BAB II

TINJAUAN PUSTAKA

2.1 Landasan Teori

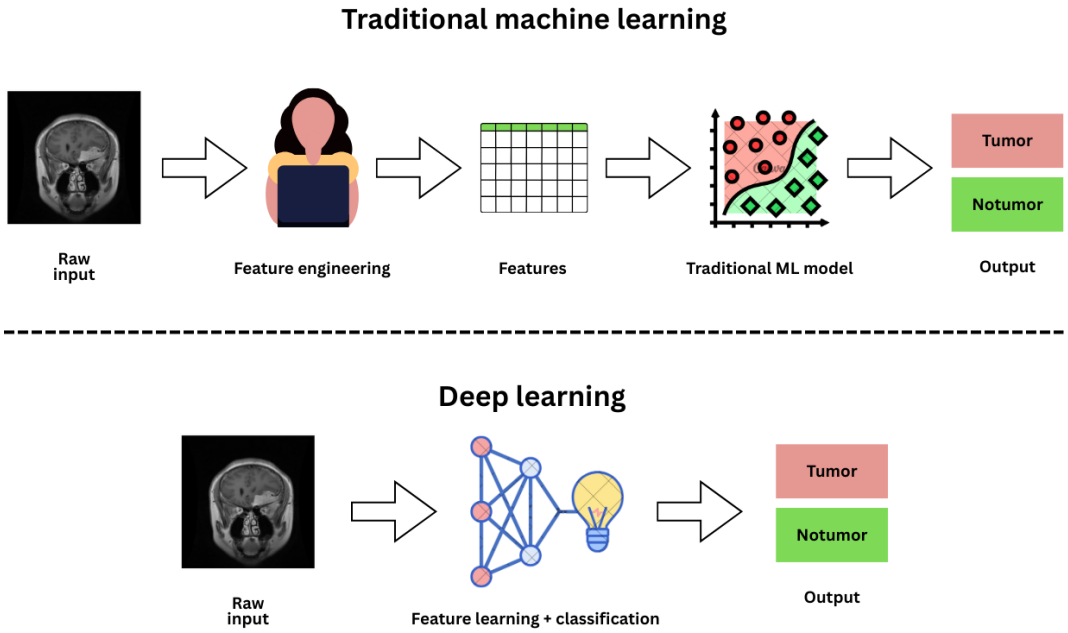
Landasan teori merupakan bagian yang menjelaskan konsep-konsep dasar, hasil penelitian sebelumnya serta teori yang menjadi dasar dalam pelaksanaan penelitian. Penyajian landasan teori bertujuan untuk memberikan pemahaman yang lebih jelas mengenai topik yang dikaji serta memperkuat argumen.

2.1.1 *Deep learning*

Deep learning merupakan sub bidang dalam machine learning yang menggunakan arsitektur neural network yang terdiri dari banyak lapisan (*layers*) untuk mempelajari representasi data yang semakin kompleks (Goodfellow dkk., 2016). Pada *deep learning* data input akan melewati serangkaian lapisan neuron yang semakin dalam dan kompleks untuk memahami pola-pola yang terkandung dalam data tersebut. *Deep learning* dapat mengekstraksi fitur-fitur yang lebih abstrak dan kompleks dari data dengan menggunakan lapisan tersebut,. Hal ini memungkinkan menghasilkan model yang lebih baik dalam tugas-tugas seperti klasifikasi, deteksi objek, dan lainnya (Benyahia dkk., 2022).

Deep learning berbeda dengan *machine learning traditional.* Pada gambar 2.1 menunjukan perbedaan mendasar antara cara kerja *deep learning* dan *machine learning traditional* (Wang Senzhangand Cao, 2021). Pada *machine learning traditional*, pemrogram harus secara spesifik menentukan fitur atau pola yang harus dikenali komputer untuk menghasilkan keputusan atau prediksi. Sebaliknya, *deep*

learning mampu secara otomatis mengenali dan mengekstraksi fitur dari data, sehingga model dapat belajar dan membuat keputusan tanpa perlu intervensi manusia.



Gambar 2. 1 Perbedaan machine learning traditional dan *deep learning* (Wang Senzhangand Cao, 2021)

Algoritma *deep learning* paling populer untuk klasifikasi gambar yaitu *Convolutional Neural Networks* (CNN). Selain itu, ada *Vision Transformers* (ViT) yang terinspirasi dari kesuksesan nya pada *Natural Language Processing* (NLP), ViT bisa bekerja lebih baik dibanding CNN disebabkan mekanisme self-attention yang membuat seluruh isi gambar dapat diakses dari lapisan tertinggi hingga lapisan terendah (Maurício dkk., 2023b).

