IMPLEMENTASI METODE *VISION TRANSFORMER* (VIT) DALAM KLASIFIKASI CITRA MRI PENYAKIT *ALZHEIMER*

(Skripsi)

Oleh

M. DWI PRATAMA

JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2024

ABSTRACT

IMPLEMENTATION OF VISION TRANSFORMER (VIT) METHOD IN MRI IMAGE CLASSIFICATION OF ALZHEIMER'S DISEASE

By

M. Dwi Pratama

Alzheimer's is a disease that causes a decline in cognitive brain function (dementia) with 10 million new cases growing every day. Early detection of Alzheimer's disease is needed to prevent dementia from getting worse. So there is a need for a technology that can diagnose Alzheimer's disease quickly and accurately. Alzheimer's can be classified into 4 classes non demented, very mild demented, mild demented, and moderate demented. Deep learning is a technology that is popular and often used in performing various machine learning tasks such as image classification. One of the methods in deep learning to perform image classification tasks is Vision Transformer (ViT). The advantage of ViT is the self-attention mechanism that can understand the relationship between parts of the data in an unlimited order. In this study, the classification of MRI images of Alzheimer's disease using a pre-trained ViT model was carried out. The research stages are data collection, data pre-processing, training, testing, and performance evaluation of the model and interpretation. The study used Alzheimer's disease MRI image data sourced from Kaggle with the results of the metric evaluation of model performance, namely accuracy, sensitivity, specificity, and F1-score in order 98.19%, 96.34%, 98.80%, and 96.37%. Based on the results obtained, it can be concluded that the pre-trained Vision Transformer model is very good at performing the task of classifying MRI images of Alzheimer's disease.

Keywords : Image Classification, ViT, Deep Learning

ABSTRAK

IMPLEMENTASI METODE *VISION TRANSFORMER* (VIT) DALAM KLASIFIKASI CITRA MRI PENYAKIT *ALZHEIMER*

Oleh

M. Dwi Pratama

Alzheimer merupakan penyakit penyebab penurunan fungsi kognitif otak (demensia) dengan pertumbuhan 10 juta kasus baru setiap hari. Pendeteksian dini penyakit *Alzheimer* diperlukan untuk mencegah demensia semakin parah. Sehingga diperlukannya suatu teknologi yang dapat melakukan diagnosis penyakit *Alzheimer* secara cepat dan akurat. *Alzheimer* secara umum dikategorikan menjadi 4 tahap penyakit yaitu *non demented*, *very mild demented*, *mild demented*, dan *moderate demented. Deep learning* merupakan teknologi yang sedang ramai dan biasanya banyak digunakan dalam berbagai tugas *machine learning* seperti klasifikasi citra. Salah satu metode dalam *deep learning* untuk melakukan tugas klasifikasi citra adalah *Vision Transformer* (ViT). Kelebihan dari ViT adalah mekanisme *self attention* yang mampu memahami hubungan antara bagian dari data dalam urutan yang tidak terbatas. Pada penelitian ini, dilakukan klasifikasi citra MRI penyakit *Alzheimer* menggunakan model *pre-trained* ViT. Tahapan penelitian yang dilakukan adalah pengumpulan data, *pre-processing* data, proses *training*, *testing*, dan evaluasi kinerja dari model dan interpretasi. Penelitian menggunakan data citra MRI penyakit *Alzheimer* yang bersumber dari Kaggle dengan hasil evaluasi metrik dari kinerja model yaitu akurasi, sensitivitas, spesifitas, dan *F1-score* secara berurutan adalah 98,19%, 96,34%, 98,80%, dan 96,37%. Berdasarkan hasil yang telah diperoleh, dapat disimpulkan bahwa model *pre-trained Vision Transformer* memiliki performa unggul dalam melakukan tugas klasifikasi citra MRI penyakit *Alzheimer*.

Kata Kunci : Klasifikasi Gambar, ViT, Deep Learning

IMPLEMENTASI METODE *VISION TRANSFORMER* (*VIT*) DALAM KLASIFIKASI CITRA MRI PENYAKIT *ALZHEIMER*

Oleh
M. DWI PRATAMA
2017031049

Skripsi

Sebagai Salah Satu Syarat untuk Mencapai Gelar
SARJANA MATEMATIKA

Pada

Jurusan Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Lampung

JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2024

Judul Skripsi	:	**IMPLEMENTASI METODE *VISION TRANSFORMER* (VIT) DALAM KLASIFIKASI CITRA MRI PENYAKIT *ALZHEIMER***
Nama Mahasiswa	:	**M. Dwi Pratama**
Nomor Pokok Mahasiswa	:	**2017031049**
Jurusan	:	**Matematika**
Fakultas	:	**Matematika dan Ilmu Pengetahuan Alam**

MENYETUJUI

1. Komisi Pembimbing

Dr. Dian Kurniasari, S.Si., M.Sc.
NIP. 196903051996032001

Dr. rer. nat. Akmal Junaidi, S.Si., M.Sc.
NIP. 197101291997021001

2. Ketua Jurusan Matematika

Dr. Aang Nuryaman, S.Si., M.Si.
NIP. 19740316 200501 1001

MENGESAHKAN

1. Tim Penguji

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

Sekretaris : **Dr. rer. nat. Akmal Junaidi, S.Si., M.Sc.** .....

Penguji
Bukan Pembimbing : **Dr. Ahmad Faisol. S.Si., M.Sc.** .....

2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam

Dr. Eng. Heri Satria, S.Si., M.Si.
NIP. 19711001 200501 1 002

Tanggal Lulus Ujian Skripsi: **13 Juni 2024**

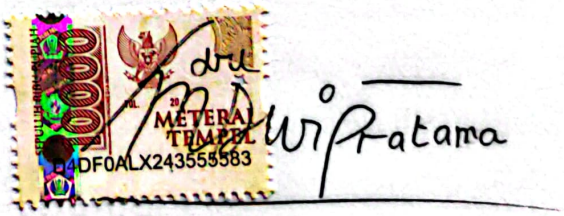
PERNYATAAN SKRIPSI MAHASISWA

Saya yang bertanda tangan di bawah ini:

Nama	:	**M. Dwi Pratama**
Nomor Pokok Mahasiswa	:	**2017031049**
Jurusan	:	**Matematika**
Judul Skripsi	:	**IMPLEMENTASI METODE *VISION TRANSFORMER* (VIT) DALAM KLASIFIKASI CITRA MRI PENYAKIT *ALZHEIMER***

Dengan ini menyatakan bahwa skripsi ini adalah hasil pekerjaan saya sendiri dan semua tulisan yang tertuang dalam skripsi ini telah mengikuti kaidah karya penulisan ilmiah Universitas Lampung.

Bandar Lampung, 13 Juni 2024
Penulis



M. Dwi Pratama
NPM. 2017031049

RIWAYAT HIDUP

Penulis memiliki nama lengkap M. Dwi Pratama lahir dan besar di Kota Bandar Lampung, provinsi Lampung pada tanggal 17 Januari 2002. Penulis adalah anak kedua dari pasangan Bapak Sam'un dan Ibu Sri Wati.

Penulis mengawali pendidikan di Pendidikan Anak Usia Dini (PAUD) Raudhatul Jinan 2007-2008. Kemudian menempuh Pendidikan Sekolah Dasar (SD) di SDN 1 Langkapura yang diselesaikan pada tahun 2014. Menyelesaikan pendidikan Sekolah Menegah Pertama (SMP) di SMPN 7 Bandar Lampung pada tahun 2017, dan melanjutkan pendidikan Sekolah Menengah Kejuruan (SMK) di SMKN 4 Bandar Lampung dengan mengambil jurusan Teknik Komputer Jaringan (TKJ) pada tahun 2017-2020.

Pada tahun 2020, penulis melanjutkan Pendidikan S1 di Jurusan Matematika FMIPA Unila melalui jalur SBMPTN dan berkesempatan untuk melanjutkan dunia perkuliahan dengan mendapatkan bantuan Pendidikan KIP-Kuliah. Penulis aktif mengikuti organisasi Koperasi Mahasiswa (KOPMA) Unila sebagai anggota Bidang 1 dan beberapa kali mengikuti kegiatan kepanitiaan serta ditunjuk sebagai Ketua Pelaksana kegiatan. Sebagai bentuk penerapan ilmu perkuliahan penulis melakukan Kerja Praktik (KP) di kantor asuransi AJB Bumiputera 1912 Kantor Wilayah Lampung. Penulis juga aktif dalam mengembangkan potensi diri dengan beberapa kali mengikuti kegiatan Studi Independen yang diselenggarakan pihak Kampus Merdeka yaitu mempelajari *Python for Data Science* oleh Hacktiv8, dan *Machine Learning* oleh Bangkit Academy led by Google, GoTo, dan Traveloka, dengan mendapat peringkat *Distinction Graduate.*

KATA INSPIRASI

مَنْ سَارَ عَلىَ الدَّرْبِ وَصَلَ

"*Barang siapa yang berjalan di jalannya niscaya ia akan sampai pada tujuan.*"

(Pepatah Arab)

"*Apa yang melewatkanku tidak akan pernah menjadi takdirku, dan apa yang ditakdirkan untukku tidak akan pernah melewatkanku."*

(Umar ibn Khattab)

"Your thoughts are the foundation of your power, with thoughts comes ideas, and with ideas comes realizations."

(mdpraz)

PERSEMBAHAN

Dengan mengucapkan rasa syukur atas segala puji dan ke hadirat Allah SWT yang telah melimpahkan segala karunia-Nya sehingga skripsi ini dapat diselesaikan dengan baik. Saya persembahkan karya sederhana ini untuk :

Diri Sendiri

Terima kasih karena telah berdedikasi tinggi, sabar, dan bertahan di setiap keadaan juga berusaha semaksimal mungkin menyelesaikan semuanya.

Orang Tua

Terima kasih atas setiap kasih sayang, pengorbanan, setiap keringat yang berjatuhan serta nasihat yang telah membentuk karakter diri penulis hingga dapat menjadi pribadi yang baik. Terima kasih selalu ada menjadi tempat pulang bagi penulis dalam setiap keadaan.

Dosen Pembimbing dan Pembahas

Terima kasih teruntuk para dosen pembimbing dan pembahas yang telah berjasa dalam membantu, menasihati, dan memberikan masukkan serta ilmu yang bermanfaat dalam penyelesaian skripsi ini.

Seluruh Keluargaku

Sahabat-sahabatku

Almamater Tercinta, Universitas Lampung

SANWACANA

Puji syukur kehadirat Allah SWT, atas segala rahmat dan karunia-Nya. Sholawat serta salam selalu tercurahkan kepada junjungan besar Nabi Muhammad SAW, sehingga penulis dapat menyelesaikan skripsi dengan judul "**Implementasi Metode *Vision Transformer* (ViT) dalam Klasifikasi Citra MRI Penyakit *Alzheimer***". Dalam menyelesaikan skripsi ini, penulis mendapat dukungan, bimbingan dan bantuan dari beberapa pihak. Oleh karena itu, penulis ingin mengucapkan terima kasih kepada :

1. Ibu Dr. Dian Kurniasari, S.Si., M.Sc., selaku Dosen Pembimbing I yang selalu memberikan arahan, bimbingan, motivasi, serta saran yang mendukung sehingga penulis dapat menyelesaikan skripsi ini.
2. Bapak Dr. rer. nat. Akmal Junaidi, S.Si., M.Sc., selaku Dosen Pembimbing II atas bimbingan dan saran kepada penulis selama proses penyusunan skripsi ini.
3. Bapak Dr. Ahmad Faisol, S.Si., M.Sc., selaku Dosen Pembahas yang telah memberikan kritik dan saran yang membangun proses penyusunan skripsi.
4. Bapak Dr. Aang Nuryaman, S.Si., M.Si., selaku Ketua Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
5. Bapak Dr. Eng. Heri Satria, S.Si., M.Si., selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
6. Bapak dan Ibuku yang tak henti-hentinya mencurahkan doa, ridho, kasih sayang, dan cinta untuk penulis hingga saat ini.
7. Saudara kandungku tersayang dan keluarga besar yang selalu mendukung penulis dalam menempuh pendidikan.
8. Syifa Zalfa yang selalu menemani, membantu, dan memberikan dukungan kepada penulis dalam menyelesaikan skripsi ini.

9. Teman-teman terbaikku anggota Cisulum 15+, teman-teman bimbingan skripsi, teman-teman Osis dan semasa sekolah yang selalu memberikan motivasi.
10. Seluruh pihak yang tidak bisa penulis sebutkan satu per satu yang telah membantu penulis dalam menyelesaikan skripsi ini.

Penulis menyadari bahwa skripsi ini masih jauh dari kata sempurna dan masih terdapat banyak kekurangan baik dalam penyajian maupun teknik penulisan. Oleh sebab itu, saran dan kritikan yang membangun senantiasa penulis harapkan agar dapat digunakan sebagai bahan perbaikan kedepannya.

