopsiaftis, G., Doulamis, A., & Doulamis, N. (2021), Multiclass confusion matrix reduction method, *Technologies*, 9(4), 81. doi: 10.3390/technologies9040081.

Mufidatuzzainiya, A., & Faisal, M. (2025). Penggunaan Teknik Transfer Learning pada Metode CNN untuk Pengenalan Tanaman Bunga. *JISKA (Jurnal Informatika Sunan Kalijaga), 10*(2), 195–206.

Pangestu, A., Purnama, B., & Risnandar, R. (2024). Vision Transformer untuk Klasifikasi Kematangan Pisang. *Jurnal Teknologi Informasi dan Ilmu Komputer, 11*(1), 75–84. https://doi.org/10.25126/jtiik.20241117389

Peng, J., Li, Y., & Zhang, L. (2024). Multi-class robust brain tumor segmentation with hybrid classification techniques. *Journal of Biomedical Imaging, 2024*, Article ID 102345.

Pradana, A. G., Setiadi, D. R. I. M., & Muslikh, A. R. (2024). Fine tuning model Convolutional Neural Network EfficientNet-B4 dengan augmentasi data untuk klasifikasi penyakit kakao. *Journal of Information System and Application Development, 2*(1), 1–11. https://doi.org/10.26905/jisad.v2i1.11899

Pratama, N., Liebenlito, M., & Irene, Y. (2024). Perbandingan Model Klasifikasi Transfer Learning Convolutional Neural Network Tumor Otak Menggunakan Citra Magnetic Resonance Imaging. *Jurnal Sehat Indonesia: JUSINDO, 6*(1), 308–318. https://doi.org/10.59141/jsi.v6i01.81

Priatna, W., Prasetyo, S. Y. J., Wijono, S., Maria, E., & Manongga, D. (2025). Deteksi Anomali dalam Penipuan E-commerce Menggunakan Hybrid Autoencoder-Transformer Frameworks. *JEPIN (Jurnal Edukasi dan Penelitian Informatika), 11*(1), 33–38.

Rahman, A. B., Islam, M. T., Islam, M. R., Sohrawordi, M., & Sultan, M. N. (2023, December). Enhanced Brain Tumor Classification from MRI Images Using Deep Learning Model. *2023 International Conference on Computer and Information Technology (ICCIT)*. https://doi.org/10.1109/ICCIT60459.2023.10441064

Ramadani, I., Hasibuan, M. S., Simatupang, K. N., Azali, I. (2026). Pendekatan Bayesian untuk prediksi permintaan produk makanan ringan UMKM dalam kondisi ketidakpastian. *Jurnal Publikasi Ilmu Komputer dan Multimedia*, 5(1), 107-116.

Rajinikanth, V., Kadry, S., Damasevicius, R., Sujitha, R. A., Balaji, G., & Mohammed, M. A. (2022). Detection in Brain MRI using Pre-trained Deep-Learning Scheme. *Proceedings of the 3rd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT)* (pp. 180-185).

Rodrigo, M., Cuevas, C., & García, N. (2024). Comprehensive comparison between vision transformers and convolutional neural networks for face recognition tasks. *Scientific Reports, 14*(1), 23392. https://doi.org/10.1038/s41598-024-73803-7

Salehi, A. W., Khan, S., Gupta, G., Alabduallah, B. I., Almjally, A., Alsolai, H., Siddiqui, T., & Mellit, A. (2023). A study of CNN and transfer learning in

medical imaging: Advantages, challenges, future scope. *Sustainability*, 15(7), 5930. https://doi.org/10.3390/su15075930

Sanhaji, G., Kurnia, D. P., & Hidayat, V. D. N. (2025). Implementasi algoritma convolutional neural network untuk klasifikasi dan deteksi tumor otak. *Jurnal Tekno Kompak, 19*(2), 1–13.

Sanjaya, I., Lelita, T., & Yustiana, I. (2025). Application of Vision Transformer for Identifying Indonesian Herbal Plants Based on Visual Images. *Jurnal Media Computer Science, 4*(2), 385–402. https://doi.org/10.37676/jmcs.v4i2

Sathyanarayanan S. & Tantri B. R., "Confusion matrix-based performance evaluation metrics," *African Journal of Biomedical Research*, vol. 27, no. 4s, pp. 4023–4031, 2024, doi: 10.53555/AJBR.v27i4S.4345.