2.1.2 Computer Vision

Computer vision merupakan cabang ilmu komputer yang bertujuan untuk memberikan kemampuan kepada komputer agar dapat melihat dan memahami

informasi visual dari gambar atau video (Wiley & Lucas, 2018). Tujuan utamanya adalah mengembangkan sistem yang dapat meniru kemampuan visual manusia seperti menganalisis, mengenali, dan mengambil keputusan berdasarkan konten visual secara otomatis (Javaid dkk., 2024). *Computer vision* mencakup berbagai sub-domain yang dirancang untuk menangani beragam permasalahan visual, diantaranya meliputi:

a. Object Detection

Object detection / deteksi objek merupakan teknik untuk mengidentifikasi objek spesifik dalam gambar atau video sekaligus menentukan lokasinya menggunakan *bounding box* atau *mask* (Deng dkk., 2020). Pada implementasinya, sistem deteksi objek terdiri dari dua tahap, yaitu lokalisasi dan klasifikasi. Lokalisasi berfokus pada penetuan posisi objek dalam gambar dengan menghasilkan koordinat yang mempresentasikan area tempat objek tersebut berada. Sementara, klasifikasi bertugas mengelompokkan objek yang telah terdeteksi ke dalam kategori atau kelas tertentu. Kemajuan teknologi dalam bidang *deep learning*, terutama melalui arsitektur *Convolutional Neural Networks* (CNN) telah meningkatkan kemampuan deteksi objek secara signifikan.

b. Image segmentation

Image segmentation merupakan teknik membagi citra menjadi beberapa bagian atau kelompok piksel yang memiliki karakteristik serupa, sehingga lebih mudah untuk dianalisis (Zaitoun & Aqel, 2015). Teknik ini berperan penting dalam berbagai aplikasi visi komputer karena

memungkinkan komputer untuk mengekstraksi informasi yang lebih spesifik dari suatu gambar dengan memisahkan objek utama dari latar belakang atau elemen lain yang tidak relevan. Sistem komputer dapat mengenali struktur visual dengan lebih baik dan meningkatkan efisiensi dalam berbagai tugas analisis citra dengan segmentasi gambar.

