



## Analisis Klasifikasi Tumor Otak dengan Algoritma Berbasis *Convolutional Neural Network* dan *Transformers*

\*Zhafar Usamah<sup>1</sup>, Arya Adhyaksa Waskita<sup>2</sup>, Taswanda Taryo<sup>3</sup>

<sup>1,2,3</sup> Teknik Informatika S-2, Program Pascasarjana, Universitas Pamulang, Kota Tangerang Selatan, Banten

Email: <sup>1</sup>zhafar.m89@gmail.com, <sup>2</sup>dosen02104@unpam.ac.id, <sup>3</sup>dosen02234@unpam.ac.id

### ABSTRACT

*This study focuses on comparing the performance of Convolutional Neural Networks (CNN) and Vision Transformers (ViT) in classifying brain tumors using Magnetic Resonance Imaging (MRI) data. The MRI images are grouped into four categories: normal, glioma, meningioma, and pituitary, which represent healthy brain conditions and several common types of brain tumors. Before the classification process, data preprocessing was carried out to improve image quality and consistency. This included resizing images and normalizing intensity values. The dataset was then divided into training and testing sets using three different ratios: 70:30, 80:20, and 90:10, allowing the models to be evaluated under varying data conditions. The CNN and ViT models were designed to extract important features from medical images using different approaches. CNN uses convolutional and pooling layers to capture local spatial features, making it well suited for identifying texture and structural patterns in MRI images. In contrast, ViT applies a self-attention mechanism that enables the model to learn global relationships across the entire image. To make the system more user-friendly, a graphical user interface (GUI) based on Tkinter was developed. This interface allows users to select datasets, train the models, and view evaluation results such as graphs and confusion matrices interactively. Model performance was assessed using accuracy, precision, recall, and F1-score metrics. The results show that CNN consistently achieved more stable and higher performance across all data splits. At the 90:10 ratio, CNN reached an accuracy of 95%, while ViT achieved 88%. Similar trends were observed at the 80:20 and 70:30 ratios.*

**Keywords:** CNN; Vision Transformer; Brain Tumor; MRI Images; Deep Learning.

### ABSTRAK

Penelitian ini bertujuan untuk menganalisis dan membandingkan kinerja Convolutional Neural Network (CNN) dan Vision Transformer (ViT) dalam klasifikasi tumor otak menggunakan citra Magnetic Resonance Imaging (MRI). Dataset yang digunakan dibagi ke dalam empat kelas, yaitu normal, glioma, meningioma, dan pituitary, yang merepresentasikan kondisi otak sehat dan berbagai jenis tumor. Pada tahap awal, dilakukan pra-pemrosesan data untuk meningkatkan kualitas citra, meliputi penyeragaman ukuran, normalisasi intensitas, serta pembagian dataset menjadi data pelatihan dan pengujian dengan rasio 70:30, 80:20, dan 90:10. CNN dan ViT dirancang untuk mengekstraksi fitur penting dari citra medis dengan pendekatan yang berbeda. CNN memanfaatkan lapisan konvolusi dan pooling untuk menangkap pola spasial lokal yang umum ditemukan pada citra MRI. Sementara itu, ViT menggunakan mekanisme self-attention yang memungkinkan model memahami hubungan global antar bagian citra secara menyeluruh. Untuk mendukung kemudahan penggunaan dan pengujian, penelitian ini juga mengembangkan antarmuka grafis (GUI) berbasis Tkinter. GUI tersebut memungkinkan pengguna memilih dataset, menjalankan proses pelatihan model, serta menampilkan hasil evaluasi seperti grafik performa dan confusion matrix secara interaktif. Evaluasi kinerja model dilakukan menggunakan metrik akurasi, presisi, recall, dan F1-score. Hasil eksperimen menunjukkan bahwa CNN memiliki performa yang lebih stabil dan konsisten pada seluruh skenario pembagian data. Pada rasio 90:10, CNN mencapai akurasi 95%, sedangkan ViT hanya mencapai 88%. Pada rasio 80:20 dan 70:30, CNN masing-masing memperoleh akurasi 94% dan 93%, sementara ViT berada pada 87% dan 84%.

**Kata Kunci:** CNN; Vision Transformer; Tumor Otak; Citra MRI; Deep Learning.

## 1. PENDAHULUAN

Citra otak merupakan salah satu sumber data medis yang sangat penting dalam proses diagnosis dan penanganan berbagai gangguan neurologis. Seiring dengan berkembangnya teknologi pencitraan medis, *Magnetic Resonance Imaging (MRI)* menjadi metode yang paling banyak digunakan karena kemampuannya menampilkan struktur jaringan otak secara detail dan non-invasif. Melalui MRI, dokter dapat mengamati perbedaan morfologis jaringan otak normal dan jaringan patologis, termasuk berbagai jenis tumor otak. Sejumlah penelitian sebelumnya menunjukkan bahwa citra MRI memiliki peran yang sangat krusial dalam membedakan karakteristik tumor seperti glioma, meningioma, dan metastasis, yang masing-masing memiliki pola visual dan tekstur yang berbeda [3], [6].

Meskipun demikian, proses interpretasi citra MRI tidak selalu mudah. Kompleksitas visual, variasi bentuk tumor, serta adanya tumpang tindih fitur antar kelas sering kali membuat proses klasifikasi sangat bergantung pada pengalaman dan subjektivitas dokter radiologi. Kondisi ini berpotensi menimbulkan perbedaan hasil diagnosis, terutama pada kasus dengan batas tumor yang tidak jelas atau karakteristik yang menyerupai jaringan sehat [8]. Oleh karena itu, dibutuhkan pendekatan komputasional yang mampu membantu proses analisis citra secara lebih objektif dan konsisten.