Bandar Lampung, 13 Juni 2024
Penulis

M. Dwi Pratama
NPM. 2017031049

DAFTAR ISI

Halaman

DAFTAR TABEL xiv
DAFTAR GAMBAR xvi
DAFTAR KODE PROGRAM xvii
I. PENDAHULUAN 1
1.1 Latar Belakang dan Masalah 1
1.2 Rumusan Masalah 5
1.3 Tujuan Penelitian 5
1.4 Manfaat Penelitian 5
II. TINJAUAN PUSTAKA 6
2.1 Penelitian Terkait 6
2.2 *Alzheimer* 9
2.3 Citra Digital 11
2.4 Klasifikasi dalam *Machine Learning* 14
2.5 Augmentasi 15
2.6 *Artificial Neural Network* 15
2.7 *Deep Learning* 17
2.8 *Vision Transformer* 19
2.8.1 *Patch Embeddings* 20
2.8.2 *Positional Embeddings* 21
2.8.3 *Transformer Encoder* 22
2.8.4 *Self Attention* 23
2.8.5 *Multi Head Attention* 24
2.8.6 Lapisan *Normalization* 26
2.8.7 *Multilayer Perceptron Head* (MLP Head) 27
2.8.8 Fungsi Aktivasi 27
2.8.9 *Optimizer : Adam Optimizer* 30

2.9 *Confusion Matrix* 30
III. METODE PENELITIAN 32
3.1 Waktu dan Tempat Kerja Praktik 32
3.2 Alat Penelitian 33
3.3 Data Penelitian 34
3.4 Metode Penelitian 34
IV. HASIL DAN PEMBAHASAN 36
4.1 Proses Input Data 36
4.2 *Pre-processing* Data 37
4.2.1 *Resize* Data 38
4.2.2 Normalisasi Data 38
4.2.3 Augmentasi Data 39
4.2.4 Transformasi Gambar Menjadi *Tensor* 43
4.3 Pembuatan *DataLoader* 44
4.4 Komponen Dalam Arsitektur *Vision Transformer* (ViT) 44
4.4.1 Contoh Operasi Manual *Patch Embeddings* 45
4.4.2 Contoh Operasi Manual *Positional Embeddings* 49
4.5 *Pre-trained* Model ViT 52
4.6 Proses *Training* Model 54
4.7 Perbandingan Hasil *Training* dari Beberapa Parameter Berbeda 57
4.8 Perbandingan Hasil *Training* dari Parameter yang Digunakan dengan Proporsi Data Berbeda 58
4.9 Proses *Testing* Model dengan Parameter Terbaik 59
4.10 Evaluasi Model 62
4.11 *Misprediction* Target Klasifikasi 68
4.12 Interpretasi Hasil dan Perbandingan Dengan Penelitian Terdahulu 69
V. KESIMPULAN DAN SARAN 73
5.1 Kesimpulan 73
5.2 Saran 74
DAFTAR PUSTAKA 75
LAMPIRAN 81

DAFTAR TABEL

Tabel Halaman

1. Penelitian Terkait 6
2. Intensitas Citra *Grayscale* 13
3. Kategori Nilai Kinerja Model Arsitektur 31
4. Perbandingan Hasil *Training* Antara Beberapa Kombinasi Parameter 57
5. *Benchmark* antara Hasil Penelitian dengan Penelitian Rujukan 70

DAFTAR GAMBAR

Gambar Halaman

1. MRI *Alzheimer Very Mild Demented* 9
2. MRI *Alzheimer Mild Demented* 10
3. MRI *Alzheimer Moderate Demented* 10
4. MRI *Alzheimer Non Demented* 11
5. Ilustrasi Transformasi Citra Digital 13
6. Ilustrasi Teknik Augmentasi 15
7. Ilustrasi Neuron Pada *Artificial Neural Network* 16
8. Ilustrasi Representasi *Deep Learning* 18
9. Model *Vision Transformer* 19
10. *Transformer Encoder* 22
11. *Self Attention* 23
12. *Multi Head Attention* 25
13. Ilustrasi (a) Lapisan *Post-LN Layer*; (b) Lapisan *Pre-LN Layer* 26
14. Fungsi Aktivasi *Sigmoid* 28
15. Fungsi Aktivasi ReLU 28
16. Fungsi Aktivasi *Softmax* 29
17. Fungsi Aktivasi GELU 29

18. Ilustrasi *Confusion Matrix* 30
19. *Flowchart* Metode Penelitian 35
20. Hasil *Splitting* Data 37
21. Ilustrasi *RandomResizedCrop* 40
22. Ilustrasi *RandomHorizontalFlip* dan *RandomVerticalFlip* 40
23. Ilustrasi *RandomRotation* 41
24. Ilustrasi *ColorJitter* 41
25. Komponen Arsitektur *Vision Transformer* (ViT) 45
26. Ilustrasi *Patching* 46
27. Proses *Flat Patching* 47
28. *Splitting* Citra MRI *Alzheimer* 54
29. Hasil *Training* 55
30. Grafik *Loss* Data *Training* dan Validasi 56
31. Grafik Akurasi Data *Training* dan Validasi 56
32. Ilustrasi (a) Proporsi Data 1; (b) Proporsi Data 2 59
33. *Confusion Matrix* Prediksi Penyakit *Alzheimer* 60
34. *Misprediction* Label Kelas *NonDemented* dan *VeryMildDemented* 69

DAFTAR KODE PROGRAM

Kode Program Halaman

1. *Input* Data 36
2. Normalisasi Data 39
3. *Syntax* Augmentasi dengan *PyTorch* 42
4. Proses *Loader* Data *Training* 44
5. *Syntax* Pembentukan Model *Pre-trained* ViT 53

I. PENDAHULUAN

1.1 Latar Belakang dan Masalah

Alzheimer merupakan suatu gangguan *neurodegeneratif b*erupa penyusutan volume *hippocampus* dan korteks serebral otak menyebabkan terjadinya kondisi demensia seperti menurunnya kemampuan dalam pemecahan masalah, penurunan daya ingat, dan kesulitan berbicara pada seseorang (Almumtazah dkk., 2023). Penyakit ini pertama kali ditemukan oleh seorang ahli Psikiatri asal Jerman yaitu Alois Alzheimer pada tahun 1907. Ia mengawali penelitian tersebut ketika mengamati aktivitas dari seorang wanita berumur 51 tahun, yang mengalami penurunan fungsi kognitif otak seperti kemampuan berpikir dan gangguan pada daya ingatnya. Dengan diagnosis akhir bahwa wanita tersebut mengidap penyakit *Alzheimer* berdasarkan hasil pemeriksaan *neuropatologi* otak. Terdapat beberapa kondisi atau tanda penurunan fungsi kognitif yang dapat menjadi suatu gejala bahwa seseorang mengidap penyakit *Alzheimer* ini diantaranya yaitu kesulitan dalam berbicara, penurunan kemampuan dalam perencanaan sesuatu, terjadi perubahan pola perilaku sehari-hari, depresi, dan kesulitan dalam mengendalikan emosi (Sianturi, 2021). Selain itu, terdapat beberapa faktor yang meningkatkan risiko seseorang terjangkit penyakit ini yaitu kurangnya aktivitas fisik, kebiasaan merokok, obesitas, diabetes, malnutrisi, cedera otak dan usia (Alzheimer's Association, 2023).

Menurut data *World Heath Organization* (WHO) pada tahun 2017 terjadi peningkatan yang dinamis diseluruh dunia pada penyakit *Alzheimer* dengan perkiraan sekitar 46,8 atau 50 juta didiagnosis dengan demensia *Alzheimer* dan total pertumbuhan 10 juta kasus baru setiap hari. Indonesia sendiri, jumlah orang yang

menderita penyakit ini pada tahun 2016 mencapai 1,2 juta kasus, dan jumlah ini diperkirakan akan terus meningkat menjadi 2 juta kasus pada tahun 2030. *Alzheimer* seringkali sulit dideteksi, terutama pada tahap awal penyakit. Penyakit ini sendiri dapat didiagnosis dengan menggunakan fitur citra *Magnetic Resonance Imaging* (MRI) oleh seorang ahli radiologi (Almumtazah dkk., 2023). Namun sayangnya, menurut Humas Fakultas Kedokteran UI pada tahun 2021 jumlah ahli radiologi yang ada di Indonesia hanya terdapat 1646 orang yang melayani lebih dari 270 juta jiwa penduduk Indonesia. Jumlah tersebut tentunya sangat sedikit, terutama mengingat perkembangan ilmu pengetahuan dan teknologi yang sangat pesat belakangan ini. Oleh karena itu, diperlukannya suatu inovasi yang menggunakan kecerdasan buatan dalam membantu proses diagnosis penyakit *Alzheimer*, memungkinkan diagnosis dilakukan lebih dini dan mempercepat proses penanganan (Fabrizio dkk., 2021).

Untuk mengklasifikasikan penyakit *Alzheimer* dapat dilakukan dengan MRI. Menurut Setiyoningsih dan Apriantoro (2023), MRI adalah alat kedokteran dalam bidang pemeriksaan diagnostik radiologi yang menggunakan medan magnet berkekuatan tinggi dan resonansi getaran terhadap inti atom hidrogen. Secara umum, MRI adalah teknik pencitraan yang menghasilkan gambar bagian tubuh manusia secara *transversal*, *koronal*, *sagittal*, dan *pra aksial* tanpa menggunakan sinar X. (Astuti, 2019). Pemindaian MRI digunakan dalam penelitian *Alzheimer* karena MRI memiliki kontras yang tinggi, resolusi spasial yang baik, dan aksesibilitas yang tinggi (Khan dkk., 2019). Menurut Sailaja dan Gopal (2023), pada penyakit *Alzheimer* dikelompokkan menjadi 4 kategori berdasarkan tingkat kondisi pengidap yaitu *nondemented, very mild demented, mild demented,* dan *moderate demented*.

Perkembangan teknologi yang saat ini sudah sangat maju dan mendukung komputasi untuk berjalan dengan cepat sehingga penerapan di bidang medis sudah banyak dilakukan contohnya seperti diagnosis suatu penyakit. Teknik yang dapat digunakan untuk melakukan diagnosis suatu penyakit gangguan neurologis seperti *Alzheimer* adalah *deep learning* (Gautam dan Sharma, 2020). Diagnosis dilakukan

dengan cara melakukan klasifikasi atau penggolongan suatu objek dengan ciri-ciri yang sama atau berbeda. Dan metode yang saat ini sedang ramai diperbincangkan yaitu bidang *computer vision. Computer vision* adalah teknologi yang memungkinkan komputer untuk memahami dan menganalisis gambar serta video (Guntara, 2022). Adapun metode yang saat ini sering digunakan untuk proses klasifikasi terutama pada data gambar yaitu menggunakan algoritma *Convolutional Neural Network* (CNN) dan juga *Vision Transformer* (ViT).

Convolutional Neural Network (CNN) merupakan algoritma yang telah digunakan secara luas dalam melakukan proses ekstraksi fitur pada data citra dan mencapai kinerja yang tinggi dalam tugas klasifikasi dan deteksi objek (Alwanda dkk., 2020). Namun, CNN dalam melakukan tugasnya seringkali membutuhkan jumlah parameter dan waktu pelatihan yang lama untuk mencapai kinerja yang maksimal (Ilahiyah dan Nilogiri, 2018). Menurut Kim dkk. (2020), CNN memiliki keterbatasan lain dalam mendeteksi suatu pola spesifik di lokasi tertentu dan untuk mengatasi permasalahan ini dapat menerapkan teknik *patching* pada arsitektur CNN. Teknik *patching* memungkinkan CNN untuk memahami detail gambar melalui fitur-fitur kecil dalam gambar seperti tepi, tekstur, atau pola spesifik di lokasi tertentu. Selain itu, teknik ini harus dilakukan secara tepat agar dapat memenuhi kebutuhan dari arsitektur CNN.

Algoritma lain yang sering digunakan untuk melakukan tugas *computer vision* adalah *Vision Transformer* (ViT). Model ViT adalah suatu metode dengan pendekatan *transformer* yang sebelumnya sudah terkenal pada pengolahan metode bahasa alami (NLP) kemudian diimplementasikan ke dalam tugas pengolahan data gambar (Dosovitskiy dkk., 2020). Model ViT memiliki teknik khusus dalam melakukan *patching* sehingga arsitektur ini tidak perlu lagi melakukan *patching* secara terpisah. Pada ViT, terdapat komponen kunci dalam arsitektur *transformer* yaitu *self-attention* yang memungkinkan untuk memberikan bobot dan perhatian pada bagian yang berbeda dari input kemudian dikombinasikan dengan *patch* dan *positional embeddings* (Dosovitskiy dkk., 2020). Kelebihan dari metode ini adalah arsitektur tidak menggunakan konvolusi seperti CNN melainkan prinsip *attention*

sehingga dapat mengurangi jumlah parameter yang digunakan (Heo dkk., 2021). Penelitian terkait yang menggunakan arsitektur ViT dan klasifikasi penyakit *Alzheimer* telah dilakukan sebelumnya, seperti penelitian yang dilakukan oleh Almumtazah dkk. (2023), pada penelitian tersebut data yang digunakan yaitu data citra MRI *Alzheimer* dengan proses segmentasi fitur data menggunakan *Fuzzy C-Means* (FCM) dan ekstrasi dilakukan menggunakan *Adaptive Neuro Fuzzy Inference Syste*m (ANFIS) mendapatkan akurasi sebesar 88,60%, spesifitas sebesar 87,15%, dan sensitifitas 90,27%.

Penelitian kedua dilakukan oleh Ghazal dkk. (2022), menggunakan *Transfer Learning* dengan model AlexNet untuk klasifikasi citra penyakit *Alzheimer* didapatkan hasil akurasi sebesar 91,70%, sensitivitas sebesar 93,70% dan nilai spesifitas sebesar 96%. Penelitian lain yang dilakukan oleh Wu dkk. (2021), dengan menggunakan arsitektur ViT pada pengenalan penyakit diabetes didapatkan hasil yang baik dengan akurasi 91,40%, spesifitas 97,70% dan sensitivitas 92,60%.

Penelitian keempat dilakukan Figo dkk. (2023) dalam mendeteksi penyakit Covid-19 dari citra *X-ray* paru-paru manusia menggunakan arsitektur ViT, didapatkan hasil akurasi sebesar 90,60%. Penelitian selanjutnya dilakukan oleh Shin dkk. (2023), membandingkan antara dua metode *computer vision* yaitu ViT dengan VGG19 untuk klasifikasi *penyakit Alzheimer* secara berurutan yaitu akurasi, *precision*, *recall*, dan *F1-Score* menggunakan ViT yaitu 80%, 60%, 75%, dan 66,67%, sedangkan untuk metode VGG19 yaitu 73%, 60%, 60%, dan 60%.

Berdasarkan keterbatasan yang dimiliki arsitektur CNN untuk mengklasifikasi penyakit *Alzheimer* oleh karena itu, dalam tugas akhir ini akan dilakukan implementasi arsitektur ViT dalam melakukan tugas klasifikasi *multi-class* pada citra MRI penyakit *Alzheimer*. Adapun eksperimen yang dilakukan yaitu menggunakan model *pre-trained* pada arsitektur ViT, serta untuk meningkatkan peforma dari arsitektur akan dilakukan proses augmentasi data pada tahap *pre-processing*. Kinerja dari aristektur tersebut akan dievaluasi berdasarkan dari nilai akurasi, sensitivitas, spesifitas dan *F1-score*.

1.2 Rumusan Masalah

Rumusan masalah dari penelitian ini adalah sebagai berikut :

1. Bagaimana membangun arsitektur ViT yang efektif untuk mengklasifikasikan citra data MRI penyakit *Alzheimer* ?
2. Bagaimana kinerja dari arsitektur ViT dalam melakukan klasifikasi citra data MRI penyakit *Alzheimer* ?

1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut :

1. Membangun arsitektur ViT untuk mengklasifikasikan citra data MRI penyakit *Alzheimer*.
2. Mengevaluasi kinerja dari arsitektur ViT dengan melihat nilai akurasi, sensitivitas, spesifitas, dan *F1-score* dalam melakukan klasifikasi citra data MRI penyakit *Alzheimer*.