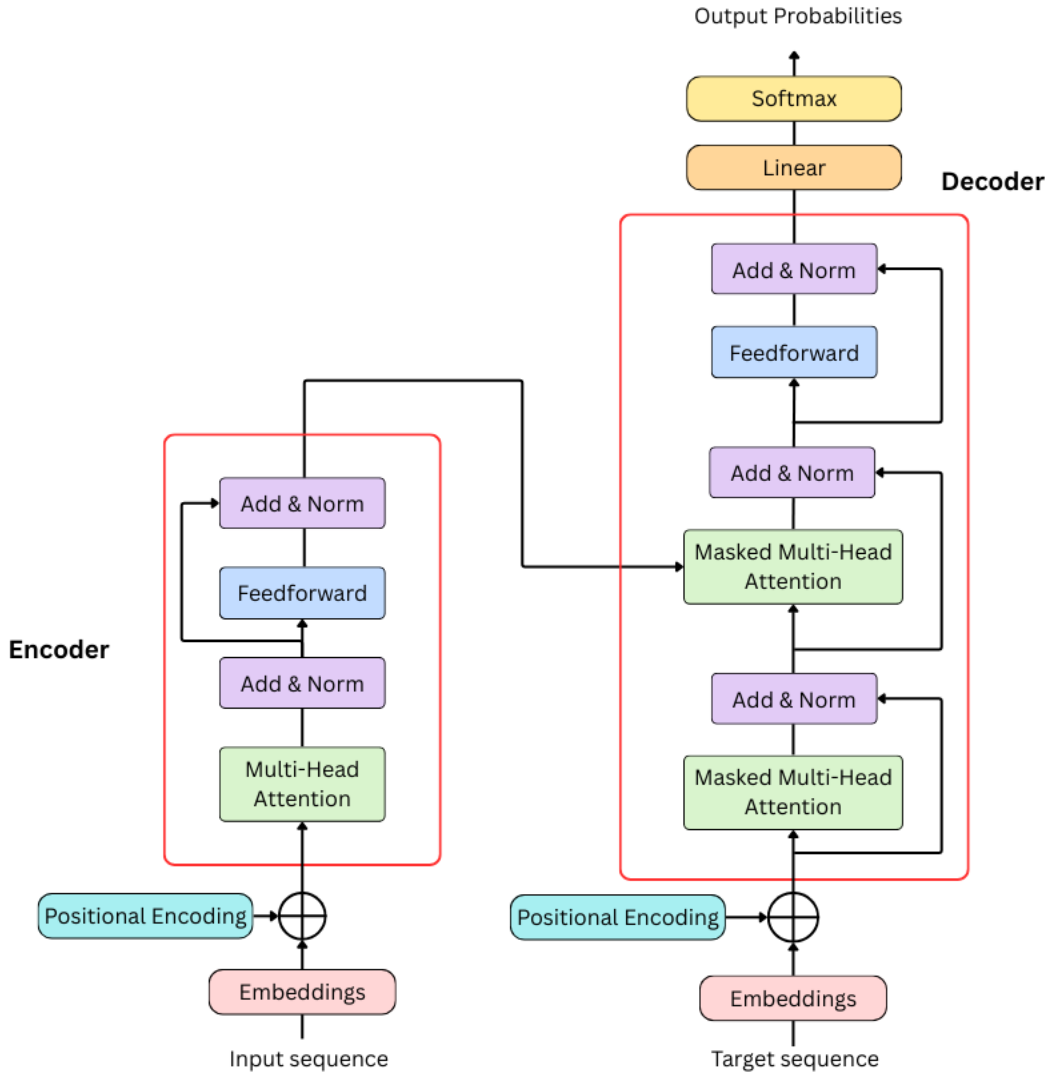
c. Image recognition

Image recognition atau pengenalan gambar adalah cabang dari visi komputer yang berfokus pada identifikasi dan klasifikasi objek, pola, atau fitur dalam gambar digital (X. Zhao, 2022). Teknologi ini memungkinkan komputer mengenali pola visual secara otomatis menggunakan teknik seperti ekstraksi fitur, machine learning, dan *deep learning*. Salah satu metode yang paling efektif dalam image recognition adalah Vision Transformer yang mampu mengekstrak fitur penting dari gambar dan melakukan klasifikasi dengan tingkat akurasi yang tinggi (Maurício dkk., 2023b).

2.1.3 Transformer

Transformer merupakan salah satu inovasi terbesar dalam bidang *Natural Language Processing* (NLP) yang telah merevolusi cara model kecerdasan buatan memahami dan menganalisis teks. Arsitektur ini dirancang untuk mengatasi keterbatasan model sebelumnya seperti *Long Short-Term Memory* (LSTM) dan *Recurrent Neural Networks* (RNN) yang memiliki kesulitan dalam menangani ketergantungan jarak jauh serta proses komputasi yang berurutan. Penggunaan *attention mechanism* khususnya *self-attention*, Transformer mampu memproses

data secara paralel, meningkatkan efisiensi dan akurasi dalam berbagai tugas pemrosesan bahasa (Vaswani dkk., 2017). Mekanisme *self-attention*



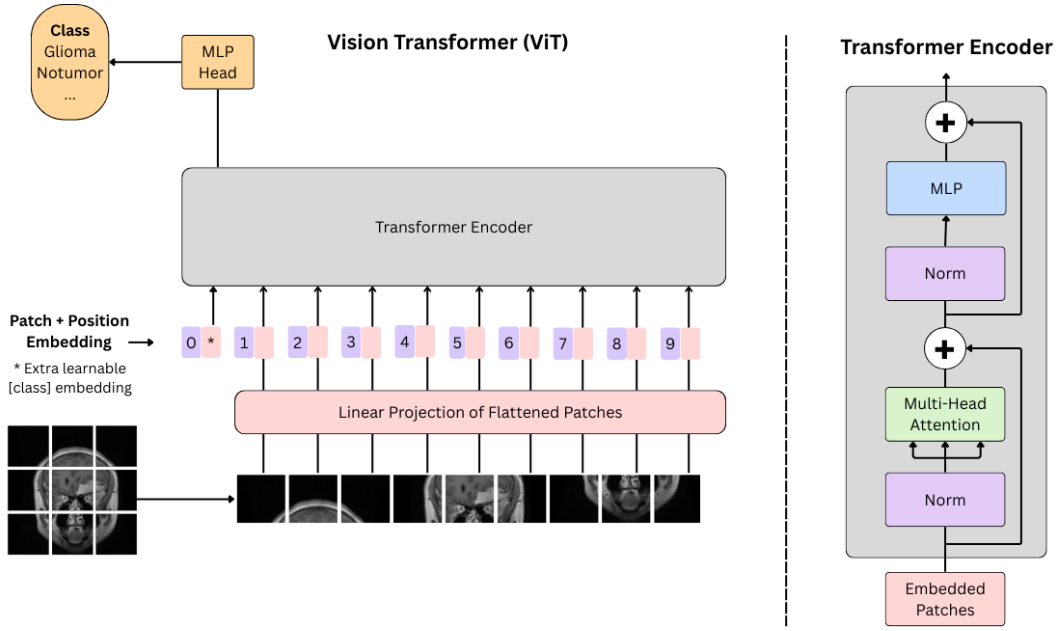
Gambar 2. 2 Arsitektur transformer (Vaswani dkk., 2017)

pada transformer berfungsi sebagai panduan model untuk fokus kepada satu kata yang akan diproses. Pada proses ini *attention function* akan melakukan mapping dari tiga vektor yaitu *query, key, dan value* ke dalam bentuk output. Hasil dari self-attention diperoleh melalui *weighted sum* dari nilai yang diberikan bobot berdasarkan relevansi setiap kata dalam suatu urutan. Arsitektur transformer dapat

dilihat pada gambar 2.2. Terdapat dua komponen utama dalam arsitektur transformer, yaitu *encoder* dan *decoder*. *Encoder* bertugas mengolah input teks dengan membangun representasi fitur yang kaya, sedang *decoder* menggunakan representasi tersebut untuk menghasilkan keluaran yang sesuai. Setiap komponen terdiri dari beberapa lapisan yang menerapkan multi-head self-attention dan feed-forward neural networks untuk menangkap hubungan kompleks dalam data. Pada bagian *decoder,* digunakan masked multi-head attention untuk memastikan model tidak melihat informasi dari token-token berikutnya selama proses pelatihan, sehingga dapat melakukan prediksi secara bertahap. Setelah itu, model menerapkan lapisan normalisasi, feed-forward layer, serta softmax layer untuk menentukan probabilitas output.