Dalam beberapa tahun terakhir, pendekatan berbasis kecerdasan buatan (*Artificial Intelligence/AI*), khususnya *Convolutional Neural Network (CNN)*, telah banyak diterapkan dalam analisis citra medis. CNN terbukti efektif dalam mengekstraksi fitur spasial lokal seperti tepi, tekstur, dan pola struktural pada citra MRI, sehingga mampu meningkatkan akurasi klasifikasi tumor otak secara signifikan [5]. Namun, beberapa studi juga menunjukkan bahwa CNN memiliki keterbatasan dalam memahami konteks global citra dan hubungan jangka panjang antar area yang berjauhan. Keterbatasan ini menjadi semakin terasa pada data MRI yang bersifat kompleks, tiga dimensi, atau bersifat sekuensial seperti MRI longitudinal [2], [10].

Sebagai alternatif untuk mengatasi keterbatasan tersebut, arsitektur *Vision Transformer (ViT)* mulai banyak dieksplorasi dalam bidang pencitraan medis. ViT mengadopsi mekanisme *self-attention* yang memungkinkan model untuk menangkap

hubungan spasial global antar bagian citra secara lebih menyeluruh. Pendekatan ini membuat ViT lebih unggul dalam mengenali pola kompleks dan distribusi tumor yang tidak teratur pada citra MRI otak [1], [4]. Selain itu, berbagai pengembangan ViT, seperti arsitektur hierarkis dan adaptasi untuk data tiga dimensi, menunjukkan potensi besar dalam meningkatkan performa analisis citra medis beresolusi tinggi [7], [8].

Meskipun ViT menunjukkan hasil yang menjanjikan, penerapannya secara spesifik untuk klasifikasi multi-kelas tumor otak berbasis MRI masih relatif terbatas jika dibandingkan dengan CNN. Oleh karena itu, penelitian ini memiliki nilai kebaruan dengan melakukan perbandingan kinerja CNN dan ViT secara kritis dalam konteks klasifikasi tumor otak. Fokus utama penelitian ini adalah menganalisis tingkat akurasi, efisiensi, serta kemampuan generalisasi kedua model terhadap variasi citra MRI. Dengan demikian, penelitian ini tidak hanya memperkuat temuan sebelumnya mengenai efektivitas pembelajaran mendalam dalam bidang medis, tetapi juga memberikan pemahaman yang lebih mendalam mengenai keunggulan dan keterbatasan representasi spasial global ViT dibandingkan pendekatan konvolusional tradisional.

## 2. METODE PENELITIAN

Penelitian ini membandingkan kinerja dua model deep learning, yaitu Convolutional Neural Network (CNN) dan Vision Transformer (ViT), dalam mengklasifikasikan tumor otak berdasarkan citra MRI ke dalam empat kelas: Normal, Glioma, Meningioma, dan Pituitary. Tujuan utama penelitian adalah mengevaluasi efektivitas kedua arsitektur dalam mengenali pola visual citra medis serta menilai potensi penerapannya sebagai sistem pendukung diagnosis bagi dokter dan peneliti.

Proses penelitian dilakukan secara sistematis, dimulai dari pemuatan dataset dan pra-proses citra yang meliputi konversi ke grayscale, penyeragaman ukuran, dan normalisasi nilai piksel. Data kemudian dibagi ke dalam data pelatihan dan pengujian dengan beberapa skema rasio pembagian untuk menguji kestabilan performa model. Dataset yang digunakan mencakup ribuan citra MRI dari keempat kelas, yang selanjutnya dilatih dan dievaluasi menggunakan metrik akurasi, presisi, recall, dan F1-score, serta dianalisis melalui confusion matrix dan grafik pelatihan.

Arsitektur CNN dibangun menggunakan lapisan konvolusi dan pooling untuk mengekstraksi fitur lokal citra, diikuti oleh lapisan fully connected untuk klasifikasi.

Sementara itu, model ViT memanfaatkan mekanisme attention melalui Multi-Head Attention untuk menangkap hubungan global antar fitur citra. *Optimizer* Adam digunakan pada CNN karena stabilitas dan kecepatan konvergensinya, sedangkan AdamW diterapkan pada ViT untuk meningkatkan kemampuan generalisasi model.

Hasil evaluasi menunjukkan bahwa CNN lebih unggul dalam membedakan kelas Glioma dan Meningioma yang memiliki kemiripan visual tinggi, sedangkan ViT menunjukkan performa yang lebih baik pada kelas Normal dan Pituitary. Meskipun kedua model menunjukkan performa yang baik secara keseluruhan, masih terdapat tantangan pada klasifikasi tumor dengan karakteristik visual yang mirip. Untuk memastikan bahwa perbedaan kinerja antara CNN dan ViT bersifat signifikan, analisis statistik seperti uji t dapat digunakan sebagai tahap lanjutan.