1.4 Manfaat Penelitian

Terdapat manfaat dari penelitian ini adalah sebagai berikut :

1. Dapat digunakan sebagai alternatif dalam bidang medis terutama ahli radiologi untuk diagnosis awal penyakit *Alzheimer* berdasarkan hasil MRI otak.
2. Dapat menjadi referensi bagi peneliti di masa depan dalam melakukan penelitian menggunakan *deep learning* pada data yang berbeda di bidang *computer vision*.

II. TINJAUAN PUSTAKA

2.1 Penelitian Terkait

Penelitian yang dilakukan ini penulis menggunakan beberapa penelitian lain yang terkait sebagai referensi dan acuan untuk memperdalam pemahaman terhadap topik penelitian yang diimplementasikan dalam penelitian ini. Berikut merupakan ringkasan dari beberapa penelitian dapat dijelaskan pada Tabel 1.

Tabel 1. Penelitian Terkait

No	Judul Penelitian	Metode	Hasil			
			Acc	*Spec*	*Sen*	*F1*
1.	Vision Transformer Approach for Classification of Alzheimer's Disease Using 18F-Florbetaben Brain Images (Shin dkk., 2023)	Metode Klasifikasi Gambar menggunakan Vision Transformer	80%	-	-	66,6%
2.	Klasifikasi Alzheimer Berdasarkan Data Citra MRI Otak Menggunakan FCM dan ANFIS (Almumtazah dkk., 2023)	Metode Klasifikasi Gambar menggunakan metode FCM dan ANFIS	88,6%	87,1%	90,2%	-
3.	Deteksi Covid-19 dari Citra X-	Metode Klasifikasi	90,6%	-	-	-

	ray menggunakan Vision Transformer (Figo dkk., 2023)	Gambar menggunakan Vision Transformer (Base Large-16)				
4.	Vision Transformer-Based Recognition Of Diabetic Retinopathy Grade (Wu dkk., 2023)	Metode Klasifikasi Gambar Menggunakan ViT	91,4%	97,7%	92,6%	-
5.	Alzheimer Disease Detection Empowered with Transfer Learning AlexNet (Ghazal dkk., 2022)	Metode Klasifikasi Gambar Menggunakan Transfer Learning AlexNet	91,7%	93,7%	96%	-

Beberapa penelitian yang dijadikan bahan referensi dalam melakukan penelitian ini, yaitu penelitian yang dilakukan oleh Shin dkk. (2023), membahas tentang klasifikasi citra penyakit *Alzheimer* menggunakan metode *Vision Transformer*. Data yang digunakan yaitu data citra MRI *Alzheimer* dengan jumlah total 383 yang terbagi menjadi tiga kategori yaitu *alzheimer disease, mild cognitive,* dan *healthy control.* Nilai akurasi diperoleh sebesar 80% dan *F1-Score* sebesar 66%, dikatakan bahwa kinerja klasifikasi tidak ideal mungkin karna *overfitting* dan bias yang kurang tepat karena keterbatasan data gambar PET.

Kemudian penelitian yang dilakukan oleh Almumtazah dkk. (2023), membahas terkait klasifikasi citra penyakit *Alzheimer* menggunakan metode FCM dan ANFIS. Data yang digunakan merupakan citra MRI *Alzheimer* dengan total sebanyak 1000 citra, yang terbagi menjadi kategori *alzheimer* dan *non-alzheimer.* Penelitian tersebut diperoleh hasil akurasi sebesar 88,6%, spesifitas sebesar 87,1% dan sensitifitas sebesar 90,2%, dan dapat disimpulkan dari hasil tersebut model yang telah dibuat dengan segmentasi FCM dan klasifikasi ANFIS telah mendapatkan

kinerja akurasi yang tinggi pada uji coba menggunakan *k-fold cross validation* dengan nilai *k = 5* dan *epochs* sebesar 50.

Penelitian selanjutnya dilakukan oleh Figo dkk. (2023), membahas terkait dengan pendeteksian penyakit Covid-19 dari citra *X-ray* menggunakan metode *Vision Transformer*. Data yang digunakan pada penelitian ini yaitu data CT scan paru-paru dengan total sebanyak 30.530 citra, yang terbagi menjadi kategori normal, *pneumonia*, dan covid-19. Penelitian tersebut didapatkan hasil akurasi sebesar 90,6% dari arsitektur *vision transformer large-16.* Namun, terdapat *overfitting* pada model yang dibangun tersebut ditandakan dengan perbedaan yang signifikan antara nilai akurasi data validasi yaitu 90,6% dan 75,9% pada data *testing*.

Penelitian lainnya yang dijadikan referensi dilakukan oleh Wu dkk. (2021), membahas terkait dengan pendeteksian *diabetic retinopathy* dengan melakukan klasifikasi menggunakan metode *Vision Transformer* dengan data yang digunakan sebanyak 35.126 dan terbagi menjadi beberapa kategori yaitu *no diabetic retinopathy, mild, moderate, severe,* dan *proliferative diabetic retinopathy.* Penelitian tersebut didapatkan hasil akurasi sebesar 91,4%, spesifitas 97,7%, dan sensitivitas sebesar 92,6% dan dapat dikatakan bahwa model yang dibangun sudah memprediksi penyakit *diabetic retinopathy* dengan baik.

Penelitian terakhir yaitu dilakukan oleh Ghazal dkk. (2022), membahas terkait dengan pendeteksian penyakit *Alzheimer* dengan melakukan klasifikasi citra menggunakan metode *transfer learning* (AlexNet). Data yang digunakan pada penelitian ini yaitu data citra MRI penyakit *Alzheimer* yang berjumlah sebanyak 6.393 citra dan terbagi menjadi beberapa kategori yaitu *non demented, very mild demented, mild demented,* dan *moderate demented.* Pada penelitian tersebut mendapatkan kinerja akurasi, sensitivitas, dan spesifitas secara berurutan yaitu 91,7%, 93,7, 96% dan dapat dikatakan bahwa model yang dibangun sudah baik dalam memprediksi penyakit *Alzheimer.*

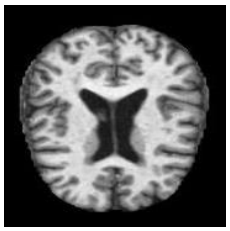
2.2 *Alzheimer*

Alzheimer bukan suatu hal yang normal ketika terjadi proses penuaan. Meskipun sebagian besar penderita *Alzheimer* merupakan orang yang berusia 65 tahun keatas, namun penyakit ini dianggap sebagai penyakit *Alzheimer* ketika menyerang seorang yang berusia di bawah 65 tahun (Alzheimer's Association, 2023). Menurut *Alzheimer's Association* biasanya orang yang menderita penyakit ini dapat berada di tahap awal, tengah atau akhir dari penyakit ini. *Alzheimer* adalah penyakit yang menyebabkan terjadinya degeneratif otak dan penyebab utama dari penyakit demensia, istilah umum ini mengacu pada tanda yang dialami oleh pengidap penyakit tersebut seperti penurunan memori, keterampilan dalam pemecahan masalah, kesulitan dalam mengendalikan emosi, dan penurunan kemampuan untuk melakukan kegiatan sehari-hari (Sianturi, 2021).

Bersumber dari penelitian terkait dengan *neuroimaging*, pada studi demensia atau gangguan kognitif data citra MRI penyakit *Alzheimer* umumnya terbagi menjadi 4 kelas yang menandakan tingkat keparahan penyusutan volume otak (Sailaja dan Gopal, 2023) yaitu :

1. *Very Mild Demented*

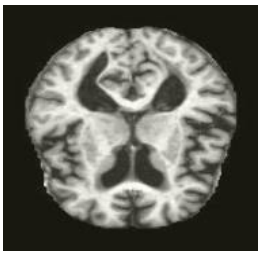
   Very Mild Demented merupakan jenis gejala paling awal dalam tingkat penyakit demensia *Alzheimer*. Pada tahap ini kebanyakan orang dapat menjalankan berbagai aktivitasnya seperti biasa tanpa mengalami banyak kendala yang berkaitan dengan fungsi kognitif otak, gejala ringan yang dialami pada tahap ini sepeeri kesulitan mengingat barang yang baru saja diletakkan (Alzheimer's Association, 2023). Gambar 1 menunjukkan kelas *very mild demented* pada penyakit *Alzheimer*.



Gambar 1. MRI *Alzheimer Very Mild Demented*

2. *Mild Demented*

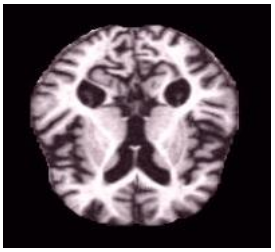
Mild Demented merupakan jenis penyakit demensia ringan, pada tingkat ini umumnya seseorang sudah mulai mengalami kesulitan ketika mengingat informasi, kesulitan dalam menemukan kata-kata yang tepat dalam berbicara, penurunan kemampuan dalam perencanaan sesuatu, dan mulai terjadi perubahan perilaku (Alzheimer's Association, 2023). Gambar 2 menunjukkan kelas *mild demented* pada penyakit *Alzheimer*.



Gambar 2. MRI *Alzheimer Mild Demented*

3. *Moderate Demented*

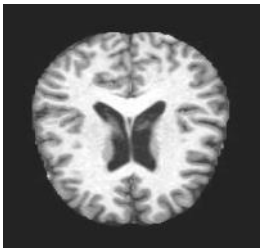
Moderate Demented merupakan jenis penyakit demensia sedang, gejala yang biasanya terjadi ketika sampai di tahap ini yaitu penurunan fungsi otak lebih signifikan dan mengalami lebih banyak masalah yang mempengaruhi kegiatan sehari-hari seseorang, seperti kesulitan dalam berbicara, kesulitan menyelesaikan tugas seperti mandi dan berpakaian, perubahan emosional yang signifikan, gangguan tidur, dan kesulitan ketika mengerjakan tugas-tugas yang sederhana (Alzheimer's Association, 2023). Gambar 3 menunjukkan kelas dari *moderate demented* pada penyakit *Alzheimer*.



Gambar 3. MRI *Alzheimer Moderate Demented*

4. *Non Demented*

Non Demented mengacu pada individu yang tidak menunjukan gejala-gejala serta gangguan kognitif seperti penyakit *Alzheimer* (Sianturi, 2021). Mereka memiliki kognitif serta fungsi otak yang normal sesuai dengan usia mereka. Beberapa karakteristiknya yaitu fungsi otak yang cendrung stabil, tidak mengalami kendala dalam mengerjakan aktivitas sehari-hari, dan tidak ada perubahan perilaku secara signifikan (Alzheimer's Association, 2023). Gambar 4 menunjukkan kelas *non demented* pada penyakit *Alzheimer*.



Gambar 4. MRI *Alzheimer Non Demented*

2.3 Citra Digital

Citra adalah kumpulan dari elemen-elemen gambar yang merekam suatu objek atau adegan melalui indera visual (Sumijan dan Purnama, 2021). Pada teknologi digital citra tidak dapat langsung diproses dan dianalisis oleh komputer, citra harus direpresentasikan menjadi *pixel* yang memuat informasi warna kemudian membentuk suatu imitasi objek dan tekstur tertentu (Kirana, 2021). Menurut Fitriyah dan Wihandika (2021), citra digital merupakan data dua dimensi dimana satuan terkecilnya yaitu *pixel* yang memiliki informasi terkait warna dan tata letak. Biasanya secara umum proses tahap pengolahan citra digital yaitu konversi warna citra, peningkatan citra, segmentasi, representasi dan *filtering* (Fitriyah dan Wihandika, 2021).

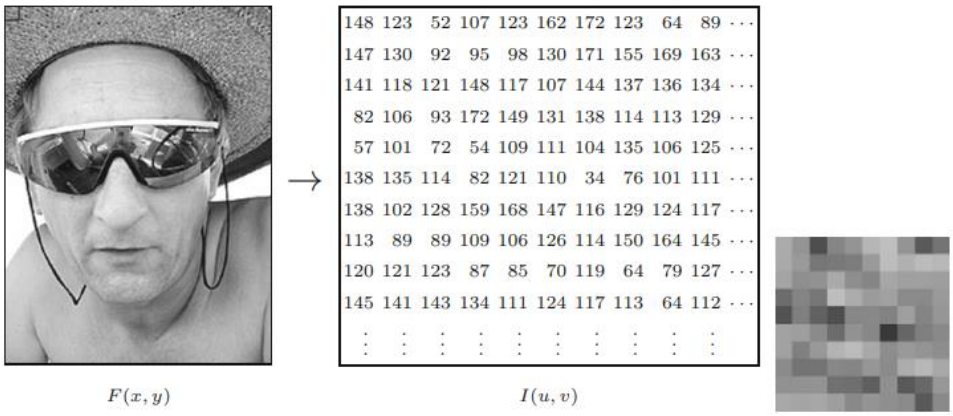
Pada awalnya, pengolahan citra digital hanya digunakan untuk mengubah gambar analog menjadi digital serta meningkatkan kualitas dari gambar, tetapi seiring waktu dan perkembangan zaman alat pendukung serta aplikasi pengolahan citra

menjadi lebih beragam. Pada saat ini didukung oleh kecepatan proses komputer yang meningkat pesat memungkinkan pengolahan dilakukan secara realtime dan membuat hasil citra analog dikodekan menjadi citra warna digital yang membuatnya mendekati warna aslinya. Menurut Asmara (2018), tipe-tipe citra yang sering digunakan dalam penelitian terbagi menjadi beberapa antara lain citra *biner*, citra *grayscale*, dan citra warna.

Citra *biner* adalah citra yang memiliki 2 warna saja yaitu hitam dan putih, serta hanya membutuhkan 1-bit memori untuk merepresentasikan warnanya (Asmara, 2018). Berbeda dengan citra *biner*, citra *grayscale* merupakan matriks data yang merepresentasikan setiap warna *pixel* dengan nilai berkisar antara 0 hingga 255, dengan setiap *pixel* membutuhkan 8-bit memori. Pada citra warna setiap *pixel*-nya memiliki tiga elemen warna yaitu merah, hijau dan biru atau sering disebut dengan RGB. Format file grafis menyimpan citra warna sebagai citra dengan representasi 24-bit, yang berasal dari komponen RGB sehingga memungkinkan citra mempunyai 16 juta kemungkinan warna (Mandala dkk., 2021). Proses transformasi citra digital direpresentasikan sebagai *array* (matrix) yang terdiri dari nilai-nilai numerik $F(x, y)$. Hal tersebut merupakan bentuk representasi yang digunakan dalam pemrosesan di komputer, dalam bentuk persamaan representasi tersebut dijelaskan oleh matriks $F_{m\times n}$ Persamaan (2.1) (Gonzales dan Woods, 2018).

$$F(x,y) = \begin{bmatrix} F(0,0) & F(0,1) & \dots & F(0,n-1) \\ F(1,0) & F(1,1) & \dots & F(1,n-1) \\ \vdots & \vdots & & \vdots \\ F(m-1,0) & F(m-1,1) & \dots & F(m-1,n-1) \end{bmatrix} \quad (2.1)$$

Berikut ini contoh dari citra *grayscale* yang mengilustrasikan proses transformasi dari citra *grayscale* kontinu menjadi citra digital berbentuk diskrit, dijelaskan pada Gambar 5 (Burger dan Burge, 2016).