2.1.4 Vision Transformer

Transformers pertama kali diperkenalkan dalam konteks NLP oleh (Vaswani dkk., 2023) melalui arsitektur *encoder-decoder* yang mengandalkan *self-attention* untuk memproses data sekuensial. *Self-attention* memungkinkan model menghitung bobot interaksi antar elemen sekuens, menangkap dependensi jarak jauh tanpa rekurensi. Transformers menggantikan RNN/CNN dalam tugas seperti mesin terjemahan. Adaptasi ke domain visual dilakukan oleh (Dosovitskiy dkk., 2020) dalam *Vision Transformer* (ViT). Ide utamanya adalah memperlakukan citra sebagai rangkai *patch* (potongan gambar) yang mirip dengan token dalam NLP dapat dilihat pada gambar 2.2. Setiap patch diproyeksikan ke vektor dan diolah oleh transformer encoder, menggantikan operasi konvolusi pada CNN. Pendekatan ini memungkinkan model memahami hubungan global antar bagian gambar.



Gambar 2. 3 Arsitektur vision transformer (Dosovitskiy dkk., 2020)

a. Patch + Position Embedding (inputs)

Proses ini dibagi menjadi *patch* berukuran tetap seperti 16x16 piksel. Setiap patch diratakan (*flattened*) menjadi vektor 1D, menghasilkan urutan vektor seperti kata dalam NLP. *Position embedding* ditambahkan ke setiap patch untuk mempertahankan informasi spasial. *Position embedding* ini bersifat *learnable* atau dipelajari selama pelatihan dan memungkinkan model memahami hubungan spasial antar-patch. Selain itu, sebuah *class token* (vektor khusus) ditambahkan ke urutan input sebagai representasi global gambar untuk tugas klasifikasi.

$$z_0 = \left[x_{class};\ x_p^1 E;\ x_p^2 E;\ \ldots;\ x_p^N E\right] + E_{pos},\quad E \in R^{(P^2.C)\times D},\ E_{pos} R^{(N+1)\times D} \qquad (1)$$

Persamaan 1 berkaitan dengan token kelas, embedding patch dan embedding posisi. Dalam bentuk vektor, embedding terlihat seperti:

$$x_{input}$$
$$= \left[class_{token}, \quad image_{patch_1}, image_{patch2}, \dots\right]$$
$$+ \left[class_{toke_{position}}, image_{patch_{1_{position}}}, image_{patch_{2_{position}}}, \dots\right]$$

b. Linear Projection of Flattened Patches

Proyeksi linear mengubah patch x_p^i menjadi vektor dimensi D. operasi ini direpresentasikan oleh matriks E pada persamaan 1. Hasilnya adalah $x_p^i E$ yang menggabungkan antara informasi visual dan posisi.

c. Norm

Norm merupakan singkatan dari layer normalization sebuah teknik untuk mengurangi *overfitting*. Pada ViT, norm diterapkan sebelum *multi-head attention* (MHA) dan *multi-layer perceptron* (MLP) di setiap blok encoder.

d. Multi-Head Attention (MHA)

MHA adalah inti dari mekanisme *self-attention* pada ViT. Setiap *head* menghitung skor attention antara semua pasangan *patch* untuk menangkap ketergantungan global. Secara matematis, untuk setiap *head*, vektor *query* (Q), *key* (K), dan *value* (V) dihasilkan dari input. MHA didefinisikan pada persamaan 2.

$$MHA\ (Q, K, V) = Concat(head_1, \dots, head_h) \qquad (2)$$

Dimana $head_i = Attention\left(QW_i^Q, KW_i^K, VW_i^V\right)$

e. MLP (Multilayer perceptron)

MLP berisi dua layer linear dan fungsi aktivasi non-linear GELU (Gaussian Error Linear Unit). MLP beroperasi secara independen pada setiap token termasuk class token.

f. Transformer Encoder

Transformer encoder adalah kumpulan layer norm, MHA, dan MLP. Terdapat dua koneksi lompatan di dalam encoder transformer yaitu simbol "+" yang merepresentasikan *residual connection* yaitu operasi penjumlahan element-wise antara input awal suatu sub-lapisan (MHA atau MLP) dengan output sub-lapisan tersebut.

g. MLP Head

Setelah melalui encoder, *class token* diproses oleh MLP head untuk klasifikasi. MLP head biasanya terdiri dari lapisan linear dengan aktivasi, lapisan linear akhir yang memetakan ke jumlah kelas target.

