

Perbandingan Kinerja Model ResNet50V2 dan VGG16 untuk Klasifikasi Tumor Otak pada Citra MRI

Teuku Nadhif Al Fath^{1*}, Rahmat Sufri² Teuku Rizky Noviandy³

^{1*,2,3} Program Studi Sistem Informasi, Fakultas Teknik, Universitas Abulyatama, Aceh Besar, Aceh, Indonesia.

Corresponding Email: nadhifalfath6@gmail.com^{1*}

Histori Artikel:

Dikirim 21 Agustus 2026; Diterima dalam bentuk revisi 06 September 2026; Diterima 10 September 2026; Diterbitkan 28 September 2026. Semua hak dilindungi oleh Lembaga Penelitian dan Pengabdian Masyarakat (LPPM) STMIIK Indonesia Banda Aceh.

Abstrak

Tumor otak merupakan pertumbuhan sel abnormal pada jaringan otak yang dapat membahayakan keselamatan pasien apabila tidak terdeteksi sejak dini. Proses diagnosis umumnya dilakukan melalui pengamatan citra *Magnetic Resonance Imaging* (MRI) oleh radiolog, namun metode tersebut memerlukan waktu yang lama dan berpotensi menimbulkan subjektivitas. Penelitian ini bertujuan membandingkan kinerja dua arsitektur *Convolutional Neural Network* (CNN), yaitu ResNet50V2 dan VGG16, dalam mengklasifikasikan tumor otak pada citra MRI, serta menerapkan model dengan kinerja terbaik ke dalam prototipe aplikasi berbasis Streamlit. Dataset yang digunakan adalah *Brain Tumor MRI* yang terdiri atas empat kelas, yaitu glioma, meningioma, pituitary, dan *no tumor*. Kedua model dilatih dengan pendekatan *transfer learning* dan dievaluasi menggunakan metrik *accuracy*, *precision*, *sensitivity*, *specificity*, dan *F1-score*. Hasil penelitian menunjukkan bahwa ResNet50V2 memperoleh kinerja yang lebih tinggi dibandingkan dengan VGG16 pada seluruh metrik evaluasi, dengan *accuracy* 88,17% berbanding 86,19%. Keunggulan tersebut diduga berkaitan dengan mekanisme *residual connection* pada ResNet50V2 yang membantu mengatasi *vanishing gradient*, sedangkan VGG16 memiliki struktur yang lebih sederhana. Model dengan kinerja terbaik kemudian diimplementasikan sebagai prototipe aplikasi berbasis Streamlit untuk klasifikasi citra MRI tumor otak. Prototipe tersebut masih bersifat riset awal dan memerlukan pengujian lebih lanjut menggunakan data klinis serta evaluasi oleh tenaga medis sebelum dapat dipertimbangkan sebagai alat bantu diagnosis.

Kata Kunci: *Tumor Otak; MRI; Deep Learning; CNN; ResNet50V2; VGG16; Klasifikasi Citra.*

Abstract

Brain tumors are abnormal cell growths in brain tissue that can endanger patient safety if not detected early. Diagnosis is generally performed by radiologists through the examination of *Magnetic Resonance Imaging* (MRI) images; however, this approach requires considerable time and may introduce subjectivity. This study aims to compare the performance of two *Convolutional Neural Network* (CNN) architectures, namely ResNet50V2 and VGG16, in classifying brain tumors on MRI images, and to implement the best-performing model into a Streamlit-based application prototype. The dataset used was the *Brain Tumor MRI* dataset, which consists of four classes: glioma, meningioma, pituitary, and *no tumor*. Both models were trained using the *transfer learning* approach and evaluated using the metrics *accuracy*, *precision*, *sensitivity*, *specificity*, and *F1-score*. The results show that ResNet50V2 achieved higher performance than VGG16 across all evaluation metrics, with an *accuracy* of 88.17% compared with 86.19%. This advantage is likely related to the *residual connection* mechanism in ResNet50V2, which helps address the *vanishing gradient* problem, whereas VGG16 has a simpler architecture. The best-performing model was then implemented as a Streamlit-based application prototype for classifying brain tumor MRI images. However, the prototype remains at an early research stage and requires further testing using clinical data and evaluation by medical professionals before it can be considered a diagnostic support tool.

Keyword: *Brain Tumor; MRI; Deep Learning; CNN; ResNet50V2; VGG16; Image Classification.*

1. Pendahuluan

Tumor otak merupakan pertumbuhan sel abnormal di dalam atau di sekitar jaringan otak yang berlangsung secara tidak terkendali dan dapat menimbulkan dampak serius terhadap fungsi neurologis pasien. Kondisi tersebut menjadi salah satu tantangan besar dalam dunia medis karena tingkat mortalitasnya cenderung tinggi apabila tidak terdeteksi dan ditangani sejak awal (Rasheed *et al.*, 2023). Deteksi dini menjadi faktor penting dalam meningkatkan peluang keberhasilan pengobatan, mengurangi risiko komplikasi, serta membantu tenaga medis menentukan strategi penanganan yang tepat. Salah satu pendekatan utama dalam proses diagnosis tumor otak adalah melalui pemeriksaan citra medis, seperti X-Ray, *Computed Tomography* (CT) Scan, dan *Magnetic Resonance Imaging* (MRI). Di antara ketiga modalitas tersebut, MRI lebih sering digunakan karena mampu menghasilkan citra dengan resolusi tinggi dan kontras jaringan yang baik, sehingga perbedaan antara jaringan normal dan jaringan abnormal dapat diamati dengan lebih jelas (Ahmed *et al.*, 2024).

Meskipun MRI memiliki keunggulan dalam menampilkan struktur jaringan otak, proses interpretasi hasil citra masih banyak bergantung pada pemeriksaan manual oleh radiolog. Metode manual tersebut memerlukan waktu yang cukup lama, terutama ketika jumlah citra yang harus dianalisis sangat banyak. Selain itu, interpretasi citra juga berpotensi dipengaruhi oleh subjektivitas pengamat, terutama pada kasus dengan karakteristik citra yang kompleks atau memiliki kemiripan visual antarkelas. Kondisi tersebut mendorong perlunya pendekatan otomatis yang mampu membantu proses analisis citra medis secara lebih cepat, konsisten, dan objektif (Rasheed *et al.*, 2023).