Gambar 5. Ilustrasi Transformasi Citra Digital (Burger dan Burge, 2016)

Dari ilustrasi di atas menjelaskan terkait dengan transformasi dari sebuah citra *grayscale F*(*x*,*y*) menjadi bentuk citra digital diskrit *I*(*u*,*v*). Dengan mengasumsikan gambar persegi panjang dengan ukuran sebuah gambar ditentukan secara langsung dari lebar *M* (jumlah kolom) dan tinggi *N* (jumlah baris) dari matriks citra *I* (Burger dan Burge, 2016). Informasi dari setiap warna tersebut direpresentasikan dari nilai *pixel* yang secara praktis selalu berupa nilai biner dengan panjang *k*, menurut (Burger dan Burge, 2022) nilai *k* tersebut merupakan kedalaman bit yang menginformasikan tata letak dan intensitas yang tepat dari setiap *pixel*. Nilai dari tingkat kedalaman intensitas *k* dari citra *grayscale* berdasarkan tipe dan kegunaannya dapat dilihat pada Tabel 2.

Tabel 2. Intensitas Citra *Grayscale*

Kernel	*Bits/Pixel*	Rentang	Digunakan
1	1	[0, 1]	Gambar *Biner* : dokumen, ilustrasi, fax
1	8	[0, 255]	Umum : poto, scan, print
1	12	[0, 4095]	Kualitas Tinggi : poto, scan, print
1	14	[0, 16383]	Profesional: photo, scan, print
1	16	[0, 65535]	Kualitas Tertinggi : obat-obatan, astronomi

2.4 Klasifikasi dalam *Machine Learning*

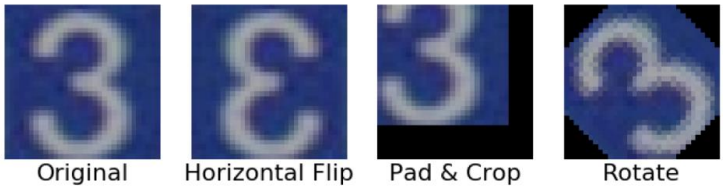
Di dalam konteks *machine learning* sendiri, teknik klasifikasi merujuk pada penggunaan algoritma pembelajaran mesin untuk menggolongkan atau mengategorikan data ke dalam kelas atau kelompok tertentu berdasarkan fitur atau atribut tertentu (Bachtiar dkk., 2019). Tujuan dari klasifikasi sendiri yaitu untuk memprediksi kelas ataupun label dari suatu objek yang belum diketahui dengan membuat suatu model atau fungsi yang bisa membedakan kelas data (Ninditama dkk., 2021).

Bidang *machine learning* secara umum dapat dikelompokkan ke dalam tiga kategori utama yaitu *supervised learning*, *unsupervised learning*, dan *reinforcement learning* (Abijono dkk., 2021). Metode *supervised learning* adalah metode klasifikasi dengan setiap data dalam kumpulan data telah diberi label, dan tujuannya adalah untuk mengklasifikasikan kelas-kelas yang tidak diketahui berdasarkan pola yang dipelajari dari data yang sudah diberi label (Roihan dkk., 2020). Tujuan dari klasifikasi *supervised* yaitu dengan menggunakan contoh data pelatihan untuk merancang pengklasifikasian yang dapat digunakan untuk memprediksi contoh data baru secara akurat (Sarker, 2021).

Berbeda dengan metode *supervised*, teknik *unsupervised learning* tidak bergantung dari kesamaan karakteristik pada contoh-contoh yang telah diberi label sebelumnya, dan teknik *unsupervised learning* sering disebut *clustering* karena tidak memerlukan pemberian label pada data, dan hasilnya tidak mengidentifikasi data berada dalam kelas yang telah ditentukan sebelumnya (Roihan dkk., 2020). Sedangkan metode *reinforcement learning* adalah jenis algoritma pembelajaran mesin yang memungkinkan perangkat lunak dan mesin untuk secara otomatis dapat menilai perilaku optimal dalam suatu konteks atau lingkungan tertentu guna meningkatkan efisiensinya (Sarker, 2021).

2.5 Augmentasi

Augmentasi merupakan strategi yang memungkinkan peningkatan signifikan pada keragaman dari data yang tersedia untuk model pelatihan tanpa perlu mengumpulkan data baru (Ho dkk., 2019). Teknik augmentasi data yang dilakukan ini seperti pemotongan, penerjemahan, pencerminan yang biasanya digunakan untuk melatih model *deep neural network* (DNN) dan diilustrasikan pada Gambar 6 berikut (Sanjaya dan Ayub, 2020).

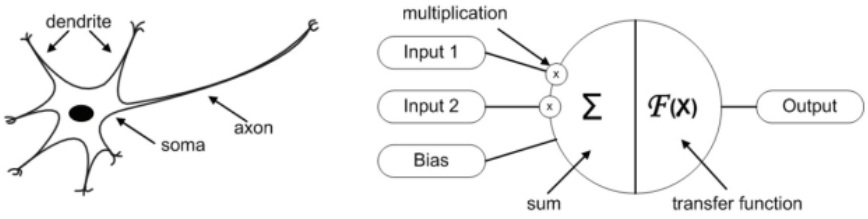


Gambar 6. Ilustrasi Teknik Augmentasi (Sanjaya dan Ayub, 2020)

Teknik ini mengubah data seperti gambar untuk meningkatkan keragaman datanya, dan ini menunjukkan kinerja yang sangat baik dalam tugas seperti klasifikasi gambar (Alomar dkk., 2023). Terdapat pertimbangan utama ketika menerapkan teknik ini yaitu dalam pemilihan fungsi atau teknik augmentasi yang akan digunakan, karena apabila dilakukan secara berlebihan akan mempengaruhi lama waktu dalam pemrosesan gambar (Alomar dkk., 2023).

2.6 *Artificial Neural Network*

Artificial Neural Network (ANN) adalah model pengelola informasi yang mirip dan mencoba untuk menyimulasikan fungsi sistem saraf biologis otak manusia (Abiodun dkk., 2018). Setiap neuron dapat terhubung dan berkomunikasi dengan satu sama lain melalui akson yang berfungsi sebagai koneksi penghubung, ilustrasi yang menggambarkan neuron dapat dijelaskan pada Gambar 7 berikut.



Gambar 7. Ilustrasi Neuron pada *Artificial Neural Network*

Neuron-neuron tersebut saling terhubung, memungkinkannya untuk dapat saling mengirimkan sinyal satu sama lain, mirip dengan koneksi sinapsis dalam otak biologis manusia (Mijwil, 2021). Menurut Mijwil (2021), kekuatan koneksi ANN yang dikenal sebagai bobot, disesuaikan selama proses pembelajaran, memungkinkan jaringan untuk belajar dari data input dan meningkatkan kinerjanya dari waktu ke waktu. Penjelasan terkait rumus dari pembobotan model ANN dapat diuraikan pada Persamaan (2.2) (Aggarwal, 2023) :

$$y(k) = F(\sum_{i=0}^{m} w_i(k) \cdot x_i(k) + b) \quad (2.2)$$

Keterangan :

$x_i(k)$ = adalah nilai input dalam waktu *k* dimana *i* dari 0 hingga ke m,
$w_i(k)$ = adalah nilai bobot dalam waktu k dimana *i* dari 0 hingga ke m,
$y_i(k)$ = adalah nilai output dalam waktu k,
F = adalah fungsi transfer,
b = adalah bias.

Struktur di dalam lapisan ANN terdapat lebih dari satu lapisan komputasi, yang pada umumnya terdiri dari lapisan *input*, tersembunyi dan *output* (Mijwil, 2021). Menurut Aggarwal. (2023), setiap lapisan tersebut saling terhubung untuk melakukan pemrosesan data, Adapun lapisan pada ANN sebagai berikut :

a. Lapisan *Input*

Merupakan lapisan awal pada ANN yang mentransmisikan data ke lapisan selanjutnya yaitu lapisan tersembunyi yang kemudian akan dilakukan pemrosesan dan semua komputasi sepenuhnya terlihat oleh pengguna (Aggarwal, 2023). Jumlah neuron dalam lapisan ini biasanya menyesuaikan jumlah fitur ataupun atribut pada data.

b. Lapisan Tersembunyi

Lapisan Tersembunyi adalah lapisan perantara antara lapisan *input* dan lapisan *output* yang berisi kumpulan neuron-neuron kecil yang melakukan pemrosesan data (Uzair dan Jamil, 2020). Jumlah lapisan tersembunyi dalam ANN biasanya beragam dan memiliki banyak neuron di dalamnya.

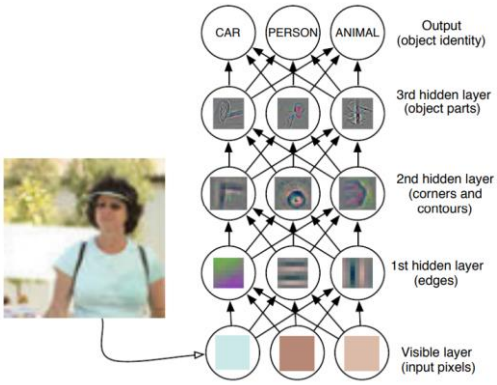
c. Lapisan *Output*

Merupakan lapisan akhir dari jaringan ANN memiliki hubungan langsung dengan kompleksitas masalah atau tugas yang dilakukan (Uzair dan Jamil, 2020). Biasanya jumlah neuron pada lapisan ini bergantung pada tugas yang dilakukan seperti klasifikasi, regresi, dan lainnya.

2.7 *Deep Learning*

Faktor kesulitan utama dalam banyak penerapan kecerdasan buatan adalah banyaknya jumlah faktor yang bervariasi dalam mempengaruhi setiap bagian data yang diamati pada suatu objek data (Bengio dkk., 2021). Sebagai contoh, *pixel* dari gambar mobil berwarna navy akan terlihat sangat mirip dengan warna hitam saat berada di kondisi malam hari. Hal ini menjadi permasalahan dalam pembelajaran representasi. Untuk mengatasi permasalahan tersebut *deep learning* memberikan solusi dalam pembelajaran representasi dengan memperkenalkan representasi yang diekspresikan dalam bentuk yang lebih sederhana. *Deep learning* memungkinkan komputer untuk mengembangkan konsep yang kompleks dari konsep-konsep sederhana.

Deep Learning merupakan aspek dari ANN yang memanfaatkan banyak lapisan dan bertujuan untuk melakukan tugas *machine learning* yang digunakan manusia untuk mendapatkan pengetahuan tertentu (Choldun R dan Surendro, 2018). Contoh dasar dari model *Deep Learning* adalah *multilayer perceptron* (MLP). MLP adalah salah satu jenis arsitektur ANN. Konsep MLP adalah fungsi matematika yang memetakan sekumpulan nilai *input* ke nilai *output*. Arsitektur *Deep Learning* di representasikan pada Gambar 8 sebagai berikut.



Gambar 8. Ilustrasi Representasi *Deep Learning* (Bengio dkk., 2021)

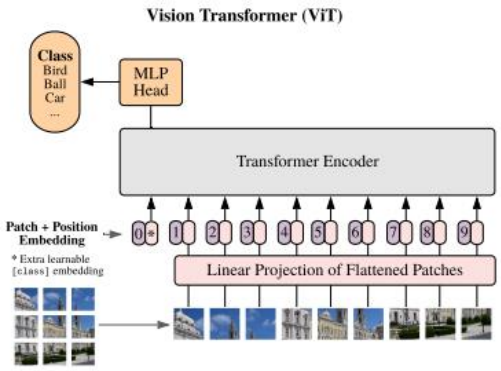
Berdasarkan Gambar 8 dapat diketahui bahwa setiap representasi dari suatu data gambar diwakili oleh neuron. Pada lapisan pertama, *pixel* lapisan tersebut dapat mengidentifikasi tepi dari gambar, selanjutnya pada lapisan kedua dapat dengan mudah untuk menentukan sudut ataupun kontur yang diperluas dan dapat dikatakan sebagai kumpulan tepi. Representasi dari lapisan kedua maka lapisan ketiga dapat mendeteksi seluruh bagian pada suatu objek tertentu ketika menemukan kumpulan tepi dan kontur tertentu. Sehingga pada akhirnya representasi gambar yang dihasilkan dapat digunakan untuk mengenali objek yang ada di dalam suatu data gambar (Bengio dkk., 2021).

Kemampuan *deep learning* dalam menganalisis data yang kompleks membuatnya sangat berguna dibandingkan dengan algoritma lain seperti *machine learning* yang lebih banyak melakukan penekanan pada rekayasa dan seleksi fitur yang penting untuk digunakan. Penerapan *deep learning* telah menjangkau banyak bidang seperti bidang medis, ekonomi, dan social. Contoh aplikasi penerapan *deep learning* seperti pendeteksian objek, prediksi harga saham, klasifikasi suatu penyakit, dan lainnya. Beberapa algoritma dalam *deep learning* seperti *vision transformer* (ViT). Arsitektur ViT merupakan salah satu jenis arsitektur yang telah mencapai kinerja memuaskan dalam menyelesaikan tugas *computer vision* seperti klasifikasi citra, segmentasi, *object detection*, dan lainnya (Gheflati dan Rivaz, 2021).