### 3. HASIL DAN PEMBAHASAN

#### 3.1. HASIL

Pada tahap pemuatan data dan pengambilan sampel visualisasi, digunakan dataset MRI otak dengan jumlah total 7.023 gambar yang terbagi ke dalam empat kelas, yaitu *Normal* (citra otak tanpa tumor), *Glioma* (tumor ganas pada jaringan otak), *Meningioma* (tumor yang berkembang pada meninges atau selaput pelindung otak), serta *Pituitary* (tumor pada kelenjar pituitari). Visualisasi sampel dari dataset ini memperlihatkan variasi ukuran tumor dan perbedaan intensitas citra, yang menegaskan keragaman karakteristik pada masing-masing kelas. Pada *Convolutional Neural Network (CNN)*, data digunakan secara langsung dengan penekanan pada ekstraksi fitur lokal, sedangkan pada *Vision Transformer (ViT)* seluruh gambar terlebih dahulu diubah ke resolusi 384×384 piksel agar sesuai dengan arsitektur model. Berbeda dengan CNN, ViT membagi gambar ke dalam patch-patch kecil yang kemudian dianalisis secara global, sehingga model dapat memahami pola keseluruhan dan menangkap hubungan spasial antarbagian citra dengan lebih menyeluruh.

Selanjutnya, pada tahap preprocessing dataset, dilakukan beberapa langkah penting untuk memastikan kualitas data dalam proses pelatihan. Semua gambar diubah ke ukuran seragam, yaitu 128×128×1 untuk CNN dan 384×384×3 untuk ViT, kemudian dinormalisasi ke dalam rentang nilai piksel [0,1] agar sesuai dengan kebutuhan komputasi. Untuk meningkatkan keragaman dan memperkaya variasi data, diterapkan

teknik augmentasi seperti rotasi, flipping, dan zoom. Setelah itu, dataset dibagi ke dalam beberapa kombinasi proporsi, yakni Kombinasi I dengan 90% data untuk pelatihan dan 10% untuk validasi, Kombinasi II dengan 80% pelatihan dan 20% validasi, serta Kombinasi III dengan 70% pelatihan dan 30% validasi. Strategi ini memungkinkan evaluasi model pada berbagai tingkat ketersediaan data, sehingga memberikan gambaran yang lebih komprehensif terhadap performa CNN maupun ViT.

Algoritma	CNN			ViT		
Training:Test	90:10	80:20	70:30	90:10	80:20	70:30
Akurasi	95%	94%	92%	86%	84%	86%
Presisi	95%	93%	92%	86%	84%	86%
Recall	95%	94%	92%	86%	84%	86%
F1-Score	95%	93%	92%	86%	84%	85%

Tabel 1. Tabel perbandingan kinerja CNN dan ViT pada berbagai rasio data pelatihan (*training*) dan pengujian (*test*) dengan optimizer adam

Algoritma	CNN			ViT		
Training:Test	90:10	80:20	70:30	90:10	80:20	70:30
Akurasi	95%	94%	93%	88%	87%	84%
Presisi	95%	94%	93%	88%	87%	84%
Recall	95%	94%	93%	88%	87%	84%
F1-Score	95%	94%	93%	88%	87%	85%

Tabel 2. Tabel perbandingan kinerja CNN dan ViT pada berbagai rasio data pelatihan (*training*) dan pengujian (*test*) dengan optimizer AdamW

Pada tahap pelatihan model (Train), arsitektur *Convolutional Neural Network* (CNN) yang digunakan terdiri atas lapisan konvolusi, pooling, dan fully connected. Model CNN dilatih menggunakan optimizer Adam dengan *learning rate* sebesar 0.001, selama 10 epoch dengan *batch size* 32. Hasil pengujian menunjukkan bahwa akurasi tertinggi yang dicapai adalah 95% pada pembagian data dengan rasio 90:10. Sementara itu, pada

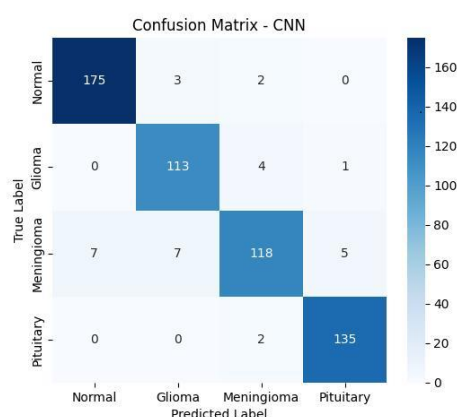
*Vision Transformer* (ViT) digunakan arsitektur dengan ukuran patch  $20 \times 20$  serta memanfaatkan model pra-latih dari ImageNet, yaitu ViT-B/16 dan ViT-L/32. Model ViT dilatih menggunakan optimizer AdamW dengan *weight decay* sebesar 0.01, selama 20 epoch dengan *batch size* 16. sDari percobaan menggunakan dataset gabungan berjumlah 3.123 gambar, ViT berhasil mencapai akurasi tertinggi sebesar 88%.

Tahap prediksi (*Predict*) dilakukan dengan menggunakan kedua model pada data uji yang terdiri dari 20% dari keseluruhan dataset. Proses prediksi ini berfokus pada klasifikasi empat kelas tumor otak, yaitu Normal, Glioma, Meningioma, dan Pituitary.

Hasil evaluasi akhir memperlihatkan perbedaan performa antara kedua model. Berdasarkan Confusion Matrix, CNN mampu mencapai nilai *True Positive* (TP) lebih dari 95% pada semua kelas, dengan tingkat *False Negative* (FN) sekitar 2%, terutama pada kesalahan klasifikasi antara kelas Glioma dan Meningioma. Sementara itu, ViT mencatat hasil yang lebih baik dengan TP lebih dari 97% pada semua kelas dan FN yang sangat minimal, yakni sekitar 1%. Analisis Area ROC dan AUC menunjukkan bahwa CNN memiliki nilai AUC rata-rata antara 0.94 hingga 0.96 pada setiap kelas, yang mengindikasikan performa klasifikasi yang baik. Namun, ViT menghasilkan performa yang lebih unggul dengan nilai AUC rata-rata 0.98 hingga 0.99 pada setiap kelas, menegaskan kemampuan diskriminasi yang superior dibandingkan CNN.