Perkembangan *deep learning* telah memberikan kontribusi besar dalam bidang analisis citra medis. Salah satu arsitektur yang paling banyak digunakan adalah *Convolutional Neural Network* (CNN), karena memiliki kemampuan untuk mengekstraksi fitur spasial secara otomatis dari citra tanpa memerlukan perancangan fitur secara manual. Kemampuan tersebut menjadikan CNN unggul dalam berbagai tugas klasifikasi citra, termasuk klasifikasi tumor otak, dan telah terbukti mampu meningkatkan akurasi serta konsistensi hasil dibandingkan metode konvensional (Wong *et al.*, 2025). Dalam penerapannya, CNN tidak hanya efisien dalam mengenali pola visual, tetapi juga mampu mempelajari representasi fitur yang lebih kompleks dari data latih yang tersedia.

Di antara berbagai arsitektur CNN yang telah banyak digunakan, VGG16 dan ResNet50V2 menjadi dua model yang cukup populer dalam penelitian klasifikasi citra medis. VGG16 dikenal memiliki struktur yang bertingkat dan sederhana dengan penggunaan kernel konvolusi berukuran 3×3 secara berulang. Struktur tersebut membuat model relatif mudah dipahami dan sering dijadikan pembandingan dalam berbagai penelitian. Namun, jumlah parameternya yang besar menyebabkan kebutuhan sumber daya komputasi menjadi tinggi, sehingga proses pelatihan dapat memerlukan waktu yang lebih lama (Özkaraca *et al.*, 2023; Sarshar *et al.*, 2024). Meski demikian, VGG16 tetap banyak digunakan karena performanya yang stabil pada berbagai jenis dataset dan kemampuannya untuk di-*fine-tune* dengan cukup baik.

Sementara itu, ResNet50V2 mengadopsi pendekatan *residual mapping* melalui mekanisme *skip connection* pada setiap blok jaringan. Arsitektur tersebut dirancang untuk mengatasi masalah *vanishing gradient* yang umum terjadi pada jaringan yang sangat dalam. Dengan adanya *skip connection*, informasi penting dapat mengalir lebih lancar selama proses pelatihan, sehingga model menjadi lebih stabil dan lebih efektif dalam mempelajari pola yang kompleks. Beberapa penelitian menunjukkan bahwa ResNet50V2 mampu memberikan kinerja yang tinggi pada klasifikasi tumor otak maupun pada tugas klasifikasi citra medis lainnya (Younis *et al.*, 2024). Karakteristik tersebut menjadikan ResNet50V2 sebagai salah satu model yang layak dipertimbangkan dalam analisis citra MRI tumor otak.

Walaupun kedua arsitektur tersebut sama-sama telah banyak digunakan, hasil kinerjanya pada klasifikasi tumor otak tidak selalu seragam. Perbedaan hasil sering dipengaruhi oleh variasi dataset, teknik praproses data, strategi augmentasi, pembagian data latih dan uji, serta parameter pelatihan yang digunakan. Sejumlah studi menunjukkan bahwa ResNet memiliki performa lebih baik pada

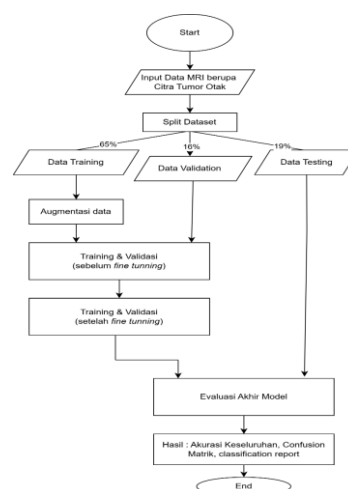
dataset MRI tertentu, sedangkan studi lainnya melaporkan bahwa VGG16 masih mampu memberikan hasil yang kompetitif karena strukturnya yang stabil dan proses *fine-tuning* yang relatif mudah (Huang & Prakash, 2025; Khan & Khan, 2025). Perbedaan tersebut menunjukkan bahwa perbandingan antarmodel tidak dapat dilakukan hanya berdasarkan hasil dari penelitian yang menggunakan dataset dan protokol evaluasi berbeda. Oleh karena itu, diperlukan penelitian komparatif yang menggunakan dataset yang sama, tahapan pelatihan yang seragam, serta metrik evaluasi yang konsisten agar hasil yang diperoleh dapat dibandingkan secara lebih objektif.

Pemilihan VGG16 dan ResNet50V2 pada penelitian ini didasarkan pada pertimbangan bahwa keduanya merupakan arsitektur CNN yang telah mapan, banyak digunakan sebagai *baseline* dalam literatur klasifikasi citra medis, dan mewakili dua pendekatan desain yang berbeda. VGG16 merepresentasikan arsitektur konvolusional klasik yang bertumpuk secara berurutan, sedangkan ResNet50V2 mewakili arsitektur modern berbasis *residual connection*. Perbandingan keduanya penting untuk mengetahui sejauh mana mekanisme residual mampu memberikan peningkatan kinerja dalam tugas klasifikasi tumor otak pada citra MRI. Selain itu, arsitektur yang lebih baru, seperti EfficientNet (Tan & Le, 2019), DenseNet (Huang *et al.*, 2017), maupun Vision Transformer (Dosovitskiy *et al.*, 2021), masih sangat potensial untuk diteliti pada studi lanjutan sebagai pembanding tambahan.

Berdasarkan uraian tersebut, penelitian ini bertujuan untuk mengimplementasikan dan membandingkan kinerja model ResNet50V2 dan VGG16 dalam mengklasifikasikan tumor otak pada citra MRI yang terdiri atas empat kelas, yaitu glioma, meningioma, pituitary, dan *no tumor*. Evaluasi dilakukan menggunakan beberapa metrik, yaitu *accuracy*, *precision*, *sensitivity*, *specificity*, dan *F1-score*. Selain itu, model dengan kinerja terbaik akan diterapkan ke dalam prototipe aplikasi berbasis Streamlit sebagai bentuk penerapan hasil penelitian. Dengan adanya penelitian ini, diharapkan diperoleh gambaran yang lebih jelas mengenai model yang paling sesuai untuk klasifikasi tumor otak berbasis citra MRI, sekaligus membuka peluang pengembangan sistem pendukung diagnosis yang lebih efektif pada tahap berikutnya.

2. Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen komparatif untuk membandingkan kinerja dua model *deep learning*, yaitu VGG16 dan ResNet50V2, dalam klasifikasi citra MRI otak berdasarkan metrik evaluasi *accuracy*, *precision*, *sensitivity*, *specificity*, dan *F1-score*. Tahapan penelitian secara ringkas disajikan pada Gambar 1.