2.8 *Vision Transformer*

Kesuksesan *transformer* dalam pengolahan data di *Natural Language Processing* (NLP), membuka inspirasi untuk perluasan ke tugas-tugasnya di *computer vision*. Fitur utama dari *transformer* yaitu mekanisme *self-attention* yang membantu model untuk mempelajari hubungan antar kata atau bagian dari data dalam urutan yang sama. Hal ini memungkinkan model untuk "memperhatikan" atau "memahami" ketergantungan yang berbeda antara elemen dalam urutan data (Vaswani dkk., 2023). Belum lama ini mekanisme berbasis perhatian seperti *transformer* telah berhasil diterapkan ke berbagai tugas *computer vision* sebagai *vision transformer* (ViT) (Dosovitskiy dkk., 2020). ViT adalah salah satu jenis arsitektur dalam *deep learning* dengan fokus utama pengolahan citra yang saat ini sedang dikembangkan dan baru-baru ini sukses diterapkan ke dalam tugas *computer vision* seperti klasifikasi citra (Zhou dkk., 2021).

Beberapa tugas di *Computer Vision* seperti pengenalan gambar, deteksi objek, *segmentasi*, *sintesis* gambar teks, dan menjawab pertanyaan visual ViT telah mencapai kinerja yang setara atau dapat dikatakan lebih baik dibandingkan dengan model CNN yang sudah ada sebelumnya dan bertahan cukup lama dalam mengelola citra (Khan dkk., 2022). Sumber kekuatan dari ViT dan *transformer* pada umumnya berasal dari mekanisme *attention* dan variasi parameter seperti banyaknya lapisan *encoder*, dimensi yang tidak terlihat, dan jumlah *head attention.* Struktur model *Vision Transformer* di ilustrasikan pada Gambar 9 sebagai berikut.



Gambar 9. Model *Vision Transformer* (Dosovitskiy dkk., 2021)

Gambar 9 menggambarkan terkait struktur dari arsitektur *Vision Transformer*, dan secara umum ViT terbagi ke dalam 3 proses utama yaitu penyisipan *patching* dengan *positional embedding* dari gambar yang telah di-*patching* sebelumnya, kemudian dilakukan *transformasi encoder*, dan *classification head* dengan MLP. Proses *input* gambar yang dilakukan dalam arsitektur ViT akan dipecah atau dilakukan *splitting* menjadi *patch-patch* kecil biasanya tiap *patch* berukuran 16×16 kemudian diratakan secara linear (Dosovitskiy dkk., 2020). Setelah melalui proses *patch embedding* dilakukan pemberian token input atau penyematan *positional embedding* sehingga dapat diproses pada tahap *transformasi encoder*. Proses selanjutnya yaitu pada lapisan *Multilayer Perceptron* (MLP) dilakukan serangkaian proses untuk mendapatkan klasifikasi.

2.8.1 *Patch Embeddings*

Patch embeddings merupakan salah satu konsep penting dalam arsitektur ViT. Hal ini melibatkan proses mengubah gambar menjadi *patch* dan direpresentasikan dalam vektor, yang memungkinkan ViT untuk memproses gambar menjadi urutan token menggunakan pendekatan berbasis transformer (Dosovitskiy dkk., 2020). Pada ViT, *patch* merupakan konsep terkait dengan pemrosesan citra dengan ukuran yang tersegmentasi atau dipecah menjadi bagian-bagian kecil. Pada arsitektur ini, *input* gambar diubah menjadi bentuk *patch-patch* kecil yang disusun secara linear dan tidak bertumpuk (Dosovitskiy et al., 2020). *Input* gambar dibagi menjadi sebanyak *n* patch dan dihitung dengan Persamaan (2.3) berikut (Dosovitskiy dkk., 2020):

$$n = \frac{hw}{h_p w_p} \tag{2.3}$$

dengan (h, w) merupakan ukuran atau resolusi dari *input* gambar dan *height* (h) adalah tinggi gambar, dengan *width* (w) adalah lebar gambar, dimana $(h_p w_p)$ adalah resolusi dari setiap *patch* gambar dengan ukuran persegi $m \times m$, dan *n* adalah jumlah *patch* yang terbentuk (Dosovitskiy dkk., 2020). *Input* dipartisi menjadi bagian yang tidak tumpang tindih, kemudian diratakan menjadi vektor

satu dimensi, dan diproyeksikan ke dalam ruang fitur beridimensi tinggi menggunakan lapisan linear dengan dimensi embedding **D** (A. Khan dkk., 2023). Pendekatan ini dapat memungkinkan arsitektur ViT untuk mempelajari ketergantungan antar *patch* yang berbeda dapat dijelaskan pada Persamaan (2.4) berikut (A. Khan dkk., 2023).

$$\boldsymbol{X}_{patch}^{\boldsymbol{N}\times\boldsymbol{D}} = R\left(\boldsymbol{I}_{image}^{A\times B\times C}\right) \tag{2.4}$$

dengan *input* Gambar $\boldsymbol{I}$ *image* dengan ukuran $A \times B \times C$, dan $R()$ merupakan fungsi pembentukan untuk menghasilkan jumlah *n patches* dengan ukuran $D = P \times P \times C$ dan C merupakan jumlah *channels* serta P merupakan *patch size* (Khan dkk., 2023).

2.8.2 *Positional Embeddings*

Pada ViT, konsep *positional embeddings* merupakan representasi vektor yang ditambahkan menjadi tiap-tiap token atau elemen dalam urutan data seperti *patch* gambar untuk memberikan informasi terkait dengan posisi atau urutan relatif dari setiap elemen dalam data (Vaswani dkk., 2023). *Positional embeddings* tersebut diperlukan karena ViT tidak memiliki struktur yang memperhitungkan urutan alamiah dalam input gambar yang dimasukkan. Transformer melihat *patch* gambar yang telah disematkan sebagai satu kumpulan *patch* tanpa mengetahui urutannya.

Sebelum urutan dari *patch-patch* dimasukkan ke dalam *encoder*, setiap *patch* diproyeksikan secara linear menjadi vektor model dimensi D menggunakan matriks *embeddings* yang telah dipelajari. Kemudian representasi yang telah disematkan digabungkan dengan token klasifikasi (X_c) yang diperlukan untuk menyelesaikan tugas klasifikasi. Untuk mempertahankan pengaturan spasial *patch* sebagaimana ditunjukkan pada Gambar 9, informasi posisi E_{pos} *patch* harus dikodekan dan dimasukkan ke representasi *patch*. Hasil urutan *embeddings* dari *patches* dengan token z_0 diberikan dengan Persamaan (2.5) berikut (Bazi dkk., 2021) :

$$z_0 = [V_{class}; X_1E; X_2E; \dots; X_nE] + E_{pos}, \quad E \in \mathbb{R}^{(p^2c)\times D} E_{pos} \in \mathbb{R}^{(n+1)\times D} \tag{2.5}$$

dengan V_{class} merupakan *one hot encoding* untuk matriks token yang dibuat oleh komputer, X_n mewakili *patch* ke-*n* dalam bentuk matriks, sedangkan E adalah matriks *embeddings* atau representasi *patch*, dan E_{pos} adalah *position enconding* yang berbentuk matriks. Selanjutnya pengkodean dari masing-masing posisi yang diberikan oleh fungsi *sinus* dan *cosinus*, hal ini dilakukan untuk mendapatkan posisi dari nilai pengkodean pada frekuensi berbeda yang dapat dijelaskan pada Persamaan (2.6) dan (2.7) (Al-hammuri dkk., 2023).

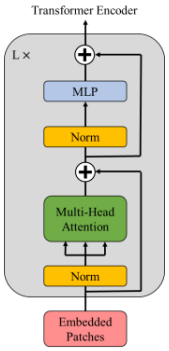
$$P(x, 2j) = sin\left(\frac{x}{10000\frac{2j}{d}}\right), dengan\ -1 \leq P(x, 2j+1) \leq 1 \tag{2.6}$$

$$P(x, 2j+1) = cos\left(\frac{x}{10000\frac{2j}{d}}\right), dengan -1 \leq P(x, 2j+1) \leq 1 \tag{2.7}$$

dengan P adalah *positional encoding*, d merupakan vektor dimensi penyisipan output, x adalah posisi objek dalam urutan input dengan nilai skalar yang telah ditetapkan yaitu 10.000, dan j untuk pemetaan ke indeks kolom dengan $0 \leq j \leq \frac{d}{2}$

2.8.3 *Transformer Encoder*

Hasil dari output *patch* adalah z_0 yang akan diteruskan ke *transformer encoder*. Ilustrasi terkait *transformer encoder* terdapat pada Gambar 10 berikut ini :



Gambar 10. *Transformer Encoder* (Dosovitskiy dkk., 2021)

Gambar 10 merupakan ilustrasi dari *transformer encoder,* terdapat beberapa komponen utama yang membentuk lapisan *transformer encoder* yaitu blok *multihead self-attention* (MSA) (Vaswani dkk., 2023). Lapisan MSA memiliki Persamaan (2.8) sebagai berikut:

$$Z'_{\ell} = MSA\big(LN(Z_{\ell-1})\big) + Z_{\ell-1,} \text{dan } \ell = 1,2,3,\dots,n \quad (2.8)$$

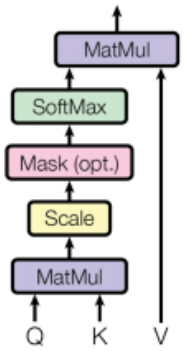
(Bazi dkk., 2021)

dengan blok *multilayer perceptron* (MLP) yang saling terhubung terlihat pada Persamaan (2.9). Blok terakhir dari arsitektur ViT biasanya terdiri dari lapisan *dense* dengan fungsi aktivasi GeLU. Masing-masing dua komponen utama *encoder* menggunakan koneksi *residual skip* dan didahului oleh lapisan *normalization* (LN) dapat diuraikan pada Persamaan (2.9) (Bazi dkk., 2021).

$$Z_{\ell} = MLP\big(LN(Z'_{\ell})\big) + Z'_{\ell,} \text{dan } \ell = 1,2,3,\dots,n \quad (2.9)$$

2.8.4 *Self Attention*

Pada lapisan *self-attention* vektor input pertama-tama diubah menjadi tiga vektor yang berbeda yang teridiri dari *query* (Q), *key* (K), dan *value* (V) dengan dimensi $d_q = d_k = d_v = d_{model} = 512$ (Han dkk., 2023). Vektor-vektor yang berasal dari input yang berbeda tersebut kemudian digabungkan menjadi tiga matriks yang berbeda yaitu Q, K, dan V. Selanjutnya, fungsi-fungsi *self-attention* antara vektor-vektor input yang berbeda tersebut dihitung dan di ilustrasikan pada Gambar 11 berikut.



Gambar 11. *Self-attention* (Han dkk., 2023)

Gambar 11 mengilustrasikan bahwa vektor-vektor tersebut dihitung dengan menggunakan operasi *matrix multiplication* (MatMul) antara Q dan seluruh K, kemudian hasil yang didapatkan dibagi masing-masing dengan $S/\sqrt{d_k}$ untuk menormalisasikan nilai guna menstabilkan *gradients* untuk meningkatkan proses pelatihan. Selanjutnya pengaplikasian fungsi aktivasi *softmax* untuk mendapatkan bobot pada nilai. Pada prosesnya, fungsi *attention* dihitung secara simultan dengan kumpulan *query* yang dikemas bersamaan dalam matriks Q, *key* dan *value* dikemas bersama dalam matriks K dan V.

Proses perhitungan matriks Q, K, dan V diperoleh dengan melakukan transformasi linear pada Z dalam Persamaan (2.5), yang dapat dijelaskan pada Persamaan (2.10), (2.11), dan (2.12) berikut (Fan dkk., 2021),

$$Q = W_q Z \tag{2.10}$$

$$K = W_k Z \tag{2.11}$$

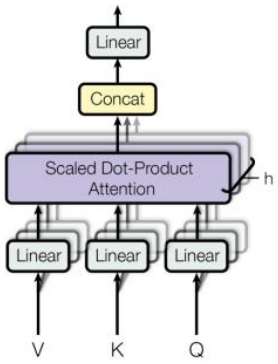
$$V = W_v Z \tag{2.12}$$

dengan $W_q, W_{k,}$dan W_v merupakan bobot transformasi linear yang pada umumnya terdiri dari angka kecil, diinilisasi secara *random* menggunakan distribusi yang sesuai seperti distribusi *Gaussian*, *Xavier,* dan *Kaiming* serta dilakukan sekali sebelum pelatihan dilakukan. Proses perhitungan matriks *output* dapat diformulasikan pada Persamaan (2.13) (Han dkk., 2023)

$$Attention\ (\mathrm{Q}, \mathrm{K}, \mathrm{V}) = softmax\left(\frac{\mathrm{Q} \cdot \mathrm{K}^{\mathrm{T}}}{\sqrt{d_k}}\right) \cdot \mathrm{V} \tag{2.13}$$

2.8.5 *Multi Head Attention*

Multi Head Attention adalah suatu mekanisme yang biasanya digunakan untuk meningkatkan kinerja dari lapisan *self attention*. Dengan menggunakan *multi head attention* akan membuka batasan kemampuan dari model untuk fokus terhadap beberapa tugas lain (Han dkk., 2023). Struktur dari *Multi Head Attention* dapat diilustrasikan pada Gambar 12 sebagai berikut.



Gambar 12. *Multi Head Attention* (Han dkk., 2023)

Berdasarkan Gambar 12 dari ilustrasi struktur *Multi Head Attention*, dapat dijelaskan bahwa *query* (Q), *key* (K), dan *value* (V) diproyeksikan secara linear sebanyak *h* kali dengan proyeksi berbeda yang dipelajari masing-masing dalam dimensi $d_q, d_k,$ dan d_v. Kemudian setiap masing-masing Q, K, dan V yang diproyeksikan fungsi *attention* selanjutnya dijalankan secara parallel dan menghasilkan *output* dimensi d_v $-$*dimensional.* Hasil dari fungsi *attention* yang dijalankan secara paralel pada masing-masing *query, key,* dan *value* yang diproyeksikan ini, kemudian digabungkan dan sekali lagi diproyeksikan sehingga menghasilkan *output* akhir seperti terlihat pada Gambar 12 (Vaswani dkk., 2017).

Multi head attention memungkinkan arsitektur untuk secara bersama-sama memperhatikan informasi dari representasi yang berbeda dalam subruang dan pada posisi yang berbeda-beda dengan satu *attention head* diuraikan pada Persamaan (2.14) dan (2.15) berikut (Vaswani dkk., 2023),

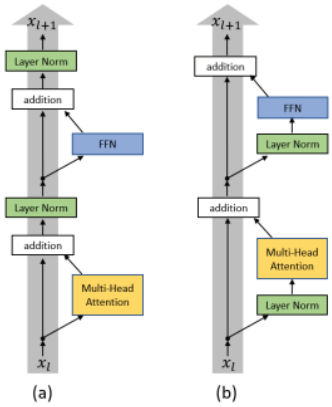
$$\text{MultiHead}\,(Q,K,V) = \text{Concat}(\text{head}_1, \text{head}_2, \ldots, \text{head}_h) \cdot W^o \qquad (2.14)$$

$$\text{dengan head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \qquad (2.15)$$

proyeksi merupakan matriks parameter $W_i^Q \in \mathbb{R}^{d_{model} \times d_k}, W_i^K \in \mathbb{R}^{d_{model} \times d_k}, W_i^V \in \mathbb{R}^{d_{model} \times d_v},$ dan $W_i^O \in \mathbb{R}^{hd_v \times d_{model}}$.