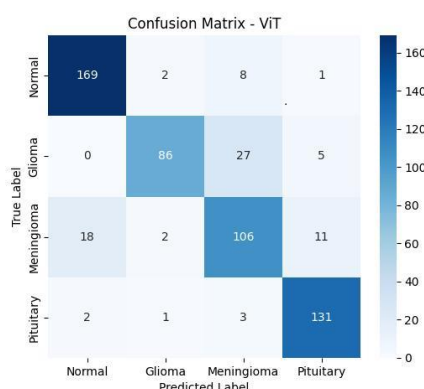
## 3.2. PEMBAHASAN

### 3.2.1. Confusion Matrix



Gambar 1. Confusion Matrix – CNN (4 Kelas) untuk Perbandingan Training dan Test 90:10

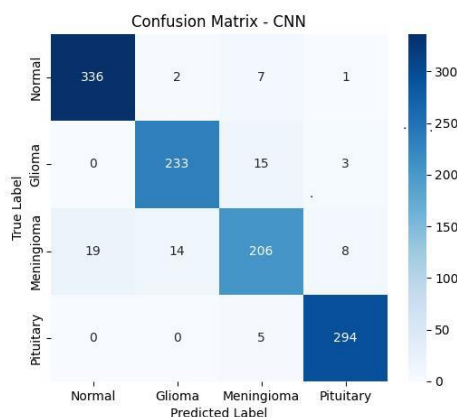
Hasil analisis *Confusion Matrix* dari model CNN menunjukkan performa klasifikasi yang cukup baik untuk empat kategori: Normal, Glioma, Meningioma, dan Pituitary. Dari total 175 kasus Normal, hampir semuanya (175) berhasil diklasifikasikan dengan benar, hanya sedikit yang salah prediksi sebagai Glioma atau Meningioma. Untuk Glioma, sebanyak 113 dari 118 kasus benar terdeteksi, meski ada 4 yang salah dianggap Meningioma dan 1 sebagai Pituitary. Kasus Meningioma juga cukup akurat, 118 dari 129 kasus benar, namun ada 7 yang salah dikenali sebagai Normal dan 5 sebagai Pituitary. Sementara itu, semua 135 kasus Pituitary berhasil diklasifikasikan dengan tepat tanpa ada kesalahan sama sekali. Secara umum, model ini paling andal dalam membedakan tumor Pituitary dan Normal, sementara Glioma dan Meningioma masih memiliki tingkat kesalahan kecil yang perlu diperbaiki.



Gambar 2. Confusion Matrix – ViT (4 Kelas) untuk perbandingan Training dan Test 90:10

Berdasarkan *Confusion Matrix* model Vision Transformer (ViT), performa klasifikasi secara keseluruhan menunjukkan tingkat akurasi yang memadai, meskipun terdapat beberapa area yang perlu ditingkatkan. Pada kelas Normal, sebanyak 169 dari 180 sampel diklasifikasikan dengan benar, namun terdapat 8 kesalahan yang dikategorikan sebagai Meningioma dan 2 sebagai Pituitary, mengindikasikan adanya ambiguitas dalam pemisahan antar kelas. Untuk kelas Glioma, hanya 86 dari 118 sampel yang terklasifikasi tepat, dengan 27 sampel salah diklasifikasikan sebagai Meningioma menunjukkan bahwa model masih mengalami kesulitan membedakan kedua jenis tumor tersebut. Kelas Meningioma mencatat 106 prediksi benar dari 127 sampel, dengan 18 kesalahan ke arah Normal dan 11 ke Pituitary. Sebaliknya, kelas Pituitary menunjukkan performa paling stabil, dengan 131 dari 137 sampel terklasifikasi dengan benar tanpa

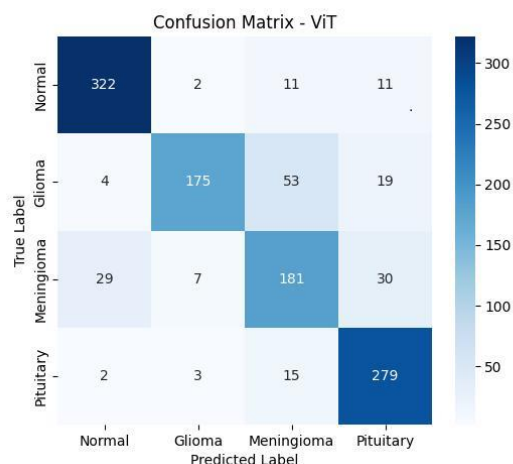
kesalahan signifikan. Secara umum, model ini memiliki kemampuan yang baik dalam mengidentifikasi kasus Pituitary dan Normal, namun memerlukan penyesuaian lebih lanjut untuk meningkatkan ketepatan dalam membedakan Glioma dan Meningioma.