2.2 Penelitian Terkait

Penelitian mengenai pengolahan gambar dalam raung lingkup kesehatan telah menggelamai perkembangan yang signifikan, sebagai hasil dari pesatnya kemajuan komputer. Beberapa studi telah mengeksplorasi pengolahan gambar menggunakan metode *machine learning* dan *deep learning* seperti CNN dan Vision Transformer. Kontribusi dan perbandingan antara penelitian sebelumnya pada tabel 2.1 adalah hasil ringkasan untuk membandingkan penelitian pemrosesan gambar dalam ruang lingkup kesehatan dengan menggunakan teknik *Deep learning*.

Tabel 2. 1 Penelitian terkait

Judul	Penulis	Algoritma	Akurasi	Fokus pembahasan
Enhanced Brain Tumor Classification through Gamma Correction in Deep Learning	(Naufal dkk., 2024b)	CNN	86.52% tanpa metode gamma correction, 88.49% dengan gamma correction	Image enhancement berguna untuk meningkatkan kualitas gambar untuk dikenali oleh *Computer-Aided Diagnosis* (CAD) *system*. Salah satu metode nya yaitu gamma correction, maka dilakukan perbandingan performa model CNN dengan dan tanpa *gamma correction*
A Hybrid Deep CNN-SVM Approach for Brain Tumor Classification	(Biswas & Islam, 2023a)	Deep CNN-SVM	96.0%	Mengembangkan model untuk klasifikasi tumor otak melalui Deep CNN dan SVM dengan pre-processed untuk meningkatkan proses ekstraksi dengan resize, *anisotropic diffusion filter* untuk noise reduction dan *adaptive histogram* equalization untuk contrast enhancement. Custom deep CNN dibangun untuk ekstraksi fitur yang mendalam

				kemudian *classifier SVM* diintegrasikan untuk tugas klasifikasi.
Brain Tumor Classification in MRI Images Using En-CNN	(Purnomo, 2021a)	En-CNN	95.5%	Mengusulkan metode baru yaitu en-CNN untuk klasifikasi tumor otak. Pendekatan ini berdasarkan pada VGG-16 arsitektur tetapi arsitektur yang lebih sederhana, dengan lebih layer dan parameter lebih sedikit.
Classification of Tumor in Brain MR Images Using Deep Convolutional Neural Network and Global Average Pooling	(Malla dkk., 2023a)	Deep CNN	98.93%	menerapkan metode *transfer learning* dengan menggunakan pre-trained VGGNet. Layer convolutional network di freeze untuk mendapat performa yang lebih baik dan menggunakan layer *Global Average Pooling* (GAP) pada output untuk menghindari permasalahan overfitting.
Efficient Brain Tumor Classification with a Hybrid CNN-SVM Approach in MRI	(Suryawanshi & Patil, 2024a)	CNN-SVM	98.01% pada dataset Brats dan 95.15% pada dataset Sartaj	Merancang model diusulkan dengan CNN untuk ekstraksi fitur dan SVM untuk klasifikasi. Selain juga dibuat model dengan *transfer learning* pre-trained yaitu VGG19 dan CNN untuk dibandingkan hasilnya.
Exploring the Power of Deep Learning: Fine-Tuned Vision Transformer for Accurate and Efficient Brain Tumor Detection in MRI Scans	(Asiri, Shaf, Ali, Shakeel, dkk., 2023b)	Fine-Tuned Vision Transformer	98.13%	Mengembangkan dan mengevaluasi model yang akurat untuk identifikasi tumor otak dari gambar dengan fine-tuned model Vision Transformer dan mengeksplorasi penggunaan *transfer learning*, data augmentation, dan teknik lain untuk meningkatkan performa dan efisiensi model.
Brain tumor classification in VIT-B/16 based on relative position encoding and residual MLP	(Hong dkk., 2024)	VIT-B/16	91.36%	Untuk improve ViT pada datasets kecil digunakan *homomorphic filtering, histogram equalization,* dan *unsharp masking* untuk memperkaya dataset kecil, meningkatkan informasi dan meningkatkan generalisasi model.