2.8.6 Lapisan *Normalization*

Menurut Figo dkk. (2023), menyatakan bahwa Lapisan *Normalization* merupakan suatu teknik untuk mengurangi penggunaan sumber daya dan waktu untuk menyelesaikan pelatihan dalam arsitektur *deep learning*. Teknik ini mencoba untuk mengurangi penggunaan sumber daya dan waktu pelatihan dengan mengurangi nilai inputan menggunakan rata-rata dan standar deviasi. Lapisan *Normalization* juga memainkan peran penting dalam kesuksesan *transformers*, pada lapisan *transformers* ini ditempatkan diantara residual blok atau setelah *sub*-lapisan dan biasa disebut dengan *Post Layer Normalization* (*Post-LN*) yang diilustrasikan pada Gambar 13 (a), berbeda dengan *Post-LN*, pada *Pre Layer Normalization* (*Pre-LN*) lapisan *normalization* ini diletakkan sebelum *sub*-lapisan residual blok yang dapat dijelaskan pada Gambar 13 (b) (Wang dkk., 2019). Arsitektur ini telah mencapai kinerja *state-of-the-art* dalam menyelesaikan berbagai tugas *machine learning* (Xiong dkk., 2020).



Gambar 13 Ilustrasi (a) Lapisan Post-LN; (b) Lapisan Pre-LN (Xiong dkk., 2020).

Persamaan (2.16) merupakan persamaan yang berisi rumus dalam menghitung lapisan *normalization* (Figo dkk., 2023) :

$$x'_{i,k} = \frac{x_{i,k} - \mu^i}{\sqrt{\sigma_i^2 + \varepsilon}} \tag{2.16}$$

Keterangan :

$x_{i,k}$ = Nilai inputan atau vektor

μ^i = Rata-rata nilai inputan

σ_i^2 = Standar deviasi nilai inputan

ε = Denominator

2.8.7 *Multilayer Perceptron Head* (MLP Head)

Pada lapisan terakhir dari arsitektur ViT yaitu *classification head* mengimplementasikan *Multi Layer Perceptron* (MLP) dengan satu lapisan tersembunyi pada saat *pre-training* dengan satu lapisan linear pada waktu *fine-tuning* (Dosovitskiy dkk., 2020). Lapisan ini berfungsi untuk menggabungkan hasil representasi yang telah diekstraksi dari setiap blok-blok *transformer* sebelumnya dalam ViT dan informasi yang didapat akan diintegrasikan menjadi representasi global yang dapat digunakan untuk tugas klasifikasi (Qayyum dkk., 2022).

Fitur-fitur hasil pemrosesan *transformer encoder* diumpankan ke dalam satu lapisan MLP yang memberikan vektor tersembunyi dapat dijelaskan pada Persamaan (2.8). Vektor ini merupakan integrasi dari urutan fitur dengan menerapkan *attention* sebelumnya (Hafiz dkk., 2021). Pada lapisan ini akan dilakukan pemrosesan dari hasil *output transformer encoder* untuk menyelesaikan berbagai tugas prediksi (Parvaiz dkk., 2023).

2.8.8 Fungsi Aktivasi

Fungsi aktivasi secara khusus digunakan di dalam ANN untuk mengubah sinyal input menjadi sinyal output yang pada gilirannya diteruskan sebagai input ke lapisan berikutnya (Sharma dkk., 2020). Menurut Hao dkk. (2020), tujuan dari fungsi aktivasi adalah untuk memungkinkan jaringan saraf dapat mempelajari dan memodelkan hubungan yang kompleks antara *input* dan *output* data, sehingga dapat meningkatkan akurasi dari model. Fungsi aktivasi ini tidak ada pada neuron di lapisan *input*, hanya terdapat pada neuron di lapisan tersembunyi dan lapisan

output. Transformasi *non-linear* yang ada pada fungsi aktivasi digunakan untuk mempelajari pola yang kompleks dalam data. Beberapa fungsi aktivasi yang biasanya digunakan di dalam *deep learning* sebagai berikut :

1. Fungsi Aktivasi Sigmoid

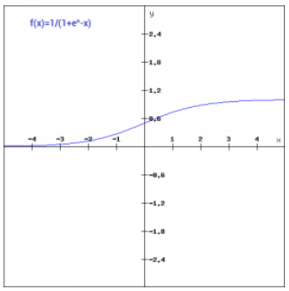
Sigmoid merupakan fungsi aktivasi yang paling banyak digunakan karena merupakan fungsi *non-linear*. Menurut Sharma dkk. (2020) fungsi *sigmoid* mengubah nilai dalam rentang 0 hingga 1. Bentuk dari *sigmoid* direpresentasikan sebagai Persamaan (2.17) dan Gambar 14 sebagai berikut :

$$f_{sigmoid}(x) = \frac{1}{e^{-x}} \tag{2.17}$$

Keterangan :

x = Nilai inputan, dimana $x \in \mathbb{R}$

e^{-x} = Eksponensial dari $-x$, dengan e bernilai 2.7182..



Gambar 14. Fungsi Aktivasi *Sigmoid* (Sharma dkk., 2020)

2. Fungsi Aktivasi *Rectified Linear Unit* (ReLU)

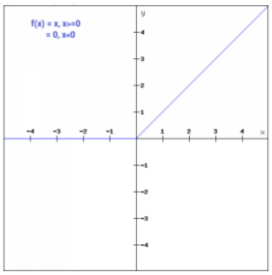
Fungsi aktivasi ReLU merupakan fungsi aktivasi *non-linear* dan merupakan salah satu fungsi aktivasi yang paling banyak digunakan dalam *deep learning*. Keuntungan dari menggunakan aktivasi ReLU adalah selama proses berjalan, tidak semua neuron diaktifkan secara bersamaan (Sharma dkk., 2020). Hal ini berarti bahwa neuron akan di non-aktifkan hanya ketika *output* dari transformasi linear adalah nol. Secara matematis fungsi ini direpresentasikan dalam persamaan (2.18) dan di ilustrasikan pada Gambar 15 sebagai berikut :

$$f_{ReLu}(x) = \max(0, x) \tag{2.18}$$

Keterangan :

x = Nilai inputan, dimana $x \in \mathbb{R}$

$\max(0, x)$ = Fungsi yang mengembalikan nilai maksimum antara 0 dan x



Gambar 15. Fungsi Aktivasi ReLU (Sharma dkk., 2020)

3. Fungsi Aktivasi *Softmax*

Fungsi aktivasi *softmax* merupakan kombinasi dari beberapa fungsi *sigmoid*, seperti yang diketahui bahwa fungsi *sigmoid* menghasilkan nilai dalam rentang 0 hingga 1, yang dapat diinterpretasikan sebagai peluang dari titik kelas tertentu. Fungsi aktivasi *softmax* secara matematis direpresentasikan sebagai Persamaan (2.19) dan direpresentasikan pada Gambar 16 sebagai berikut (Sharma dkk., 2020) :

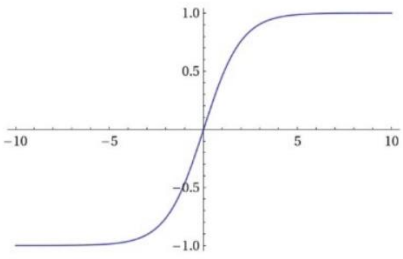
$$f_j(z) = \frac{e^{z_j}}{\sum_k e^{z_k}} \tag{2.19}$$

Keterangan :

z = Nilai inputan terdiri dari beberapa input $z = [z_1, z_2, \dots, z_k]$ dan z_j adalah elemen ke-j dari input, dimana $z \in \mathbb{R}$

e = Eksponensial yang bernilai 2.7182..

$\sum_k e^{z_k}$ = Jumlah dari eksponensial semua elemen dalam vektor z



Gambar 16. Fungsi Aktivasi *Softmax*

4. Fungsi Aktivasi *Gaussian Error Linear Unit* (GELU)

Fungsi aktivasi GELU merupakan fungsi distribusi kumulatif *Gaussian* $x\,\Phi(x)$ dengan x adalah nilai input dari neuron dikalikan dengan $\Phi(x)$ yang merupakan hasil dari nilai distribusi kumulatif *Gaussian* untuk input x. *Non-linear* GELU memberi bobot pada *input* berdasarkan persentilnya bukan berdasarkan gerbang

input seperti ReLU. Akibatnya GELU dapat dianggap sebagai ReLU yang lebih baik (Hendrycks dan Gimpel, 2023), secara matematis GELU dapat direpresentasikan sebagai Persamaan (2.20) dan divisualisasikan dalam Gambar 17 sebagai berikut (Sharma dkk., 2020) :

$$f_{GELU}(x) = x\Phi(x) = x \cdot \frac{1}{2}\left(1 + \text{erf}\left(\frac{x}{\sqrt{2}}\right)\right) \tag{2.20}$$

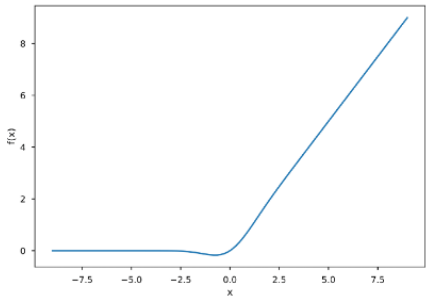
$$\approx x.\frac{1}{2}\left(1 + tanh\left(\sqrt{\frac{2}{\pi}}(x + 0{,}044715x^3)\right)\right)$$

Keterangan :

x = Nilai input, dimana $x \in \mathbb{R}$

$\Phi(x)$ = Fungsi distribusi kumulatif dari distribusi normal baku

$\text{erf}\left(\frac{x}{\sqrt{2}}\right)$ = Fungsi kesalahan (*error function*)



Gambar 17. Fungsi Aktivasi GELU

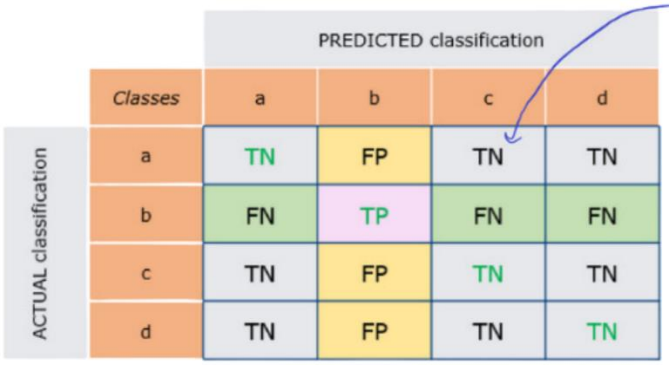
2.8.9 *Adaptive Moment Estimation* (Adam) *Optimizer*

Adaptive moment estimation merupakan salah satu algoritma atau metode yang digunakan untuk memperbarui parameter seperti bobot ataupun learning rate sehingga mencapai hasil yang optimal (Syifa dan Dewi, 2022). Teknik ini berdasarkan pada estimasi momen orde rendah adaptif dari gradien, dan mengatur kecepatan pembelajaran seiring berjalannya waktu berdasarkan gradien sebelumnya.

2.9 *Confusion Matrix*

Confusion matrix adalah representasi yang bermanfaat dan menyeluruh dari kinerja pengklasifikasi. *Confusion matrix* digunakan pada saat melakukan evaluasi model klasifikasi multi-kelas atau label tunggal. Setiap *instance* data hanya dapat dimiliki oleh satu kelas pada titik waktu tertentu (Krstinić dkk., 2023). Variabel respon Y dan prediksi $\hat{Y}$ dianggap sebagai dua variabel acak diskrit dalam klasifikasi multikelas dengan asumsi nilai-nilai kelas klasifikasi. Setiap angka menunjukkan masing-masing kelas yang berbeda.

Pada umumnya, terdapat empat ukuran yang sering digunakan dalam *confusion matrix* yaitu *true positive* (TP), *true negative* (TN), *false positive* (FP), dan *false negative* (FP) yang dapat digunakan untuk menentukan apakah seorang pasien di diagnosa penyakit tertentu ilustrasi terkait dengan *Confusion Matrix* dapat dijelaskan pada Gambar 18 berikut (Hasnain dkk., 2020).



Gambar 18. Ilustrasi *Confusion Matrix*

True positive menjelaskan terkait prediksi penyakit dan sebenarnya penyakit, *false positive* didefinisikan sebagai prediksi menderita penyakit, akan tetapi sebenarnya normal, *true negative* menjelaskan bahwa prediksi seseorang sehat dan sebenarnya sehat, dan *false negative* menjelaskan sebagai prediksi sehat, akan tetapi sebenarnya menderita penyakit (Juneja dkk., 2020). Terdapat beberapa metrik evaluasi klasifikasi untuk menentukan kinerja metode yang digunakan diantaranya yaitu akurasi, sensitivitas, spesifitas, dan *F1-score*. Akurasi dapat didefinisikan sebagai tingkat kesamaan antara hasil prediksi dan nilai sebenarnya atau persentase sampel

yang benar dan salah dalam mengidentifikasi dan dihitung dengan menggunakan Persamaan (2.21), sensitivitas mengukur tingkat keberhasilan model dalam mengidentifikasi data berlabel positif dengan benar dengan menggunakan Persamaan (2.22), sedangkan spesifitas merupakan kebalikan dari senstivitas yaitu mengukur tingkat keberhasilan model dalam mengidentifikasi data berlabel negatif dengan benar dan dapat dihitung menggunakan Persamaan (2.23), dan yang terakhir yaitu *F1-Score* adalah ukuran metriks dengan menghitung rata-rata harmonik dari presisi dan *recall* diformulasikan pada Persamaan (2.24) (Xhumari dan Haloci, 2023).

