Buletin Sistem Informasi dan Teknologi Islam
Vol 7, No 2 (2026), pp 150-159
ISSN: 2721-0901
DOI: http://dx.doi.org/10.33096/busiti.v7i2.3463

Hybrid CNN-Transformer yang Mempertimbangkan ROI untuk Klasifikasi Keberadaan Batu Ginjal pada Citra CT Aksial Heterogen

A ROI-Aware Hybrid CNN-Transformer for Kidney Stone Presence Classification on Heterogeneous Axial CT Images

Muh Ilham Akbar[a,1], Muhammad Faisal[a,2*], Desi Anggraeni[a,3], Abd Rakhim Nanda[b,4], Try Gustaf Said[c,5], Muhammad Syafaat S. Kuba[b,6]

[a] *Informatika, Universitas Muhammadiyah Makassar, Makassar, Indonesia*
[b] *Teknik Pengairan, Universitas Muhammadiyah Makassar, Makassar, Indonesia*
[c] *Pendidikan Guru Sekolah Dasar, Universitas Muhammadiyah Makassar, Makassar, Indonesia*
[1]*105841105822@student.unismuh.ac.id;* [2]*muhfaisal@unismuh.ac.id;* [3]*desianggreani@unismuh.ac.id;*
[4]*abd.rakhimnanda@unismuh.ac.id;* [5]*trygustafsaid@unismuh.ac.id;* [6]*syafaat_skuba@unismuh.ac.id*
**corresponding author*

Informasi Artikel

Diserahkan : 29 April 2026
Diterima : 14 Mei 2026
Direvisi : 28 Mei 2026
Diterbitkan : 29 Mei 2026

Kata Kunci:
batu ginjal
CT aksial
CNN-Transformer hibrida
CBAM
kalibrasi model

ABSTRAK

Batu ginjal merupakan penyebab umum nyeri pinggang akut, dan CT non-kontras menjadi standar referensi untuk mendeteksi kalkulus. Pada penelitian ini, istilah heterogen merujuk pada variasi protokol akuisisi antarrumah sakit, seperti perbedaan dosis radiasi, ketebalan irisan, rekonstruksi, dan bidang pandang, yang dapat mengubah tampilan citra serta menurunkan konsistensi pembacaan. Penelitian ini mengusulkan model hibrida CNN-Transformer yang sadar ROI (implisit) untuk klasifikasi keberadaan batu ginjal pada citra CT aksial heterogen. Arsitektur menggabungkan EfficientNet-B3, encoder Transformer ringan, dan Convolutional Block Attention Module (CBAM) tanpa anotasi ROI manual. Dataset terdiri dari 3.364 citra (1.577 batu, 1.787 non-batu) dengan pemisahan bertingkat 70/15/15. Evaluasi mencakup akurasi, presisi, sensitivitas, spesifisitas, F1, ROC-AUC, PR-AUC, inspeksi kalibrasi, dan audit Grad-CAM. Hasil menunjukkan bahwa penambahan Transformer meningkatkan kinerja dibanding baseline CNN, sedangkan CBAM menggeser profil kesalahan ke sensitivitas yang lebih tinggi. Varian Hybrid+Attention mencapai akurasi 0,9861, F1 0,9851, dan ROC-AUC 0,9967 pada set uji, dengan jumlah negatif palsu lebih rendah dibanding varian hibrida tanpa perhatian. Temuan ini menunjukkan potensi model sebagai alat bantu dokter untuk triase dan pembacaan awal yang lebih konsisten pada data lintas protokol, meskipun validasi eksternal, pemisahan berbasis pasien, dan metrik kalibrasi kuantitatif masih diperlukan sebelum klaim kesiapan klinis.