Brain Tumor Detection and Classification Using an Optimized Convolutional Neural Network	(Aamir dkk., 2024)	CNN	97%	Pengembangan dan penerapan model CNN yang dioptimalkan untuk meningkatkan akurasi diagnosis tumor otak dengan menyempurnakan hyperparameter nya yaitu batch size, jumlah layer, learning rate, activation functions, strategi pooling, padding, dan filter size.
Brain Tumor Detection Using Magnetic Resonance Imaging and Convolutional Neural Networks	(Martínez-Del-Río-Ortega dkk., 2024)	CNN	97.5%	Pengembangan dan evaluasi model CNN untuk deteksi tumor otak pada citra MRI. Metode grid search digunakan untuk mengoptimalkan hyperparameters dan memastikan performa model terbaik.
MRI brain tumor detection using deep learning and machine learning approaches	(Anantharajan dkk., 2024)	EDN-SVM	97.93%	Pengembangan metode deteksi tumor otak yang menggabungkan teknik *deep learning* dan machine learning untuk meningkatkan akurasi diagnosis. Dengan memanfaatkan kombinasi ekstraksi fitur otomatis melalui CNN dan ekstraksi fitur berbasis machine learning dengan teknik *Gray Level Cooccurrence Matrix* (GLCM) untuk mendapatkan fitur tekstur dari citra.
Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images	(Rahman dkk., 2021b)	CNN	95.11%	Eksplorasi efek dari teknik *image enhancement* untuk deteksi covid-19 menggunakan X-ray images. Terdapat 5 teknik yang di uji coba yaitu *histogram equalization* (HE), *contrast limited adaptive histogram equalization* (GLAHE), *image complement*, *gamma correction,* dan *balance contrast enhancement technique* (BCET). Teknik *gamma correction* menjadi teknik paling baik performa nya dibanding teknik lainnya.
A vision transformer machine learning model for COVID-19 diagnosis using chest X-ray images	(Chen dkk., 2024)	Vision Transformer	95.79% untuk 4 kelass,	Menganalisis dan membandingkan performa 4 arsitektur yaitu vision transformer, EfficientNet, MViT, dan EfficientViT. Hasilnya model ViT yang

			99.79% untuk 3 kelas	diusulkan dengan memodifikasi dengan mengurangi blok encoder menjadi 10 menjadi model dengan performa paling baik dibanding model lainnya.
Advancing Brain Tumor Classification through Fine-Tuned Vision Transformers: A Comparative Study of Pre-Trained Models	(Asiri, Shaf, Ali, Pasha, dkk., 2023)	Vision Transformer	98.24%	Membandingkan klasifikasi gambar tumor otak menggunakan 5 model pre-trained ViT yaitu R50-ViT-l16, ViT-l16, ViT-l32, ViT-b16, dan ViT-b32. Model mendapatkan akurasi cukup tinggi mulai dari 90.31% - 98.24%. ViT-b32 memiliki akurasi tertinggi sebesar 98.24%.
Enhancing brain tumor detection in MRI with a rotation invariant Vision Transformer	(Krishnan dkk., 2024)	Rotation invariant ViT (RVIT)	98.60%	Memperkenalkan rotation invariant ViT (RViT), RViT mengintegrasikan rotated patch embeddings, yaitu teknik yang melibatkan rotasi patch citra sebelum pemrosesan oleh model untuk meningkatkan akurasi identifikasi tumor otak.
Deteksi Tumor Otak Melalui Gambar MRI Berdasarkan Vision Transformers dengan Tensorflow dan Keras	(Supriadi dkk., 2023)	ViT	88% dan 97.9%	Mengimplementasikan dan mengevaluasi kinerja arsitektur ViT dalam mendeteksi tumor otak melalui gambar MRI. Beberapa eksperimen yang melibatkan variasi jumlah dataset menghasilkan akurasi 88% pada dataset pertama dengan data gambar sebanyak 253 dan 97.9% pada dataset kedua dengan menggabungkan dua dataset dengan gambar sebanyak 3123.

Berdasarkan tabel 2.1 mengenai penelitian terkait selama 5 tahun terakhir telah banyak penelitian yang menggunakan *deep learning* untuk deteksi penyakit pada gambar MRI seperti deteksi tumor otak. Berbagai algoritma *machine learning* dan *deep learning* seperti SVM, CNN, dan ViT telah diuji dan dibandingkan. Pada tabel 2.2 dapat dilihat matriks penelitian terdahulu mengenai pengolahan citra di raung lingkup medis.

Tabel 2. 2 Matrix penelitian

Judul	**Peneliti**	**Algoritma**			**Teknik *Image enhancement***					
		SVM	CNN	ViT	*histogram equalization*	*contrast limited adaptive histogram equalization*	*image complement*	*gamma correction*	*balance contrast enhancement technique*	*homomorphic filtering*
Enhanced Brain Tumor Classification through Gamma Correction in Deep Learning	(Naufal dkk., 2024a)	-	V	-	-	-	-	V	-	-
A Hybrid Deep CNN-SVM Approach for Brain Tumor Classification	(Biswas & Islam, 2023b)	V	V	-	-	V	-	-	-	-
Brain Tumor Classification in MRI Images Using En-CNN	(Purnomo, 2021b)	-	V	-	-	-	-	-	-	-
Classification of Tumor in Brain MR Images Using Deep Convolutional Neural Network and Global Average Pooling	(Malla dkk., 2023b)	-	V	-	-	-	-	-	-	-