Gambar 1. Alur Penelitian

Data yang digunakan berupa citra MRI otak yang bersumber dari *Brain Tumor MRI Dataset* yang dipublikasikan di Kaggle. Dataset tersebut terdiri atas empat kelas, yaitu glioma, meningioma, pituitary, dan *no tumor*. Dataset dibagi menjadi data latih dan data uji untuk masing-masing kelas, dengan rincian glioma (1.457 citra latih, 300 citra uji), meningioma (1.339 citra latih, 306 citra uji), pituitary (1.457 citra latih, 300 citra uji), dan *no tumor* (1.595 citra latih, 405 citra uji). Sebelum pelatihan, seluruh citra melalui tahap *preprocessing* berupa *resize* menjadi 224×224 piksel, normalisasi nilai piksel ke rentang 0–1, serta augmentasi data (*rotation*, *flip*, *zoom*, dan *shear*) untuk memperkaya variasi data dan mengurangi risiko *overfitting*. Tahapan tersebut sejalan dengan praktik *preprocessing* pada penelitian klasifikasi tumor otak serupa (Güler & Namlı, 2024; Rasa *et al.*, 2024).

Kedua model diterapkan menggunakan pendekatan *transfer learning* dengan bobot awal dari ImageNet. Beberapa lapisan awal dibekukan (*freeze*) untuk mempertahankan fitur dasar, sedangkan lapisan akhir disesuaikan dengan empat kelas keluaran. Proses pelatihan menggunakan parameter yang sama untuk kedua model guna menjaga konsistensi perbandingan, sebagaimana disajikan pada Tabel 1.

Tabel 1. Parameter Pelatihan Model CNN

Parameter	Nilai	Keterangan
Optimizer	Adam	Menyesuaikan laju pembelajaran secara adaptif untuk mempercepat konvergensi, sebagaimana umum digunakan pada penelitian klasifikasi tumor otak berbasis <i>transfer learning</i> (Rasa <i>et al.</i> , 2024).
Learning rate	0,0001	Dipilih sesuai praktik umum <i>fine-tuning</i> model <i>pretrained</i> pada <i>transfer learning</i> agar pembaruan bobot pada lapisan klasifikasi baru berlangsung bertahap dan stabil tanpa merusak fitur yang telah dipelajari model pada ImageNet.
Loss function	Categorical Crossentropy	Mengukur perbedaan distribusi prediksi dan label pada kasus multikelas.
Epoch	30	Jumlah iterasi penuh terhadap dataset pelatihan.

Dataset yang digunakan telah tersedia dalam struktur folder *Training* dan *Testing* sejak sumber aslinya. Folder *Training* kemudian dibagi lebih lanjut menjadi subset data latih dan data validasi menggunakan parameter *validation_split* sebesar 20% pada *ImageDataGenerator*, sedangkan folder *Testing* digunakan sepenuhnya sebagai data uji akhir. Secara keseluruhan, proporsi tersebut setara dengan sekitar 65% data latih, 16% data validasi, dan 19% data uji dari total dataset.

Data validasi digunakan untuk memantau performa model selama proses pelatihan guna mendeteksi indikasi *overfitting*, sedangkan data uji digunakan khusus untuk evaluasi akhir setelah pelatihan selesai dan tidak dilibatkan dalam proses pelatihan maupun penyesuaian *hyperparameter*. Proses pelatihan kedua model menggunakan *batch size* sebesar 32 dan dilakukan pada lingkungan Google Colaboratory dengan akselerasi GPU T4. Nilai *random seed* tidak ditetapkan secara eksplisit pada proses pelatihan, sehingga hasil pelatihan ulang berpotensi menunjukkan variasi kecil akibat inisialisasi bobot dan urutan data yang bersifat stokastik. Hal tersebut menjadi salah satu keterbatasan penelitian dari sisi *reproducibility*.

Pada kedua model, seluruh lapisan pada arsitektur dasar (*base model*) VGG16 dan ResNet50V2 dibekukan (*frozen*), sehingga bobot hasil pra-pelatihan pada ImageNet dipertahankan. Lapisan yang dilatih hanya lapisan klasifikasi baru yang ditambahkan di atasnya, yaitu lapisan *Dense* dengan 256 neuron dan aktivasi ReLU, lapisan *Dropout* sebesar 0,5, serta lapisan keluaran *Dense* dengan aktivasi *softmax*. Dengan demikian, pendekatan yang digunakan adalah *feature extraction*, bukan *fine-tuning* menyeluruh terhadap bobot *base model*.

Kedua model dilatih menggunakan mekanisme *EarlyStopping* yang memantau *validation loss* dengan *patience* 5 *epoch* dan mengembalikan bobot terbaik (*restore_best_weights*), dengan target maksimum 30 *epoch* untuk VGG16 dan 40 *epoch* untuk ResNet50V2. Perbedaan mekanisme tersebut menyebabkan jumlah *epoch* efektif pada hasil pelatihan berbeda dari target maksimum yang

ditetapkan, yaitu VGG16 berhenti pada *epoch* ke-15 dan ResNet50V2 pada *epoch* ke-21 sesuai kriteria penghentian masing-masing.

Evaluasi kinerja model dilakukan menggunakan *confusion matrix* serta metrik *accuracy*, *precision*, *sensitivity*, *specificity*, dan *F1-score*. *Accuracy* mengukur proporsi prediksi benar dari seluruh data uji; *precision* mengukur ketepatan prediksi kelas positif; *sensitivity* mengukur kemampuan model mengenali kelas positif; *specificity* mengukur kemampuan model mengenali kelas negatif; dan *F1-score* merupakan rata-rata harmonis antara *precision* dan *sensitivity* yang memberikan penilaian seimbang terhadap performa model, sebagaimana lazim digunakan pada evaluasi model *transfer learning* untuk klasifikasi tumor otak (Rasa *et al.*, 2024; Younis *et al.*, 2024).