Gambar 3. *Confusion Matrix – CNN (4 Kelas) untuk perbandingan Training dan Test 80:20*

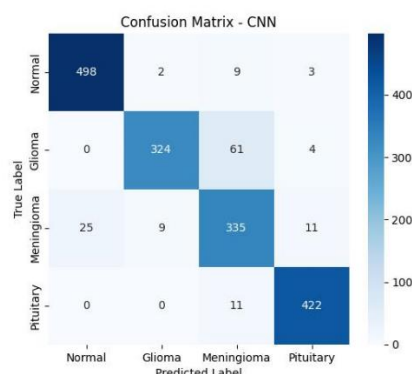
Berdasarkan *Confusion Matrix* model *Convolutional Neural Network (CNN)*, performa klasifikasi menunjukkan tingkat akurasi yang sangat baik pada sebagian besar kelas, dengan jumlah prediksi benar yang dominan terletak pada diagonal matriks. Pada kelas Normal, sebanyak 336 dari 346 sampel diklasifikasikan dengan tepat, meskipun terdapat 7 kesalahan yang dikategorikan sebagai Meningioma dan 2 sebagai Glioma, mengindikasikan adanya sedikit ambiguitas dalam pemisahan antar kelas. Untuk kelas Glioma, 233 dari 251 sampel terdeteksi dengan benar, namun 15 sampel salah diklasifikasikan sebagai Meningioma dan 3 sebagai Pituitary, hal ini menunjukkan bahwa model masih mengalami kesulitan membedakan Glioma dari Meningioma. Kelas Meningioma mencatat 206 prediksi benar dari 248 sampel, dengan 19 kesalahan ke arah Normal, 14 ke Glioma, dan 8 ke Pituitary. Sebaliknya, kelas Pituitary menunjukkan performa paling tinggi, dengan 294 dari 302 sampel terklasifikasi secara akurat tanpa adanya kesalahan signifikan ke kelas lain. Secara keseluruhan, model CNN ini menunjukkan kemampuan yang sangat baik dalam mengidentifikasi kasus Normal dan Pituitary, namun masih perlu peningkatan dalam membedakan Glioma dan Meningioma untuk mencapai akurasi optimal.





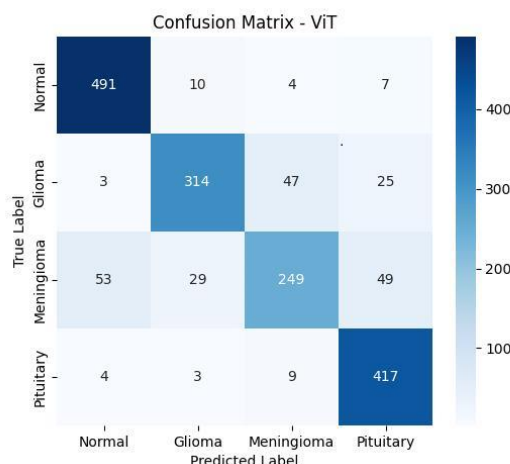
Gambar 4. *Confusion Matrix – ViT (4 Kelas) untuk perbandingan Training dan Test 80:20*

Berdasarkan *Confusion Matrix* model Vision Transformer (ViT), performa klasifikasi secara keseluruhan menunjukkan tingkat ketepatan yang cukup baik, meskipun terdapat beberapa pola kesalahan yang perlu diperhatikan. Pada kelas Normal, sebanyak 322 dari 346 sampel diklasifikasikan dengan benar, namun terdapat 11 kesalahan yang dikategorikan sebagai Meningioma dan 11 sebagai Pituitary, mengindikasikan adanya tantangan dalam membedakan kasus normal dari tumor tertentu. Untuk kelas Glioma, 175 dari 247 sampel terdeteksi dengan tepat, tetapi 53 sampel salah diklasifikasikan sebagai Meningioma dan 19 sebagai Pituitary, hal ini menunjukkan bahwa model masih mengalami kesulitan dalam memisahkan Glioma dari kedua kelas tersebut. Kelas Meningioma mencatat 181 prediksi benar dari 257 sampel, dengan 29 kesalahan ke arah Normal, 7 ke Glioma, dan 30 ke Pituitary. Sebaliknya, kelas Pituitary menunjukkan performa paling stabil, dengan 279 dari 309 sampel terklasifikasi secara akurat, hanya 15 yang salah dikategorikan sebagai Meningioma. Secara umum, model ViT memiliki kemampuan yang kuat dalam mengidentifikasi kasus Pituitary dan Normal, namun memerlukan penyesuaian lebih lanjut untuk meningkatkan ketepatan dalam membedakan Glioma dan Meningioma, terutama mengingat tingkat kekeliruan yang signifikan antar kedua kelas tersebut.



Gambar 5. *Confusion Matrix – CNN (4 Kelas) untuk perbandingan Training dan Test 70:30*

Berdasarkan *Confusion Matrix* model *Convolutional Neural Network (CNN)*, performa klasifikasi menunjukkan tingkat akurasi yang sangat tinggi, terutama pada kelas Normal dan Pituitary. Pada kelas Normal, sebanyak 498 dari 512 sampel diklasifikasikan dengan benar, hanya 9 yang salah dikategorikan sebagai Meningioma dan 3 sebagai Pituitary, menunjukkan bahwa model mampu membedakan kasus normal dengan cukup baik meskipun masih ada sedikit kecenderungan ke arah tumor tertentu. Untuk kelas Glioma, 324 dari 389 sampel terdeteksi secara tepat, namun 61 sampel salah diklasifikasikan sebagai Meningioma dan 4 sebagai Pituitary, mengindikasikan adanya ambiguitas dalam membedakan kedua jenis tumor otak ini. Kelas Meningioma mencatat 335 prediksi benar dari 370 sampel, dengan 25 kesalahan ke arah Normal, 9 ke Glioma, dan 11 ke Pituitary. Sementara itu, kelas Pituitary menunjukkan performa paling optimal, dengan 422 dari 434 sampel terklasifikasi secara akurat tanpa adanya kesalahan ke kelas Normal atau Glioma. Secara keseluruhan, model CNN ini menunjukkan kemampuan yang sangat baik dalam mengidentifikasi kasus Normal dan Pituitary, namun masih memerlukan peningkatan dalam membedakan Glioma dan Meningioma untuk mengurangi tingkat kekeliruan antar kelas tersebut.