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN}\ 100\% \tag{2.21}$$

$$Sensitivitas = \frac{TP}{(TP + FN)}\ 100\% \tag{2.22}$$

$$Spesifitas = \frac{TN}{(TN + FP)}\ 100\% \tag{2.23}$$

$$F1 - Score = \frac{2TP}{2TP + FP + FN} \tag{2.24}$$

(Krstinić dkk., 2023)

Nilai dari akurasi, sensitivitas, spesifitas, dan *F1-score* di kelompokkan ke dalam beberapa kategori yang dirangkum pada Tabel 3,

Tabel 3 Kategori Nilai Kinerja Model Arsitektur (Mishra dkk., 2016)

Nilai Kinerja (%)	Kategori
> 90	Sangat baik
81-90	Baik
71-80	Cukup
61-70	Kurang Baik
< 60	Gagal

Pada Tabel 3 kategori nilai kinerja dinilai sangat baik apabila nilai kinerjanya mencapai di atas 90%, untuk kategori baik ketika mencapai nilai 81-90%, kinerja dianggap cukup ketika mendapatkan nilai di antara 71-80%. Sementara apabila nilai kinerja berada di antara 61-70% dianggap model tersebut kurang baik, dan jika di bawah 60% maka dikategorikan sebagai gagal.

III. METODOLOGI PENELITIAN

3.1 Waktu dan Tempat Penelitian

Waktu dan tempat dilaksanakannya penelitian ini sebagai berikut :

a. Tempat Penelitian

Penelitian ini dilaksanakan secara studi literatur di Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Lampung. Dengan lokasi bertempat di Jl. Prof Dr.Soemantri Brojonegoro No.1, Gedong Meneng, Bandar Lampung, Lampung.

b. Waktu Penelitian

Waktu penelitian dilaksanakan pada semester ganjil tahun akademik 2023/2024, tepatnya pada bulan Oktober 2023. Pada penelitian ini dibagi menjadi tiga tahapan, tahap pertama yaitu studi literatur terkait topik dan tema penelitian yang dilakukan, dilanjutkan dengan pengumpulan data terkait serta penyusunan *draft* proposal penelitian. Tahap kedua yaitu tahap lanjutan dari penelitian, pada tahap ini mulai dilakukan pengerjaan program dari *exploratory data analysis*, *pre-processing data*, serta membangun model yang akan dilakukan *training* dengan beberapa kombinasi parameter berbeda dan melakukan *training* model terhadap parameter terbaik yang digunakan. Tahap ketiga adalah penyusunan hasil penelitian dan kesimpulan dari penelitian yang telah dilakukan dalam bentuk draf untuk disampaikan pada seminar hasil.

3.2 Alat Penelitian

Perangkat yang digunakan dalam penelitian ini adalah laptop dengan merek ASUS VivoBook X415 dengan tipe 64-bit *operating system*, x64-*based processor*.
Alat yang digunakan untuk digunakan dalam penelitian ini adalah sebagai berikut:

a. Perangkat Keras (*Hardware*)
   - Processor : Intel Core i3 Gen 11^{Th} dengan Intel UHD Graphics
   - Memori : SSD 512 GB
   - RAM : 12 GB

b. Perangkat Lunak (*Software*)
   - Sistem Operasi *Personal Computer* Windows 11
   - Google Colab dan Kaggle Notebooks dengan menggunakan T4 GPU dengan package sebagai berikut :
     1. Numpy (1.23.5) : Digunakan untuk operasi numerik dan manipulasi *array* pada data.
     2. Pandas (1.5.3) : Digunakan untuk manipulasi data dan membuat *dataframe.*
     3. TensorFlow (2.15.0) : Digunakan untuk membangun arsitektur dari model dan proses pengolahan data seperti augmentasi.
     4. PIL (9.4.0) : Digunakan untuk manipulasi gambar.
     5. OS(3.10.12) : Digunakan untuk memanipulasi direktori.
     6. Torch (2.1.0+cu121) : Untuk membangun model *deep learning.*
     7. Sklearn (1.2.2) : Digunakan untuk *splitting* dataset.
     8. Transformers (4.35.2) : Digunakan untuk membangun model ViT.
     9. Matplotlib (3.7.1) : Digunakan untuk membuat visualisasi seperti grafik dan plot.
     10. Seaborn (0.12.2) :Digunakan untuk membuat visualisasi *Confusion Matrix.*

3.3 Data Penelitian

Data yang digunakan dalam penelitian ini merupakan data citra yang diakses secara online melalui https://www.kaggle.com/datasets/arjunbasandrai/medical-scan-classification-dataset yang bersumber dari *Alzheimer's Disease Neuroimaging Initiative* (ADNI). Merupakan kumpulan data dari *website* Kaggle yang berisi data citra hasil MRI mengenai penyakit *Alzheimer*. Kumpulan data ini terdiri dari 8000 citra MRI otak manusia yang dikategorikan menjadi 4 kelas yaitu *non demented, mild demented, very mild demented,* dan *moderate demented.* Masing-masing dari 4 kategori kelas tersebut memiliki sebanyak 2000 gambar di dalamnya dengan resolusi tiap gambar berukuran 200 × 190, kemudian gambar tersebut akan dilakukan proses *resize* menjadi ukuran 224 × 224.

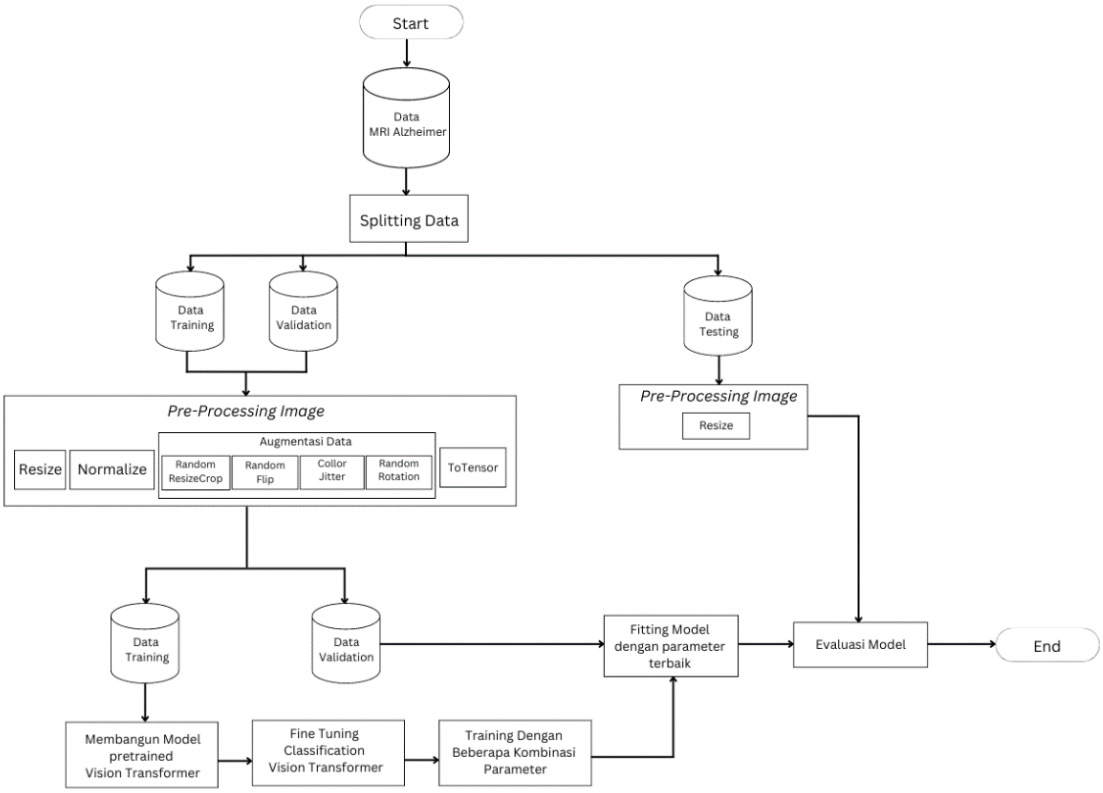
3.4 Metode Penelitian

Pada metode penelitian tugas akhir ini menggunakan sebuah metode *deep learning* untuk melakukan tugas *computer vision* yaitu metode *Vision Transformer* (ViT). Beberapa tahapan langkah yang dilakukan dalam penelitian ini adalah sebagai berikut :

1. Pengumpulan dan penginputan kumpulan data citra MRI *Alzheimer.*
2. Melakukan *splitting* data dari total 8000 citra menjadi 90% data model dan 10% data *testing*. Kemudian dari 90% data model dibagi kembali menjadi 90% untuk data *training* dan 10% untuk data validasi.
3. Pada tahap *pre-processing* dilakukan beberapa proses yaitu normalisasi data, proses *resize* pada gambar menjadi berukuran 224 × 224, augmentasi data, dan mengubah gambar menjadi bentuk *tensor PyTorch.*
4. Melakukan augmentasi pada kumpulan data seperti *RandomResizedCrop, RandomFlip, RandomRotation* dan *ColorJitter* pada data *training.*
5. Membangun model *pre-trained Vision Transformer* (ViT).
6. Melakukan *fine-tuning classification* ViT pada data training.

7. Melakukan *training* dengan beberapa kombinasi parameter *learning rate* dan *batch size* berbeda untuk mencari parameter terbaik yang digunakan.
8. Melakukan *training* pada model ViT yang telah dibangun menggunakan data *training* dan melakukan validasi menggunakan data validasi.
9. Evaluasi model dari ViT dengan menggunakan *confusion matrix* dan metrik evaluasinya yaitu akurasi, spesifitas, sensitivitas, dan *F1-score*, menggunakan data *testing*.

Berikut ini merupakan beberapa tahapan proses penelitian yang dilakukan dapat dijelaskan pada diagram *flowchart* pada Gambar 19 sebagai berikut.



Gambar 19. *Flowchart* Metode Penelitian

BAB V. KESIMPULAN DAN SARAN

5.1. Kesimpulan

Berdasarkan penelitian yang dilakukan dengan menggunakan model *pre-trained Vision Transformer* yaitu "vit-base-patch16-224" dalam melakukan tugas klasifikasi citra MRI penyakit *Alzheimer* didapatkan beberapa kesimpulan antara lain:

1. Penggunaan model *pre-trained* "vit-base-patch16-224" terbukti sangat efektif dan memiliki kinerja yang baik dalam melakukan tugas klasifikasi data citra MRI penyakit *Alzheimer*. Model ini mampu untuk secara akurat membedakan karakteristik kelas-kelas yang berbeda pada citra MRI penyakit *Alzheimer*.
2. Model *pre-trained* yang telah dilatih mampu bekerja dengan sangat baik dalam memprediksi data baru citra MRI *Alzheimer* dibuktikan dengan nilai akurasi, sensitivitas, spesifitas, dan *F1-score* secara berurutan yaitu 98,19%, 96,34%, 98,80%, dan 96,37%.
3. Hasil dari temuan ini relevan dengan tujuan penelitian dalam klasifikasi citra MRI Penyakit *Alzheimer*. Penerapan model *pre-trained* ViT membuktikan kinerja yang sangat baik dalam menganalisis data citra dan menghasilkan prediksi yang akurat.

5.2 Saran

Saran yang dianjurkan untuk penelitian selanjutnya adalah meningkatkan jumlah data yang digunakan dalam proses *training*, validasi, dan *testing* sehingga model yang digunakan dapat dimaksimalkan dan mendapat hasil yang lebih optimal. Selain itu penelitian selanjutnya dapat menggunakan beberapa model *pre-trained Vision Transformer* (ViT) lainnya yang lebih tinggi seperti varian *large* dan model *pre-trained* yang dilatih dengan data *pixel* berukuran 384×384. Tidak terbatas pada arsitektur ViT, penelitian selanjutnya juga dapat menggunakan model *transformer* lainnya untuk mengelola data citra seperti arsitektur *Data-Efficient Image Transformer* (DeiT) dan *Swin Transformer*.

DAFTAR PUSTAKA

Abijono, H., Santoso, P., dan Anggreini, N. L. 2021. Algoritma Supervised Learning Dan Unsupervised Learning dalam Pengolahan Data. *Jurnal Teknologi Terapan: G-Tech*, **4**(2), 315–318. https://doi.org/10.33379/gtech.v4i2.635

Abiodun, O. I., Jantan, A., Omolara, A. E., Dada, K. V., Mohamed, N. A., dan Arshad, H. 2018. State-Of-The-Art In Artificial Neural Network Applications: A survey. *Heliyon*, **4**(11), 1–41. https://doi.org/10.1016/j.heliyon.2018.e00938

Aggarwal, C. C. 2023. *Neural Networks and Deep Learning*. Springer International Publishing, New York, USA. 493 hlm. https://doi.org/10.1007/978-3-031-29642-0

Al-hammuri, K., Gebali, F., Kanan, A., dan Chelvan, I. T. 2023. Vision Transformer Architecture And Applications In Digital Health: A Tutorial And Survey. *Visual Computing for Industry, Biomedicine, and Art*, **6**(14), 1–28. https://doi.org/10.1186/s42492-023-00140-9

Almumtazah, N., Kiromi, M. S., Ulinnuha, N., dan Korespondensi, P. 2023. Klasifikasi Alzheimer Berdasarkan Data Citra Mri Otak Menggunakan FCM dan ANFIS. *Jurnal Teknologi Informasi dan Ilmu Komputer*. **10**(3), 613–622. https://doi.org/10.25126/jtiik.2023106826

Alomar, K., Aysel, H. I., dan Cai, X. 2023. Data Augmentation in Classification and Segmentation: A Survey and New Strategies. *Journal of Imaging*, **9**(46), 1–26. https://doi.org/10.3390/jimaging9020046

Alwanda, M. R., Ramadhan, R. P. K., dan Alamsyah, D. 2020. Implementasi Metode Convolutional Neural Network Menggunakan Arsitektur LeNet-5 untuk Pengenalan Doodle. *Jurnal Algoritme*, **1**(1), 45–56. https://doi.org/10.35957/algoritme.v1i1.434

Alzheimer's Association. 2023. *Alzheimer's & Dementia*, **19**(4), 1598–1695. https://doi.org/10.1002/alz.13016

Asmara, R.A. 2018. *Pengolahan Citra Digital*. POLINEMA Press, Malang. 112 hlm.