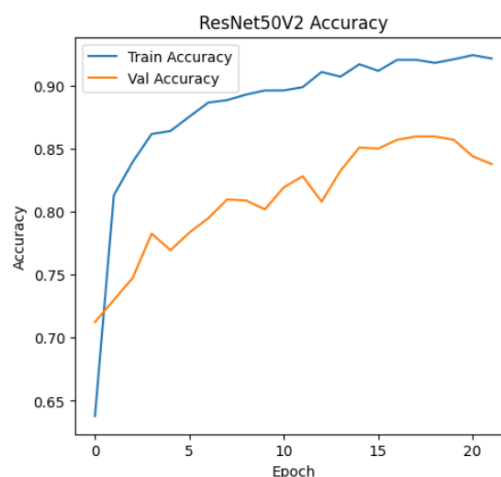
Mengingat klasifikasi yang dilakukan bersifat multikelas (empat kelas), nilai *precision*, *sensitivity*, *specificity*, dan *F1-score* yang dilaporkan merupakan hasil *macro average*, yaitu rata-rata tidak berbobot dari nilai keempat kelas. Dengan demikian, setiap kelas memberikan kontribusi yang setara terhadap nilai akhir tanpa dipengaruhi oleh perbedaan jumlah sampel antar kelas. Selain itu, dilakukan pula visualisasi *Gradient-weighted Class Activation Mapping* (Grad-CAM) untuk menginterpretasikan area citra yang menjadi dasar pengambilan keputusan model (Musthafa *et al.*, 2024; Yoon & Lin, 2025), serta implementasi model terbaik ke dalam aplikasi berbasis Streamlit sebagai antarmuka pengujian interaktif, mengikuti pendekatan serupa yang mengintegrasikan aplikasi web untuk klasifikasi tumor otak secara *real-time* (Wong *et al.*, 2025).

3. Hasil dan Pembahasan

3.1 Hasil

3.1.1 Kinerja Model ResNet50V2

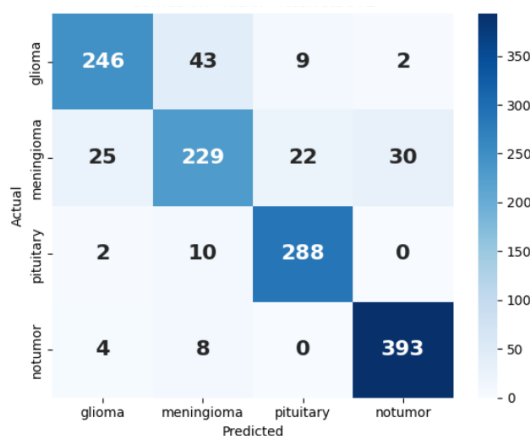
Model ResNet50V2 dilatih dengan maksimum 40 *epoch* menggunakan pendekatan *transfer learning* dari ImageNet. Namun, penerapan *Early Stopping* menghentikan proses pelatihan pada *epoch* ke-21 karena tidak terdapat peningkatan performa validasi yang memenuhi kriteria penghentian. Selama pelatihan, *training accuracy* meningkat dari sekitar 64% menjadi lebih dari 92%, sedangkan *validation accuracy* meningkat dari sekitar 71% hingga mencapai kisaran 85–86%, dengan sedikit fluktuasi pada beberapa *epoch* akhir. Pola tersebut menunjukkan bahwa model memiliki kemampuan generalisasi yang cukup stabil terhadap data validasi, meskipun masih terdapat sedikit jarak antara performa data pelatihan dan validasi pada *epoch* akhir.



Gambar 2. Grafik Akurasi Training dan Validation Model ResNet50V2

Evaluasi pada data uji menggunakan *confusion matrix* (Gambar 3) menunjukkan bahwa ResNet50V2 mampu mengklasifikasikan sebagian besar citra dengan benar pada keempat kelas.

Performa terbaik diperoleh pada kelas *no tumor* dengan 393 citra diklasifikasikan secara benar, diikuti kelas *pituitary* dengan 288 citra. Kesalahan klasifikasi paling banyak terjadi pada kelas *meningioma*, terutama citra yang diprediksi sebagai *glioma*, *pituitary*, dan *no tumor*. Selain itu, terdapat kesalahan klasifikasi antara kelas *glioma* dan *meningioma*. Hasil ini menunjukkan bahwa ResNet50V2 memiliki kemampuan klasifikasi yang baik, terutama pada kelas *no tumor* dan *pituitary*, tetapi masih menghadapi kesulitan dalam membedakan beberapa karakteristik citra *glioma* dan *meningioma*.



Gambar 3. Confusion Matrix Model ResNet50V2

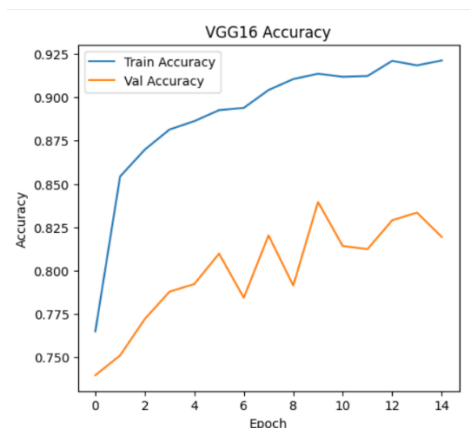
Model ResNet50V2 memperoleh accuracy sebesar 88,17%, precision 87,63%, sensitivity 87,46%, specificity 96,06%, dan F1-score 87,46%, sebagaimana disajikan pada Tabel 2. Hasil tersebut menunjukkan bahwa model mampu memberikan prediksi yang cukup akurat dan seimbang pada seluruh kelas, dengan tingkat kesalahan yang relatif rendah serta kemampuan identifikasi kelas positif yang baik.

Tabel 2. Evaluasi Performa Model ResNet50V2

Model	Accuracy	Precision	Sensitivity	Specificity	F1-score
ResNet50V2	88,17%	87,63%	87,46%	96,06%	87,46%

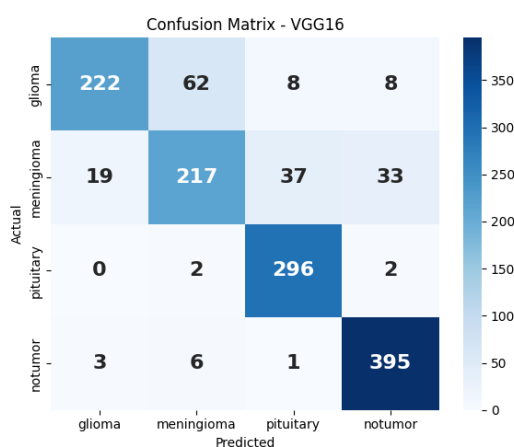
3.1.2 Kinerja Model VGG16

Model VGG16 dilatih dengan maksimum 30 epoch menggunakan pendekatan *transfer learning* dari ImageNet. Namun, penerapan *Early Stopping* menghentikan proses pelatihan pada epoch ke-15 karena tidak terdapat peningkatan performa validasi yang memenuhi kriteria penghentian. Selama pelatihan, *training accuracy* meningkat dari sekitar 76% menjadi lebih dari 92%, sedangkan *validation accuracy* meningkat dari sekitar 74% hingga mencapai kisaran 82–84% dengan beberapa fluktuasi pada epoch akhir. Perbedaan antara *training accuracy* dan *validation accuracy* yang semakin terlihat menunjukkan adanya kecenderungan *overfitting*, sehingga kemampuan generalisasi VGG16 terhadap data validasi lebih rendah dibandingkan ResNet50V2.