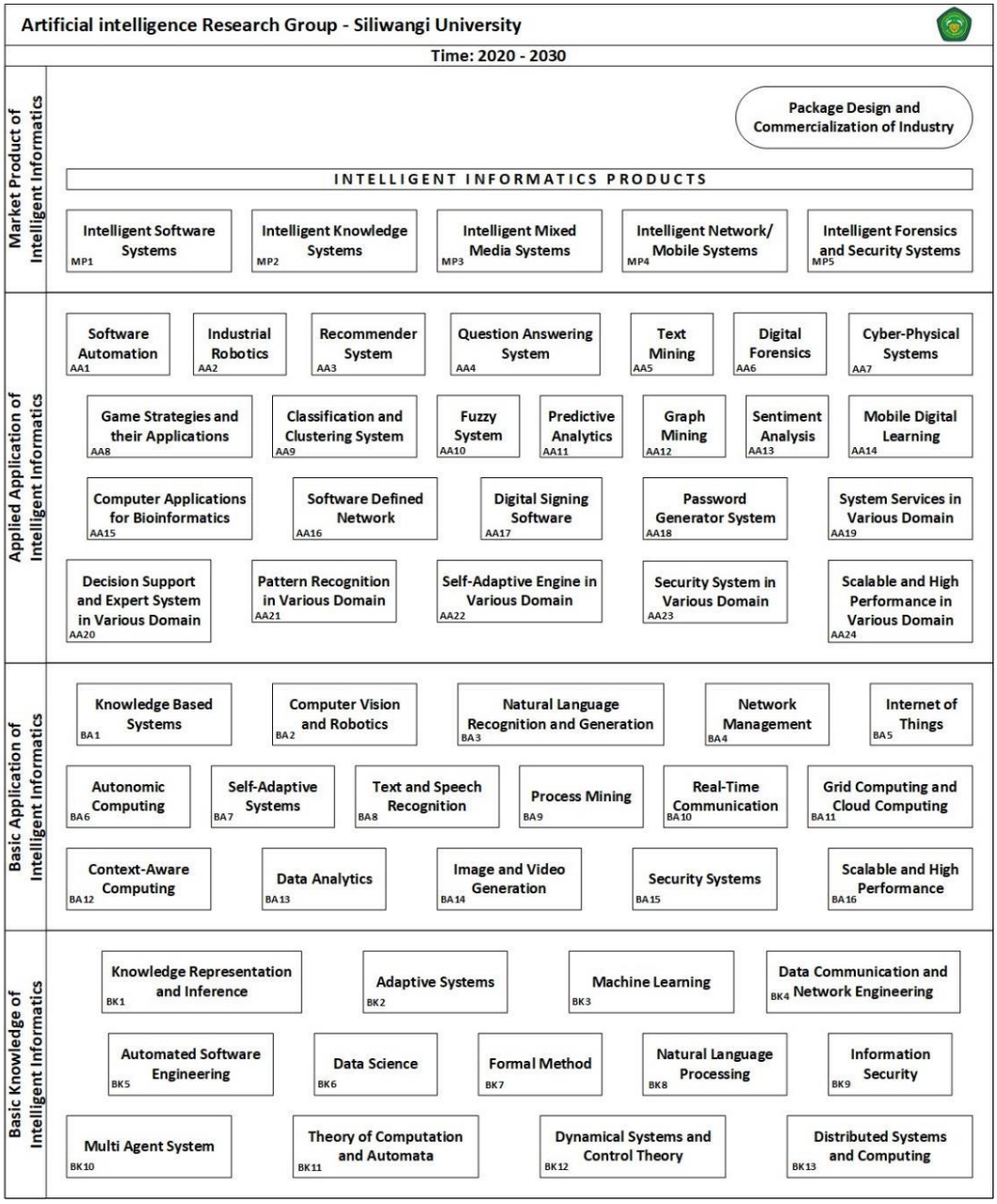
Efficient Brain Tumor Classification with a Hybrid CNN-SVM Approach in MRI	(Suryawanshi & Patil, 2024b)	V	V	-	-	-	-	-	-	-
Exploring the Power of Deep Learning: Fine-Tuned Vision Transformer for Accurate and Efficient Brain Tumor Detection in MRI Scans	(Asiri, Shaf, Ali, Shakeel, dkk., 2023b)	-	-	V	-	-	-	-	-	-
Brain tumor classification in VIT-B/16 based on relative position encoding and residual MLP	(Hong dkk., 2024)	-	-	V	V	V	-	-	-	V
Brain Tumor Detection and Classification Using an Optimized Convolutional Neural Network	(Aamir dkk., 2024)	-	V	-	-	-	-	-	-	-
Brain Tumor Detection Using Magnetic Resonance Imaging and Convolutional Neural Networks	(Martínez-Del-Río-Ortega dkk., 2024)	-	V	-	-	-	-	-	-	-
MRI brain tumor detection using deep learning and machine learning approaches	(Anantharajan dkk., 2024)	V	V	-	-	-	-	-	-	-
Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images	(Rahman dkk., 2021a)	-	V	-	V	V	V	V	V	-
A vision transformer machine learning model for COVID-19 diagnosis using chest X-ray images	(Chen dkk., 2024)	-	-	V	-	-	-	-	-	-
Advancing Brain Tumor Classification through Fine-Tuned Vision Transformers: A Comparative Study of Pre-Trained Models	(Asiri, Shaf, Ali, Pasha, dkk., 2023)	-	-	V	-	-	-	-	-	-