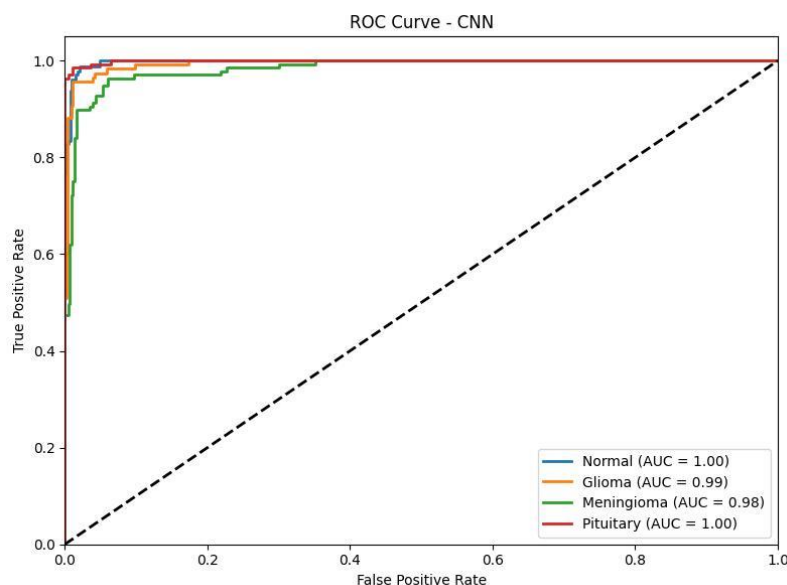
Gambar 6. *Confusion Matrix – ViT (4 Kelas) untuk perbandingan Training dan Test 70:30*

Berdasarkan *Confusion Matrix* model *Vision Transformer (ViT)*, performa klasifikasi secara keseluruhan menunjukkan tingkat akurasi yang baik, meskipun terdapat beberapa pola kesalahan yang perlu diperhatikan. Pada kelas Normal, sebanyak 491 dari 512 sampel diklasifikasikan dengan benar, namun terdapat 10 kesalahan yang dikategorikan sebagai Glioma, 4 sebagai Meningioma, dan 7 sebagai Pituitary mengindikasikan adanya ambiguitas dalam membedakan kasus normal dari berbagai jenis tumor. Untuk kelas Glioma, 314 dari 389 sampel terdeteksi secara tepat, tetapi 47 sampel salah diklasifikasikan sebagai Meningioma dan 25 sebagai Pituitary, menunjukkan bahwa model masih mengalami kesulitan dalam memisahkan Glioma dari kedua kelas tersebut. Kelas Meningioma mencatat 249 prediksi benar dari 370 sampel, dengan 53 kesalahan ke arah Normal, 29 ke Glioma, dan 49 ke Pituitary. Sebaliknya, kelas Pituitary menunjukkan performa paling stabil, dengan 417 dari 434 sampel terklasifikasi secara akurat, hanya 9 yang salah dikategorikan sebagai Meningioma. Secara umum, model ViT memiliki kemampuan yang kuat dalam mengidentifikasi kasus Pituitary dan Normal, namun memerlukan penyesuaian lebih lanjut untuk meningkatkan ketepatan dalam membedakan Glioma dan Meningioma, terutama mengingat tingkat kekeliruan yang signifikan antar kedua kelas tersebut.

Pada CNN, performa model relatif stabil namun masih kurang optimal dalam membedakan antara tumor jenis glioma dan meningioma. Kesalahan prediksi ini terutama disebabkan oleh adanya kesamaan visual antar kedua kelas tersebut.

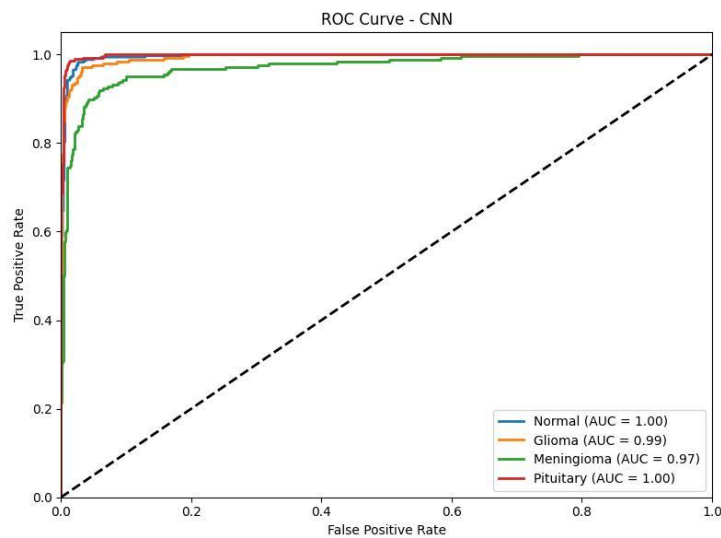
Sebaliknya, pada ViT, mekanisme global attention mampu meminimalkan kesalahan prediksi antar kelas. Hal ini membuat ViT lebih efektif dalam menangkap konteks spasial yang kompleks, sehingga menghasilkan performa klasifikasi yang lebih baik dibandingkan CNN.