Gambar 4. Grafik Akurasi Training dan Validation Model VGG16

Evaluasi pada data uji menggunakan *confusion matrix* (Gambar 5) menunjukkan bahwa VGG16 mampu mengklasifikasikan sebagian besar citra dengan benar pada keempat kelas. Performa terbaik diperoleh pada kelas *no tumor* dengan 395 citra diklasifikasikan secara benar, diikuti kelas *pituitary* dengan 296 citra. Kesalahan klasifikasi paling banyak terjadi pada kelas *meningioma*, terutama citra yang diprediksi sebagai *pituitary* dan *no tumor*. Selain itu, terdapat kesalahan klasifikasi antara kelas *glioma* dan *meningioma*. Hasil ini menunjukkan bahwa VGG16 memiliki kemampuan klasifikasi yang baik, tetapi masih menghadapi kesulitan dalam membedakan beberapa karakteristik citra *glioma* dan *meningioma*.



Gambar 5. Confusion Matrix Model VGG16

Hasil evaluasi menunjukkan bahwa model VGG16 memperoleh *accuracy* 86,19%, *precision* 85,83%, *sensitivity* 85,27%, *specificity* 95,39%, dan *F1-score* 85,18% (Tabel 3), sedikit lebih rendah dibandingkan ResNet50V2 pada seluruh metrik.

Tabel 3. Evaluasi Performa Model VGG16

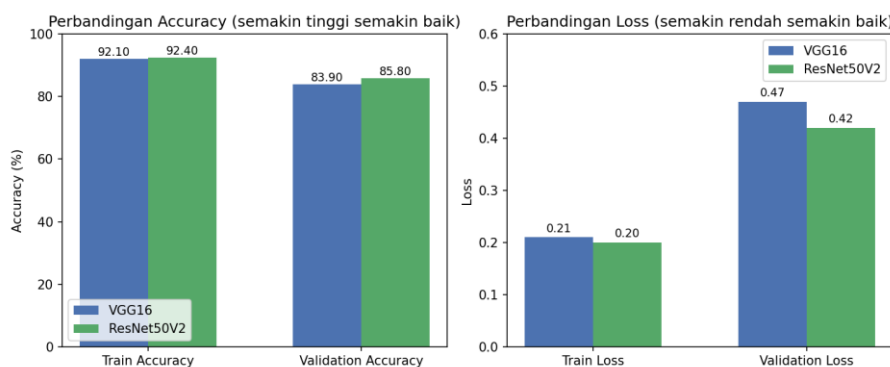
Model	Accuracy	Precision	Sensitivity	Specificity	F1-score
VGG16	86,19%	85,83%	85,27%	95,39%	85,18%

3.1.3 Perbandingan Kinerja Model

Perbandingan langsung kedua model pada seluruh metrik evaluasi disajikan pada Tabel 4. ResNet50V2 secara konsisten unggul pada seluruh metrik dibandingkan VGG16.

Tabel 4. Perbandingan Evaluasi Model ResNet50V2 dan VGG16

Model	Accuracy	Precision	Sensitivity	Specificity	F1-score
ResNet50V2	88,17%	87,63%	87,46%	96,06%	87,46%
VGG16	86,19%	85,83%	85,27%	95,39%	85,18%

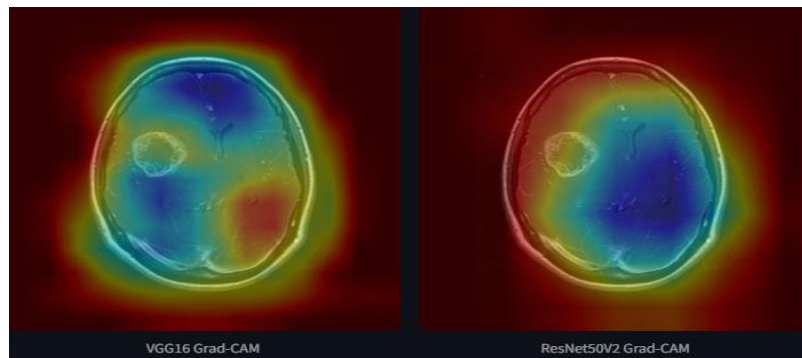


Gambar 6. Perbandingan Accuracy dan Loss Model VGG16 dan ResNet50V2 (Ditampilkan Terpisah)

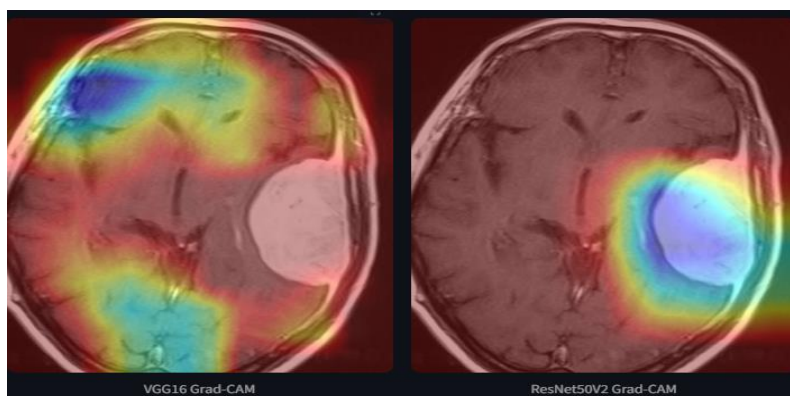
Perbandingan pada Gambar 6 sengaja ditampilkan sebagai dua panel terpisah, yaitu panel *accuracy* (semakin tinggi semakin baik) dan panel *loss* (semakin rendah semakin baik), karena kedua metrik tersebut memiliki arah interpretasi yang berlawanan sehingga penggabungannya dalam satu grafik *radar* berisiko menyesatkan pembaca. Grafik menunjukkan bahwa ResNet50V2 memperoleh *training accuracy* 92,40% dan *validation accuracy* 85,80%, dengan *training loss* 0,20 dan *validation loss* 0,42. Sementara itu, VGG16 memperoleh *training accuracy* 92,10% dan *validation accuracy* 83,90%, dengan *training loss* 0,21 dan *validation loss* 0,47.