Enhancing brain tumor detection in MRI with a rotation invariant Vision Transformer	(Krishnan dkk., 2024)	-	-	V	-	-	-	-	-	-
Deteksi Tumor Otak Melalui Gambar MRI Berdasarkan Vision Transformers dengan Tensorflow dan Keras	(Supriadi dkk., 2023)	-	-	V	-	-	-	-	-	-
Optimasi Performa Vision Transformer dengan Pengurangan Jumlah Blok Encoder untuk Klasifikasi Tumor Otak	**(Zulfan dkk, 2025)**	-	-	V	-	-	-	V	-	-

Berdasarkan penelitian terkait yang telah disajikan sebelumnya dapat disimpulkan bahwa belum ada penelitian yang secara khusus dalam klasifikasi tumor otak dengan gambar MRI menggunakan arsitektur Vision Transformer dan menggunakan teknik *image enhancement* yaitu *gamma correction*. Dengan demikian, penelitian ini bertujuan untuk mengisi gap tersebut dengan mengusulkan model menggunakan arsitektur Vision Transformer. Selain meningkatkan akurasi klasifikasi, penelitian ini juga berfokus pada penyederhanaan arsitektur Vision Transformer dengan mengurangi jumlah blok encoder untuk mengurangi biaya komputasi dan penggunaan memori.

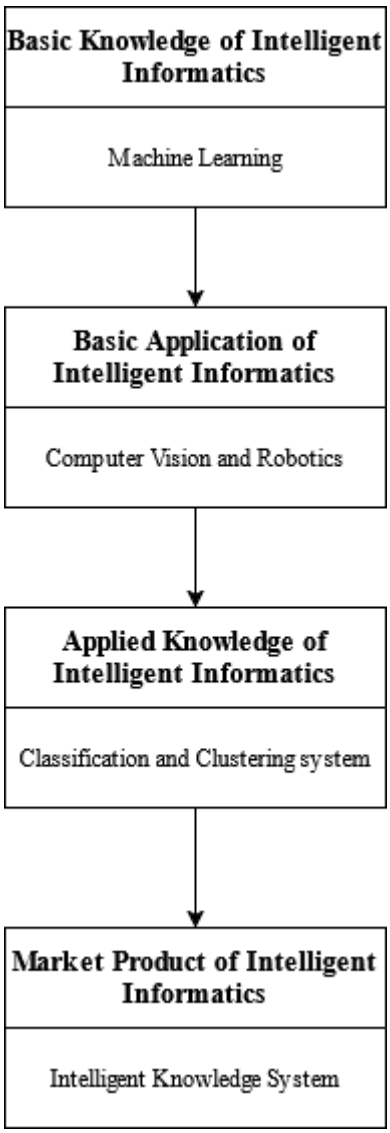
2.3 Roadmap Penelitian

Topik penelitian dalam skripsi ini merupakan bagian dari roadmap penelitian dari Kelompok Penelitian Artificial Intelligence Siliwangi (AIS), yang



Gambar 2. 4 Roadmap penelitian (AIS Universitas Siliwangi, 2020)

berada di bawah kelompok keahlian Informatika dan Sistem Inteligen (KK ISI) di Informatika Universitas Siliwangi. Kajian-kajian sebelumnya telah dijadikan landasan untuk penelitian yang diajukan dalam skripsi ini. Roadmap AIS ditunjukkan pada gambar 2.4 Topik utama penelitian yang akan dilakukan adalah *machine learning*, peta jalan penelitian ini dapat dilihat pada gambar 2.5.



Gambar 2. 5 Spesifikasi Roadmap Penelitian

Basis pengetahuan dalam penelitian ini mencakup *machine learning* yang berperan dalam pengembangan model *Vision Transformer* untuk klasifikasi tumor otak. Kemudian diimplementasikan pada tahap *computer vision* dan *robotics* yang relevan dalam pemrosesan gambar medis untuk mendeteksi atau mengklasifikasikan tumor secara otomatis. Selanjutnya penelitian ini beririsan dengan *classification and clustering system*, yang berfokus pada penggunaan model *deep learning* dalam klasifikasi jenis tumor berdasarkan karakteristiknya. Pada cakupan terakhir, yaitu *intelligent knowledge system,* penelitian ini diharapkan dapat menghasilkan sistem berbasis AI yang mampu memberikan keputusan diagnosis yang lebih akurat dan efisien. Dengan demikian, Optimasi *Vision Transformer* dalam klasifikasi tumor otak secara langsung berkaitan dengan seluruh tahapan, mulai dari dasar *machine learning* hingga dalam *intelligent knowledge system* yang diharapkan menjadi produk inovatif di bidang medis.