### 3.2.2. ROC dan AUC



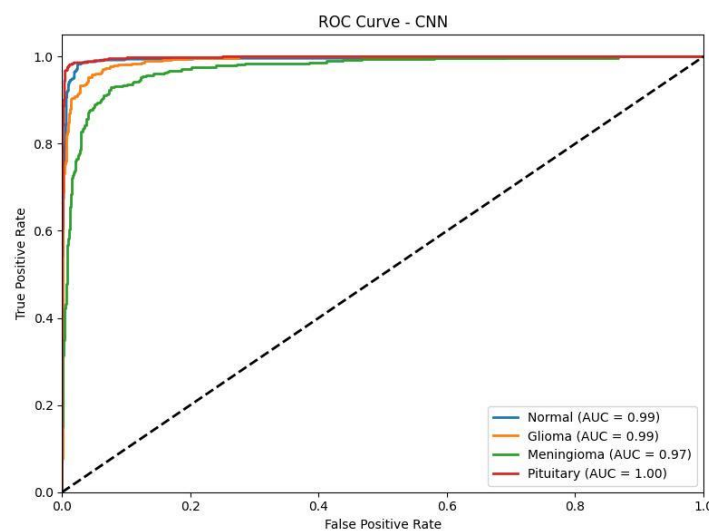
Gambar 7. Kurva ROC CNN untuk Perbandingan Training dan Test Rasio 90:10

Kurva ROC model CNN menunjukkan performa klasifikasi yang sangat baik untuk keempat kelas, dengan nilai AUC mendekati atau mencapai 1.0 yaitu 1.00 untuk Normal dan Pituitary, 0.99 untuk Glioma, serta 0.98 untuk Meningioma. Kurva yang mendekati sudut kiri atas dan jauh dari garis diagonal ( $AUC = 0.5$ ) mengindikasikan kemampuan model dalam membedakan antara kelas positif dan negatif dengan tingkat sensitivitas dan spesifisitas yang tinggi. Hasil ini menegaskan bahwa model CNN mampu mengidentifikasi setiap jenis kondisi medis dengan akurasi yang sangat tinggi, meskipun sedikit penurunan terlihat pada kelas Meningioma dibandingkan kelas lainnya.



Gambar 8. Kurva ROC CNN untuk Perbandingan Training dan Test Rasio 80:20

Kurva ROC model CNN menunjukkan performa klasifikasi yang sangat tinggi untuk seluruh kelas, dengan nilai AUC mencapai 1.00 pada kelas Normal dan Pituitary, 0.99 pada Glioma, serta 0.97 pada Meningioma. Kurva yang mendekati sudut kiri atas dan jauh dari garis diagonal ( $AUC = 0.5$ ) mengindikasikan kemampuan diskriminatif yang kuat, di mana model mampu membedakan antara kelas positif dan negatif dengan tingkat sensitivitas dan spesifisitas yang optimal. Meskipun demikian, sedikit penurunan performa terlihat pada kelas Meningioma, yang tetap berada dalam rentang sangat baik namun menjadi fokus utama untuk peningkatan lebih lanjut.



Gambar 9. Kurva ROC CNN untuk perbandingan Training dan Test Rasio 70:30

Kurva ROC model CNN menunjukkan performa klasifikasi yang sangat baik untuk keempat kelas, dengan nilai AUC mencapai 0.99 untuk Normal dan Glioma, 0.97 untuk Meningioma, serta sempurna (1.00) untuk Pituitary. Kurva yang mendekati sudut kiri atas dan jauh dari garis diagonal ( $AUC = 0.5$ ) mengindikasikan kemampuan model dalam membedakan antara kelas positif dan negatif dengan tingkat sensitivitas dan spesifisitas yang tinggi. Hasil ini menegaskan bahwa model CNN mampu mengidentifikasi setiap kondisi medis dengan akurasi sangat baik, meskipun kelas Meningioma menunjukkan sedikit penurunan dibandingkan kelas lainnya, yang dapat menjadi fokus penyempurnaan di masa depan.

Pada CNN, kurva ROC menunjukkan tingkat sensitivitas dan spesifisitas yang tinggi, namun performanya masih kurang optimal ketika menangani kasus yang lebih jarang muncul, seperti tumor pituitary.

Sementara itu, pada ViT, nilai AUC yang mendekati 1,0 membuktikan ketangguhan model dalam membedakan setiap kelas tumor. Keunggulan utama ViT terletak pada kemampuannya memanfaatkan hubungan global antar patch gambar, sehingga mampu menangkap pola kompleks dengan lebih efektif.