Selisih *validation accuracy* dan *validation loss* yang lebih baik pada ResNet50V2 sejalan dengan karakteristik umum arsitektur *residual connection* yang dilaporkan mampu mengurangi degradasi performa pada jaringan dalam (Younis *et al.*, 2024; Sarshar *et al.*, 2024). Meskipun demikian, penelitian ini tidak melakukan pengujian *ablasi* khusus, misalnya membandingkan varian ResNet dengan dan tanpa *residual connection* pada dataset yang sama, untuk mengisolasi kontribusi *residual connection* secara spesifik. Oleh karena itu, *residual connection* sebaiknya dipahami sebagai salah satu faktor yang sejalan dengan hasil yang diperoleh, bukan satu-satunya penyebab yang telah dibuktikan secara eksperimental. Faktor lain, seperti perbedaan jumlah parameter, strategi regularisasi, maupun karakteristik data, tetap dapat turut berkontribusi.

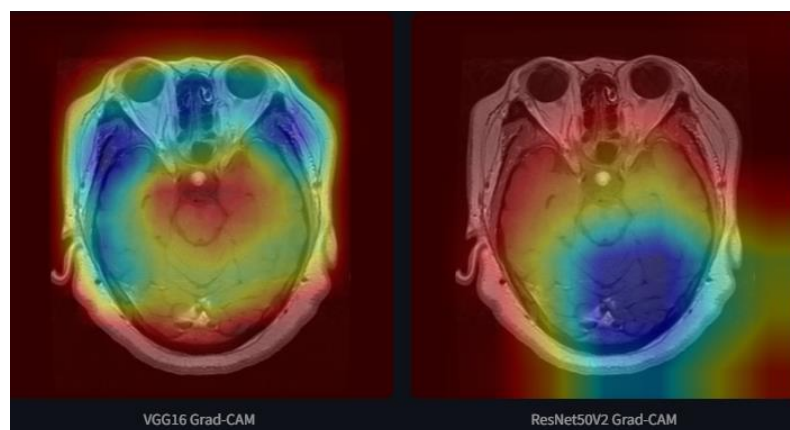
Analisis Grad-CAM pada keempat kelas turut menunjukkan bahwa kedua model memusatkan perhatian pada area yang secara medis relevan, misalnya di sekitar massa tumor pada kelas *glioma*, *meningioma*, dan *pituitary*, serta pada sebaran perhatian yang lebih merata pada kelas *no tumor*. Hal tersebut mengindikasikan tingkat interpretabilitas yang baik pada kedua arsitektur, konsisten dengan temuan bahwa ResNet50 dan VGG16 cenderung menunjukkan pola aktivasi Grad-CAM yang berbeda namun tetap terfokus pada area yang relevan secara klinis (Yoon & Lin, 2025). Visualisasi Grad-CAM pada kelas *glioma*, *meningioma*, *pituitary*, dan *no tumor* untuk kedua model ditampilkan pada Gambar 7, Gambar 8, Gambar 9, dan Gambar 10.



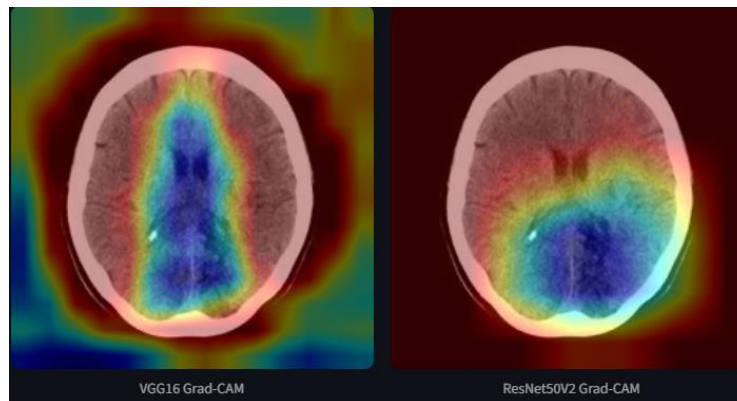
Gambar 7. Visualisasi Grad-CAM Model VGG16 dan ResNet50V2 pada Citra Kelas Glioma



Gambar 8. Visualisasi Grad-CAM Model VGG16 dan ResNet50V2 pada Citra Kelas Meningioma



Gambar 9. Visualisasi Grad-CAM pada Citra MRI Otak Kelas Pituitary



Gambar 10. Visualisasi Grad-CAM pada Citra MRI Otak Kelas No Tumor

Analisis lebih rinci per kelas menunjukkan pola yang tidak seluruhnya sejalan dengan keunggulan ResNet50V2 pada metrik agregat. Pada kelas *glioma*, ResNet50V2 mengklasifikasikan benar 246 dari 300 citra uji, sedangkan VGG16 hanya 222 citra, dengan kesalahan terbanyak pada kedua model terjadi ke arah kelas *meningioma* (43 citra pada ResNet50V2 dan 62 citra pada VGG16). Pada kelas *meningioma*, ResNet50V2 unggul tipis dengan 229 citra benar dibandingkan 217 citra pada VGG16. Namun, pada kelas *pituitary* dan *no tumor*, VGG16 justru sedikit lebih unggul: VGG16 mengklasifikasikan benar 296 citra *pituitary* dan 395 citra *no tumor*, dibandingkan ResNet50V2 yang memperoleh 288 citra *pituitary* dan 393 citra *no tumor*.

3.2 Pembahasan

Hasil penelitian menunjukkan bahwa ResNet50V2 memiliki performa yang lebih baik dibandingkan VGG16 pada klasifikasi tumor otak berbasis citra MRI. Keunggulan ini terlihat pada metrik evaluasi agregat, terutama *accuracy*, serta pada kemampuan model dalam membedakan kelas yang secara visual lebih sulit, yaitu *glioma* dan *meningioma*. Temuan ini sejalan dengan penelitian sebelumnya yang menyatakan bahwa klasifikasi tumor otak merupakan tugas yang menantang karena beberapa kelas memiliki karakteristik morfologi yang saling tumpang tindih, sehingga model memerlukan kemampuan ekstraksi fitur yang lebih kuat untuk menghasilkan prediksi yang akurat (Ahmmed *et al.*, 2023; Güler & Namlı, 2024). Pada kelas *pituitary* dan *no tumor*, kedua model menunjukkan hasil yang relatif sebanding, bahkan VGG16 sedikit lebih unggul pada beberapa citra. Hal ini menunjukkan bahwa keunggulan ResNet50V2 lebih relevan pada kelas yang kompleks, bukan pada seluruh kelas secara mutlak.