Siregar, R. R., & Husein, A. M. (2024). Klasifikasi Tumor Otak Pada Gambar Magnetic Resonance Images (MRI) Dengan Pendekatan Pembelajaran Mendalam. *Data Sciences Indonesia, 4*(1), 162–168.

Sivakumar, M., Parthasarathy, S., & Padmapriya, T. (2024). Trade-off between training and testing ratio in machine learning for medical image processing. *PeerJ Computer Science, 10*, e2245. https://doi.org/10.7717/peerj-cs.2245

Song, B., KC, D. R., Yang, R. Y., Li, S., Zhang, C., & Liang, R. (2024). Classification of mobile-based oral cancer images using the vision transformer and the Swin transformer. *Cancers, 16*(5), 987. https://doi.org/10.3390/cancers16050987

Sumaiyah, S. N., Rahmi, L., & Roni, E. S. (2024). Tumor otak primer pada anak di RSUP Dr. M. Djamil Padang pada tahun 2019–2021. *Jurnal Sains, Teknologi dan Kesehatan, 1*(4), 186–199.

Tan, Z., Pan, T., Huang, K., Yu, W., Yao, K., Jiang, C., Wang, Q., Nguyen, A., Guo, X., Cheng, Y., & Yang, X. (2024). Exploiting layer normalization fine-tuning in visual transformer foundation models for classification. https://arxiv.org/abs/2508.07577v1

Turmuzi, A. (2025). Pendekatan deep learning untuk menciptakan pengalaman belajar yang bermakna. *Jurnal Pendidikan dan Teknologi*, 8(2), 1713-1719.

Ujianto, N. T., Gunawan, F., Fadilah, H., Fanti, A. P., Saputra, A. D., & Ramadhan, I. G. (2025). Penerapan algoritma K-Nearest Neighbors (KNN) untuk klasifikasi citra medis. *IT-EXPLORE, 2*(2), 33–43.

Wani, S., Ahuja, S., & Kumar, A. (2023). Application of Deep Neural Networks and Machine Learning algorithms for diagnosis of Brain tumour. *Proceedings of International Conference on Computational Intelligence and Sustainable Engineering Solution* (pp. 106–111). Institute of Electrical and Electronics Engineers Inc.

Winarto, M. N., Mailasari, M., & Purnamawati, A. (2022). Klasifikasi jenis tumor otak menggunakan arsitekture MobileNet V2. *Jurnal SIMETRIS, 13*(2), 239–248. https://doi.org/10.24176/simet.v13i2.7938

Yunisafitri, N. L. (2025). Tatalaksana Pemeriksaan MRI Brain dengan Kontras pada Kasus Lesi Hypophysis. *PREPOTIF, 9*(2), 3931–3936.

Yun, S., Han, D., Chun, S., Oh, S. J., Choe, J., & Yoo, Y. (2019). CutMix: Regularization strategy to train strong classifiers with localizable features. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)* (pp. 6022–6031). https://doi.org/10.1109/ICCV.2019.00612

Zeng, W., Zhu, H. L., Lin, C., & Xiao, Z. Y. (2023). A survey of generative adversarial networks and their application in text-to-image synthesis. *Electronic Research Archive, 31*(12), 7142–7181. https://doi.org/10.3934/era.2023362

Zeng, W. (2024). Image data augmentation techniques based on deep learning: A survey. *Mathematical Biosciences and Engineering, 21*(8), 6190–6224. https://doi.org/10.3934/mbe.2024272

Zhang, H. Y., Cisse, M., Dauphin, Y. N., & Lopez-Paz, D. (2018). Mixup: Beyond empirical risk minimization. DOI:1710.09412.

Zhang, W., Wang, H., & Liu, S. (2024). Enhanced MRI Tumor Segmentation and Bone Color Mapping for Clinical Diagnosis. *Scientific Reports, 14*(1), 7890. https://doi.org/10.1038/s41598-024-58264-8

Zhao, Z., Alzubaidi, L., Zhang, J., Duan, Y., & Gu, Y. (2024). A comparison review of transfer learning and self-supervised learning: Definitions, applications,

advantages and limitations. *Expert Systems with Applications, 242*, 122807. https://doi.org/10.1016/j.eswa.2023.122807

Zhou, T., Fu, C., Liu, Y., & Xiang, L. (2025). Groundwater level estimation using improved transformer model: A case study of the Yellow River Basin. *Water, 17*(15), Article 2318. https://doi.org/10.3390/w17152318