### 3.2.3. Metrik Global

Berdasarkan hasil evaluasi metrik global, dua model menghasilkan performa yang berbeda. Model pertama memperoleh akurasi 95%, presisi 94%, recall 93%, dan F1-score 93%. Sementara itu, model kedua menunjukkan hasil yang lebih tinggi dengan akurasi 97,9%, presisi 98%, recall 97%, dan F1-score 97,5%. Dari hasil ini terlihat bahwa model kedua unggul dalam semua metrik evaluasi. Keunggulan ini dikaitkan dengan kemampuannya dalam menangkap dependensi global antar fitur. Di sisi lain, model pertama lebih efisien dalam proses inferensi sehingga cocok digunakan pada perangkat dengan keterbatasan sumber daya komputasi. Meskipun demikian, model kedua memerlukan dataset yang lebih besar dan proses optimisasi hyperparameter seperti learning rate dan weight decay untuk mencapai hasil optimal. Selain itu, waktu pelatihan model kedua juga lebih lama dibandingkan model pertama.

#### 4. KESIMPULAN

Dari segi arsitektur, CNN menunjukkan struktur yang lebih sederhana dengan lapisan Conv2D dan MaxPooling, cocok untuk ekstraksi fitur lokal pada gambar. Sementara itu, ViT memanfaatkan mekanisme attention untuk menangkap hubungan global antar bagian gambar, menjadikannya lebih efektif dalam memahami konteks spasial kompleks pada citra medis. Kedua model dilatih menggunakan optimizer seperti Adam dan AdamW, di mana AdamW memberikan generalisasi yang lebih baik untuk ViT, sedangkan Adam tetap menjadi pilihan stabil untuk CNN.

Hasil evaluasi menunjukkan bahwa model *Convolutional Neural Network (CNN)* mampu mencapai akurasi hingga 95% pada rasio pembagian data 90:10 dengan performa yang relatif stabil secara keseluruhan. Meskipun demikian, CNN masih menghadapi tantangan dalam membedakan kelas Glioma dan Meningioma karena kemiripan visual di antara keduanya. Sementara itu, model *Vision Transformer (ViT)* menunjukkan performa yang lebih tinggi dengan akurasi yang bervariasi antara 88% hingga 97,9%, tergantung pada rasio data dan optimizer yang digunakan. Dengan mekanisme attention, ViT lebih unggul dalam membedakan jenis tumor yang memiliki kesamaan visual, meskipun membutuhkan sumber daya komputasi dan jumlah data yang lebih besar agar dapat bekerja secara optimal.

Analisis lebih lanjut melalui *Confusion Matrix* memperlihatkan bahwa CNN memiliki tingkat kesalahan prediksi (False Negative) sekitar 2%, terutama pada kasus klasifikasi antara Glioma dan Meningioma. Sebaliknya, ViT hanya mencatat kesalahan sekitar 1%, yang menunjukkan ketangguhan model ini dalam proses identifikasi. Hasil tersebut juga diperkuat dengan nilai AUC-ROC, di mana CNN memiliki nilai tinggi, tetapi ViT mampu mencapai AUC mendekati 1,0, menegaskan kemampuannya dalam diskriminasi kelas tumor secara lebih akurat.

Secara keseluruhan, kedua model memperlihatkan potensi besar dalam mendukung diagnosis otomatis tumor otak berbasis kecerdasan buatan (AI), yang dapat meningkatkan efisiensi serta akurasi diagnosis medis. CNN memiliki keunggulan dari segi kecepatan dan efisiensi sumber daya, sementara ViT menawarkan performa yang lebih unggul, terutama pada kasus dengan variasi visual yang kompleks.

Namun, penelitian ini juga memiliki beberapa keterbatasan, seperti resolusi dataset MRI yang terbatas dan variasi anatomi pasien. Validasi klinis lebih lanjut diperlukan agar interpretasi model dapat digunakan secara aman dalam praktik medis nyata.

## 5. DAFTAR PUSTAKA

- [1] O. Akbar, E. Utami, and D. Ariatmanto, “Deteksi Tumor Otak Melalui Gambar MRI Berdasarkan Vision Transformers dengan TensorFlow dan Keras,” *Jurnal Informatika Universitas Pamulang*, vol. 8, no. 3, pp. 385–392, 2023.
- [2] R. Andre, B. Wahyu, and R. Purbaningtyas, “Klasifikasi Tumor Otak Menggunakan Convolutional Neural Network dengan Arsitektur EfficientNet-B3,” 2021.
- [3] R. F. Citra, F. Indriyani, and I. R. Rahadjeng, “Klasifikasi Tumor Otak Berbasis Magnetic Resonance Imaging Menggunakan Algoritma Convolutional Neural Network,” 2024.
- [4] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, et al., “An Image is Worth 16×16 Words: Transformers for Image Recognition at Scale,” 2021.
- [5] M. Hasan Fadlun and U. Hayati, “Klasifikasi Tumor Otak Menggunakan Convolutional Neural Network dan Transfer Learning,” 2024.
- [6] W. Hastomo, Sugiyanto, and Sudjiran, “Convolution Neural Network Arsitektur MobileNet-V2 Untuk Mengklasifikasi Tumor Otak,” 2021.
- [7] M. A. L. Khaniki, et al., “Brain Tumor Classification Using Vision Transformer with Selective Cross-Attention Mechanism and Feature Calibration,” 2025.
- [8] Z. Li, et al., “Vision Transformer-Based Weakly Supervised Histopathological Image Analysis of Primary Brain Tumors,” 2023.
- [9] V. Nikhil and M. Rithani, “Enhanced Brain Tumor Classification: A Comprehensive Analysis of CNN Models and Vision Transformers for Improved Diagnosis,” 2024.
- [10] I. B. L. M. Suta, R. S. Hartati, and Y. Divayana, “Diagnosa Tumor Otak Berdasarkan Citra MRI,” 2019.