Secara arsitektural, hasil yang lebih baik dari ResNet50V2 dapat dijelaskan melalui mekanisme *residual connection*. Mekanisme ini memungkinkan informasi dan gradien mengalir lebih lancar selama pelatihan, sehingga mengurangi risiko *vanishing gradient* yang umum terjadi pada jaringan yang lebih dalam. Penelitian-penelitian sebelumnya juga menunjukkan bahwa arsitektur berbasis residual cenderung memberikan hasil yang kuat pada klasifikasi citra medis, termasuk MRI tumor otak, karena mampu menangkap fitur yang lebih dalam dan lebih kompleks dibandingkan arsitektur CNN klasik (Younis *et al.*, 2024; Musthafa *et al.*, 2024). Dalam konteks ini, ResNet50V2 lebih efektif dalam mempelajari pola visual yang sulit dibedakan antarkelas, khususnya antara *glioma* dan *meningioma*.

Meskipun ResNet50V2 unggul, selisih performa antara kedua model tidak terlalu besar. Nilai *accuracy* ResNet50V2 sebesar 88,17% dan VGG16 sebesar 86,19%, dengan selisih 1,98 poin persentase. Perbedaan kecil ini menunjukkan bahwa VGG16 masih merupakan model yang kompetitif. Hal tersebut konsisten dengan penelitian sebelumnya yang menyatakan bahwa VGG16 tetap mampu menghasilkan performa yang baik pada klasifikasi citra medis, terutama ketika digunakan dengan pendekatan *transfer learning* dan praproses data yang tepat (Khan & Khan, 2025; Sarshar *et al.*, 2024). Dengan kata lain, meskipun arsitektur VGG16 lebih sederhana daripada

ResNet50V2, model ini masih efektif dalam menangkap pola penting pada citra MRI, terutama untuk kelas yang memiliki ciri visual yang lebih jelas.

Dari sisi evaluasi per kelas, keunggulan ResNet50V2 terutama tampak pada kelas *glioma* dan *meningioma*. Pada dua kelas ini, model mampu mengurangi kesalahan klasifikasi dibandingkan VGG16. Hasil tersebut penting karena kedua kelas tersebut dikenal sulit dibedakan, baik secara visual maupun secara klinis. Studi lain juga menunjukkan bahwa model CNN yang lebih adaptif cenderung lebih unggul dalam membedakan kelas-kelas dengan kemiripan tinggi pada data MRI, terutama ketika dataset memiliki variasi intensitas dan bentuk lesi yang besar (Islam *et al.*, 2023; Huang & Prakash, 2025). Sementara itu, VGG16 sedikit lebih baik pada kelas *pituitary* dan *no tumor*, yang kemungkinan disebabkan oleh karakteristik citra yang lebih khas sehingga model yang lebih sederhana pun sudah cukup efektif mengenalinya.

Penelitian ini juga menyoroti adanya *trade-off* antara akurasi dan efisiensi komputasi. VGG16 menghasilkan waktu inferensi yang jauh lebih cepat dibandingkan ResNet50V2, yaitu rata-rata 5,5575 detik per citra dibandingkan 21,0124 detik per citra. Perbedaan ini menunjukkan bahwa VGG16 lebih efisien secara komputasi, sedangkan ResNet50V2 lebih unggul dalam akurasi. Temuan ini sejalan dengan literatur yang menyebutkan bahwa arsitektur yang lebih dalam dan kompleks biasanya memerlukan waktu proses yang lebih besar, meskipun menawarkan kemampuan representasi fitur yang lebih baik (Tan & Le, 2019; Huang *et al.*, 2017). Oleh karena itu, pemilihan model sangat bergantung pada kebutuhan implementasi. Jika tujuan utamanya adalah kecepatan untuk mendukung *triage* awal, VGG16 bisa menjadi pilihan. Namun, jika tujuan utamanya adalah memperoleh akurasi yang lebih tinggi pada kelas yang sulit, ResNet50V2 lebih tepat digunakan.

Selain evaluasi numerik, penelitian ini juga menggunakan Grad-CAM untuk melihat area citra yang menjadi fokus model. Hasil visualisasi menunjukkan bahwa kedua model memusatkan perhatian pada area yang secara medis relevan, seperti lokasi massa tumor pada kelas *glioma*, *meningioma*, dan *pituitary*. Pada kelas *no tumor*, perhatian model tersebar lebih merata di area otak yang tidak menunjukkan lesi. Hasil ini mengindikasikan bahwa kedua model tidak hanya mampu melakukan klasifikasi, tetapi juga memiliki kemampuan interpretatif yang cukup baik. Penggunaan Grad-CAM dalam penelitian ini sejalan dengan studi sebelumnya yang menunjukkan bahwa pendekatan *explainable AI* sangat berguna untuk memverifikasi apakah model belajar dari area patologi yang benar (Musthafa *et al.*, 2024; Yoon & Lin, 2025). Dengan demikian, visualisasi ini memperkuat kepercayaan bahwa prediksi model didasarkan pada fitur yang relevan secara klinis.

4. Kesimpulan dan Saran

Berdasarkan hasil penelitian, ResNet50V2 dan VGG16 sama-sama mampu mengklasifikasikan citra MRI tumor otak ke dalam empat kelas (*glioma*, *meningioma*, *pituitary*, dan *no tumor*) dengan baik, namun ResNet50V2 menunjukkan kinerja yang lebih unggul pada seluruh metrik evaluasi, dengan *accuracy* 88,17%, *precision* 87,63%, *sensitivity* 87,46%, *specificity* 96,06%, dan *F1-score* 87,46%, dibandingkan VGG16 yang memperoleh *accuracy* 86,19%, *precision* 85,83%, *sensitivity* 85,27%, *specificity* 95,39%, dan *F1-score* 85,18%. Keunggulan ini konsisten dengan karakteristik mekanisme *residual connection* pada ResNet50V2 yang secara umum membantu mengatasi permasalahan *vanishing gradient*, meskipun penelitian ini tidak melakukan pengujian ablasi khusus untuk membuktikan hubungan sebab-akibat tersebut secara langsung, sedangkan VGG16 tetap kompetitif berkat strukturnya yang sederhana dan mudah di-fine-tuning, bahkan menunjukkan hasil sedikit lebih baik pada kelas *pituitary* dan *no tumor*. Model ResNet50V2 kemudian berhasil diimplementasikan ke dalam prototipe aplikasi berbasis Streamlit yang memungkinkan klasifikasi citra MRI tumor otak secara interaktif. Prototipe ini masih bersifat riset awal dan belum diuji pada data klinis maupun divalidasi oleh tenaga medis, sehingga potensinya sebagai alat bantu diagnosis baru dapat dipertimbangkan setelah melalui tahap pengujian dan validasi lebih lanjut bersama radiolog atau dokter spesialis. Penelitian selanjutnya disarankan untuk menggunakan dataset yang lebih besar dan bervariasi, termasuk data MRI dari institusi kesehatan, mengeksplorasi arsitektur