Astuti, L. W. 2019. Ekstrasi Fitur Citra MRI Otak Menggunakan Data Wavelet Transform (DWT) untuk Klasifikasi Penyakit Tumor Otak. *Jurnal Ilmiah Informatika Global*, **10**(2), 80–86. https://doi.org/10.36982/jiig.v10i2.854

Bachtiar, F. A., Syahputra, I. K., dan Wicaksono, S. A. 2019. Perbandingan Algoritme Machine Learning untuk Memprediksi Pengambil Matakuliah. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, **6**(5), 543–548. https://doi.org/10.25126/jtiik.2019651755

Bazi, Y., Bashmal, L., Rahhal, M. M. Al, Dayil, R. Al, dan Ajlan, N. Al. 2021. Vision Transformers for Remote Sensing Image Classification. *Remote Sensing*, **13**(3), 1–19. https://doi.org/10.3390/rs13030516

Bengio, Y., Lecun, Y., dan Hinton, G. 2021. Deep learning for AI. *Communications of the ACM*, **64**(7), 58–65. https://doi.org/10.1145/3448250

Burger, W., dan Burge, M. J. 2022. *Digital Image Processing*. Springer London. https://doi.org/10.1007/978-1-4471-6684-9

Choldun, M. I. R. , dan S. K. 2018. Klasifikasi Penelitian dalam Deep Learning. *Jurnal Ilmiah Manajemen Informatika* , **10**(1), 25–34.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., dan Houlsby, N. 2020. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *Proceedings of 2020 International Conference on Learning Representations*, 1–21.

Fabrizio, C., Termine, A., Caltagirone, C., dan Sancesario, G. 2021. Artificial Intelligence for Alzheimer's Disease: Promise or Challenge? *Diagnostics*, **11**(8), 1–19. https://doi.org/10.3390/diagnostics11081473

Fan, H., Xiong, B., Mangalam, K., Li, Y., Yan, Z., Malik, J., dan Feichtenhofer, C. 2021. Multiscale Vision Transformers. *Proceedings of IEEE/CVF International Conference on Computer Vision (ICCV)*, 6804–6815. https://doi.org/10.1109/ICCV48922.2021.00675

Figo, J., Yudistira, N., dan Widodo, A. W. 2023. Deteksi Covid-19 dari Citra X-ray menggunakan Vision Transformer. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, **7**(3), 1116–1125.

Fitriyah, H., dan W. R. C. 2021. *Dasar-Dasar Pengolahan Citra Digital*. UB Press, Malang. 176 hlm.

Gautam, R., dan Sharma, M. 2020. Prevalence and Diagnosis of Neurological Disorders Using Different Deep Learning Techniques: A Meta-Analysis. *Journal of Medical Systems*, **44**(2), 1–24. https://doi.org/10.1007/s10916-019-1519-7

Ghazal, M. T., Abbas, S., Munir, S., Khan, M. A., Ahmad, M., Issa, G. F., Zahra, S. B., Khan, M. A., dan Hasan, M. K. 2022. Alzheimer Disease Detection Empowered with Transfer Learning. *Journal Computers, Materials, & Continua*, **70**(3), 5005–5019. doi: 10.32604/cmc.2022.020866

Gheflati, B., dan Rivaz, H. 2021. Vision Transformer for Classification of Breast Ultrasound Images. *Proceedings of 2022 44th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 480–483. doi : 10.1109/EMBC48229.2022.9871809.

Gonzales, R. C., dan Woods, R.E. 2018. *Digital Image Processing*. Pearson, England. 993 hlm.

Guntara, R. G. 2022. Aplikasi Pengenalan Citra Wajah di KTP Menggunakan Google Cloud Vision API dan Kairos API Berbasis Android. *ILKOMNIKA: Journal of Computer Science and Applied Informatics*, **4**(2), 198–207. https://doi.org/10.28926/ilkomnika.v4i2.504

Hafiz, A. M., Parah, S. A., dan Bhat, R. U. A. 2021. Attention Mechanisms And Deep Learning For Machine Vision: A Survey Of The State Of The Art. *Journals of Computer Vision and Pattern Recognition*, **2160**(7550), 1–24. doi: 10.48550/arXiv.2106.07550

Han, K., Wang, Y., Chen, H., Chen, X., Guo, J., Liu, Z., Tang, Y., Xiao, A., Xu, C., Xu, Y., Yang, Z., Zhang, Y., dan Tao, D. 2023. A Survey on Vision Transformer. *Journals of IEEE Transactions on Pattern Analysis and Machine Intelligence*, **45**(1), 87–110. https://doi.org/10.1109/TPAMI.2022.3152247

Hao, W., Yizhou, W., Yaqin, L., dan Zhili, S. 2020. The Role of Activation Function in CNN. *Proceedings of 2020 2nd International Conference on Information Technology and Computer Application (ITCA)*, 429–432. https://doi.org/10.1109/ITCA52113.2020.00096

Hasnain, M., Pasha, M. F., Ghani, I., Imran, M., Alzahrani, M. Y., dan Budiarto, R. 2020. Evaluating Trust Prediction and Confusion Matrix Measures for Web Services Ranking. *Journals of IEEE Access*, **8**, 90847–90861. https://doi.org/10.1109/ACCESS.2020.2994222

Hendrycks, D., dan Gimpel, K. 2023. Gaussian Error Linear Units (GELUs). *Journals of Machine Learning*, **1606**(8415), 1–10. doi: 10.48550/arXiv.1606.08415

Heo, B., Yun, S., Han, D., Chun, S., Choe, J., dan Oh, S. J. 2021. Rethinking Spatial Dimensions of Vision Transformers. *Proceedings of 2021 IEEE International Conference on Computer Vision (ICCV)*, 11916–11925. https://doi.org/10.1109/ICCV48922.2021.01172

Ho, D., Liang, E., Stoica, I., Abbeel, P., dan Chen, X. 2019. Population Based Augmentation: Efficient Learning of Augmentation Policy Schedules. *Proceedings of 36th International Conference on Machine Learning*, **97**, 2731–2741.

Ilahiyah, S., dan Nilogiri, A. 2018. Implementasi Deep Learning Pada Identifikasi Jenis Tumbuhan Berdasarkan Citra Daun Menggunakan Convolutional Neural Network. *Jurnal Sistem Dan Teknologi Informasi Indonesia*, **3**(2), 49–56.

Juneja, M., Thakur, N., Thakur, S., Uniyal, A., Wani, A., dan Jindal, P. 2020. GC-NET for classification of glaucoma in the retinal fundus image. *Machine Vision and Applications*, **31**(5), 1–18. https://doi.org/10.1007/s00138-020-01091-4

Khan, A., Rauf, Z., Sohail, A., Khan, A. R., Asif, H., Asif, A., dan Farooq, U. 2023. A survey of the vision transformers and their CNN-transformer based variants. *Artificial Intelligence Review*, **56**(S3), 2917–2970. https://doi.org/10.1007/s10462-023-10595-0

Khan, S., Naseer, M., Hayat, M., Zamir, S. W., Khan, F. S., dan Shah, M. 2022. Transformers in Vision: A Survey. *ACM Computing Surveys*, **54**(10s), 1–41. https://doi.org/10.1145/3505244

Khan, U., Ali, A., Khan, S., Aadil, F., Durrani, M. Y., Muhammad, K., Baik, R., dan Lee, J. W. 2019. Internet Of Medical Things–Based Decision System For Automated Classification of Alzheimer's Using Three-Dimensional Views Of Magnetic Resonance Imaging Scans. *International Journal of Distributed Sensor Networks*, **15**(3), 1–15. https://doi.org/10.1177/1550147719831186

Kim, W.-S., Lee, D.-H., Kim, Y.-J., Kim, T., Hwang, R.-Y., dan Lee, H.-J. 2020. Path Detection For Autonomous Traveling In Orchards Using Patch-Based CNN. *Journals of Computers and Electronics in Agriculture*, **175**, 1–11. https://doi.org/10.1016/j.compag.2020.105620

Kirana, K. C. 2021. *Pengolahan Citra Digital Teori dan Penerapan Pengolahan Citra Citra Digital Pada Deteksi Wajah*. Ahlimedia Press, Malang. 121 hlm.

Krstinić, D., Šerić, L., dan Slapničar, I. 2023. Comments on "MLCM: Multi-Label Confusion Matrix." *Journals of IEEE Access*, **11**, 40692–40697. https://doi.org/10.1109/ACCESS.2023.3267672

Mijwil, M. M. 2021. Artificial Neural Networks Advantages and Disadvantages. *Journals of Big Data Mesopotamian*, **2021**(6), 29–31. https://doi.org/10.58496/MJBD/2021/006

Mandala, J. Y., Warsito, A., Lapono, dan Louk, A. 2021. Rancang Bangun Prototype Sistem Akuisisi Citra Digital Untuk Proses Tomografi. *Jurnal Fisika: Fisika Sains Dan Aplikasinya*, **6**(1), 32–43.

Mishra, S., Vanli, O. A., Huffer, F. W., dan Jung, S. 2016. Regularized Discriminant Analysis For Multi-Sensor Decision Fusion and Damage Detection With Lamb Waves. *Proceedings of 2016 Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace System*, **9803**, 728–741. https://doi.org/10.1117/12.2217959

Ninditama, I. P. 2021. Model Machine Learning untuk Klasifikasi Keluarga Sejahtera Study Kasus : Kecamatan Kota Palembang. *Jurnal Tekno Kompak*, **15**(2), 37–49. https://doi.org/10.33365/jtk.v15i2.1156

Parvaiz, A., Khalid, M. A., Zafar, R., Ameer, H., Ali, M., dan Fraz, M. M. 2023. Vision Transformers in Medical Computer Vision—A Contemplative Retrospection. *Journals of Engineering Applications of Artificial Intelligence*, **122**, 1–60. https://doi.org/10.1016/j.engappai.2023.106126

Pytorch. 2023. *Dokumentasi Package Torch Normalize.* www.pytorch.org. diakses pada tanggal 06 Juni 2024 pukul 13.30.

Qayyum, A., Benzinou, A., Mazher, M., dan Meriaudeau, F. 2022. Efficient Multi-model Vision Transformer Based on Feature Fusion for Classification. *Proceedings of 2022 Diabetic Foot Ulcers Grand Challenge*, **13183**, 62–75. https://doi.org/10.1007/978-3-030-94907-5_5

Roihan, A., Sunarya, P. A., dan Rafika, A. S. 2020. Pemanfaatan Machine Learning dalam Berbagai Bidang: Review paper. *IJCIT (Indonesian Journal on Computer and Information Technology)*, **5**(1), 75–82. https://doi.org/10.31294/ijcit.v5i1.7951

Sailaja, B., dan Gopal, T. V. 2023 . Alzheimer's Disease Detection Using Hybrid Transfer Learning. *Journal of Fundamental & Comparative*, **9**(2), 89–100.

Sanjaya, J., dan Ayub, M. 2020. Augmentasi Data Pengenalan Citra Mobil Menggunakan Pendekatan Random Crop, Rotate, dan Mixup. *Jurnal Teknik Informatika Dan Sistem Informasi*, **6**(2), 311–323. https://doi.org/10.28932/jutisi.v6i2.2688

Sarker, I. H. 2021. Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Computer Science*, **2**(3), 1–21. https://doi.org/10.1007/s42979-021-00592-x

Setiyoningsih, W., dan Apriantoro, N. H. 2023. Peranan Magnetic Resonance Imaging (MRI) Lumbal dalam Penegakan Diagnosis Kasus Low Back Pain: Studi Kasus di Instalasi Radiologi Rumah Sakit Satya Negara Jakarta Utara. *Jurnal Imejing Diagnostik (JImeD)*, **9**(2), 57–61. https://doi.org/10.31983/jimed.v9i2.9990

Sharma, S., Sharma, S., dan Athaiya, A. 2020. Activation Functions In Neural Networks. *International Journal of Engineering Applied Sciences and Technology*, **04**(12), 310–316. https://doi.org/10.33564/IJEAST.2020.v04i12.054

Shin, H., Jeon, S., Seol, Y., Kim, S., dan Kang, D. 2023. Vision Transformer Approach for Classification of Alzheimer's Disease Using 18F-Florbetaben Brain Images. *Journal Applied Sciences*, **13**(6), 1–14. https://doi.org/10.3390/app13063453

Sianturi, A. G. M. 2021. Stadium, Diagnosis, dan Tatalaksana Penyakit Alzheimer. *Journal Majalah Kesehatan Indonesia*, **2**(2), 39–44. https://doi.org/10.47679/makein.202132

Sumijan, dan Purnama, P. A. W. 2021. *Teori dan Aplikasi Pengolahan Citra Digital Penerapan dalam Bidang Citra Medis*. Insan Cendekia Mandiri, Solok. 176 hlm.

Syifa, A. S., dan Dewi, A. I. 2022. Arsitektur Resnet-152 dengan Perbandingan Optimizer Adam dan RMSProp untuk Mendeteksi Penyakit Paru-Paru. *Journal Multimedia Artificial Networking Database (MIND)*, **7**(2), 139–150.

Uzair, M., dan Jamil, N. 2020. Effects of Hidden Layers on the Efficiency of Neural Networks. *Proceedings of 2020 IEEE 23rd International Multitopic Conference (INMIC)*, 1–6. https://doi.org/10.1109/INMIC50486.2020.9318195

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., dan Polosukhin, I. 2023. Attention Is All You Need. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 6000–6010. doi: 10.5555/3295222.3295349

Wang, Q., Li, B., Xiao, T., Zhu, J., Li, C., Wong, D. F., dan Chao, L. S. 2019. Learning Deep Transformer Models for Machine Translation. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, **1176**, 1810–1822. https://doi.org/10.18653/v1/P19-1176

Wu, J., Hu, R., Xiao, Z., Chen, J., dan Liu, J. 2021. Vision Transformer-based recognition of diabetic retinopathy grade. *Journals of Medical Physics*, **48**(12), 7850–7863. https://doi.org/10.1002/mp.15312

Xhumari, E., dan Haloci, S. 2023. A comparative study of Credit Scoring and Risk Management Techniques in Fintech: Machine Learning vs Regression Analysis. *Proceedings of the 5th International Conference on Recent Trends and Applications in Computer Science and Information Technology*, **3402**, 13–20.

Xiong, R., Yang, Y., He, D., Zheng, K., Zheng, S., Xing, C., Zhang, H., Lan, Y., Wang, L., dan Liu, T. 2020. On Layer Normalization in the Transformer Architecture. In H. D. III & A. Singh (Eds.), *Proceedings of the 37th International Conference on Machine Learning*, **119**, 10524–10533. https://proceedings.mlr.press/v119/xiong20b.html

Zhou, D., Kang, B., Jin, X., Yang, L., Lian, X., Jiang, Z., Hou, Q., dan Feng, J. 2021. DeepViT: Towards Deeper Vision Transformer. *Journal Computer Vision and Pattern Recognition*, **2103**(11886), 1–12. doi: 10.48550/arXiv.2103.11886