deep learning lain seperti DenseNet (Huang *et al.*, 2017), EfficientNet (Tan & Le, 2019), atau Vision Transformer (Dosovitskiy *et al.*, 2021), serta mempertimbangkan pengembangan sistem ke dalam bentuk aplikasi web atau mobile yang terintegrasi dengan sistem rumah sakit setelah melalui proses validasi klinis yang memadai.

5. Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada Program Studi Sistem Informasi, Fakultas Teknik, Universitas Abulyatama, atas dukungan fasilitas dan bimbingan akademik selama penelitian ini berlangsung, serta kepada seluruh pihak yang telah membantu pelaksanaan penelitian ini.

6. Daftar Pustaka

- Ahmed, M. M., Hossain, M. M., Islam, M. R., Ali, M. S., Nafi, A. A. N., & Ahmed, F. (2024). Brain tumor detection and classification in MRI using hybrid ViT and GRU model with explainable AI in southern Bangladesh. *Scientific Reports*, 14, 22797. <https://doi.org/10.1038/s41598-024-71893-3>
- Ahmmed, S., Podder, P., Mondal, M. R. H., Rahman, S. M. A., Kannan, S., Hasan, M. J., Rohan, A., & Prosvirin, A. E. (2023). Enhancing brain tumor classification with transfer learning across multiple classes: An in-depth analysis. *BioMedInformatics*, 3(4), 1124–1144. <https://doi.org/10.3390/biomedinformatics3040068>
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations (ICLR)*. <https://arxiv.org/abs/2010.11929>
- Güler, M., & Namlı, E. (2024). Brain tumor detection with deep learning methods' classifier optimization using medical images. *Applied Sciences*, 14(2), 642. <https://doi.org/10.3390/app14020642>
- Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 4700–4708). <https://doi.org/10.1109/CVPR.2017.243>
- Huang, K. A., & Prakash, N. (2025). Evaluating the impact of attention mechanisms on a fine-tuned neural network for magnetic resonance imaging tumor classification: A comparative analysis. *Cureus*, 17(3). <https://doi.org/10.7759/cureus.80872>
- Islam, M. R., Reza, M. T., Kaosar, M., & Parvez, M. Z. (2023). Effectiveness of federated learning and CNN ensemble architectures for identifying brain tumors using MRI images. *Neural Processing Letters*, 55, 3779–3809. <https://doi.org/10.1007/s11063-022-11014-1>
- Khan, M., & Khan, M. T. (2025). Transfer learning for brain tumor MRI classification using VGG16: A comparative and explainable AI approach. *Preprints*. <https://doi.org/10.20944/preprints202504.2505.v2>

- Musthafa, M. M., Mahesh, T. R., Vinoth Kumar, V., & Guluwadi, S. (2024). Enhancing brain tumor detection in MRI images through explainable AI using Grad-CAM with ResNet50. *BMC Medical Imaging*, 24(1), 107. <https://doi.org/10.1186/s12880-024-01292-7>
- Özkaraca, O., Bağrıaçık, O. İ., Gürüler, H., Khan, F., Hussain, J., Khan, J., & Laila, U. E. (2023). Multiple brain tumor classification with dense CNN architecture using brain MRI images. *Life*, 13(2), 349. <https://doi.org/10.3390/life13020349>
- Rahman, A., Masum, M. I., Hasib, K. M., Mridha, M. F., Alfarhood, S., Safran, M., & Che, D. (2024). GliomaCNN: An effective lightweight CNN model in assessment of classifying brain tumor from magnetic resonance images using explainable AI. *Computer Modeling in Engineering & Sciences*, 140(3), 2425–2448. <https://doi.org/10.32604/cmescs.2024.050760>
- Rasa, S. M., Islam, M. M., Talukder, M. A., Uddin, M. A., Khalid, M., Kazi, M., & Kazi, M. Z. (2024). Brain tumor classification using fine-tuned transfer learning models on magnetic resonance imaging (MRI) images. *Digital Health*, 10. <https://doi.org/10.1177/20552076241286140>
- Rasheed, Z., Ma, Y.-K., Ullah, I., Al Shloul, T., Tufail, A. B., Ghadi, Y. Y., Khan, M. Z., & Mohamed, H. G. (2023). Automated classification of brain tumors from magnetic resonance imaging using deep learning. *Brain Sciences*, 13(4), 602. <https://doi.org/10.3390/brainsci13040602>
- Sarshar, N. T., Sadeghi, S., Kamsari, M., Avazpour, M., Ghouschi, S. J., & Ranjbarzadeh, R. (2024). Advancing brain MRI image classification: Integrating VGG16 and ResNet50 with a multi-verse optimization method. *BioMed*, 4(4), 499–523. <https://doi.org/10.3390/biomed4040038>
- Tan, M., & Le, Q. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. In *Proceedings of the 36th International Conference on Machine Learning (ICML)* (pp. 6105–6114).
- Wong, Y., Su, E. L. M., Yeong, C. F., Holderbaum, W., & Yang, C. (2025). Brain tumor classification using MRI images and deep learning techniques. *PLOS ONE*, 20(5). <https://doi.org/10.1371/journal.pone.0322624>
- Yoon, H. C., & Lin, L. P. (2025). Brain tumor classification in MRI: Insights from LIME and Grad-CAM explainable AI techniques. *IEEE Access*, 13, 154172–154202. <https://doi.org/10.1109/ACCESS.2025.3603272>
- Younis, A., Li, Q., Afzal, Z., Adamu, M. J., Kawuwa, H. B., Hussain, F., & Hussain, H. (2024). Abnormal brain tumors classification using ResNet50 and its comprehensive evaluation. *IEEE Access*, 12, 78843–78853. <https://doi.org/10.1109/ACCESS.2024.3403902>