Keywords:
keyword1
kidney stone
axial CT
hybrid CNN-Transformer
CBAM
model calibration

ABSTRACT

Kidney stones are a frequent cause of acute flank pain, and non-contrast CT remains the reference standard for calculus detection. In this study, heterogeneous refers to cross-hospital variation in acquisition protocols, including differences in radiation dose, slice thickness, reconstruction, and field of view, which may alter image appearance and reduce reading consistency. We propose an implicitly ROI-aware hybrid CNN-Transformer for kidney stone presence classification on heterogeneous axial CT images. The architecture combines EfficientNet-B3, a lightweight Transformer encoder, and a Convolutional Block Attention Module (CBAM) without manual ROI annotation. The dataset comprises 3,364 images (1,577 stone; 1,787 non-stone) with a stratified 70/15/15 split. Evaluation includes accuracy, precision, recall/sensitivity, specificity, F1, ROC-AUC, PR-AUC, calibration inspection, and Grad-CAM auditing. Results show that adding the Transformer improves performance over the CNN baseline, while CBAM shifts the error profile toward higher sensitivity. The Hybrid+Attention model achieves 0.9861 accuracy, 0.9851 F1, and 0.9967 ROC-AUC on the held-out test set, with fewer false negatives than the hybrid without attention. These findings suggest practical value as a clinician-support tool for triage and early review under protocol variability, although external validation, patient-level splitting, and quantitative calibration metrics are still needed before strong clinical deployment claims.

I. Pendahuluan

Penyakit batu ginjal sering menjadi pendorong nyeri panggul akut dan CT darurat, dan CT non-kontras tetap menjadi standar referensi karena kepekaannya terhadap calculus hiperpadat di seluruh saluran kemih [1], [2]. Dalam praktiknya, deteksi rawan kesalahan untuk batu kecil atau mencolok rendah dan untuk peniru batu, dan kinerja dapat bervariasi dengan pilihan akuisisi dan rekonstruksi, memotivasi dukungan keputusan yang dinilai ketahanan dalam kondisi pencitraan heterogen [3].

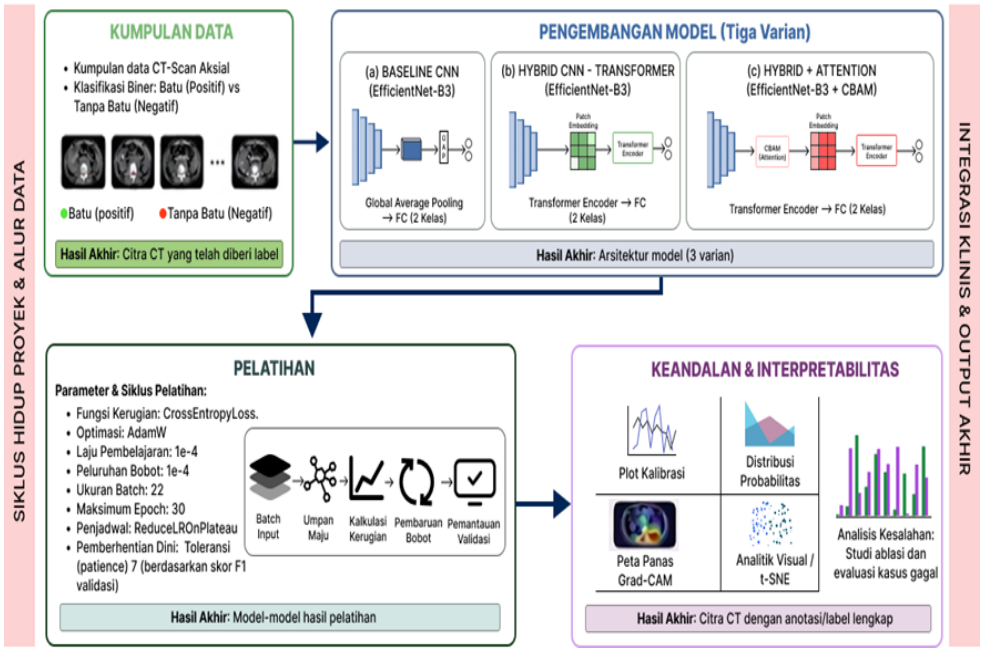
Sebagian besar model CT batu ginjal yang tersedia masih bertumpu pada CNN dan umumnya dirangkum melalui metrik diskriminasi agregat, padahal landasan tersebut belum memadai untuk penerapan klinis [4] Lokalitas CNN tidak selalu cukup ketika batu berukuran kecil, berkontras rendah, atau bercampur dengan mimickers hiperdens, karena keputusan klinis sering menuntut integrasi petunjuk lokal yang lemah dengan konteks anatomi yang lebih luas [5]. Atas dasar itu, desain CNN-Transformer hibrida dipilih untuk menggabungkan kekuatan representasi lokal CNN dengan pemodelan dependensi global Transformer [6]. Lebih penting lagi, pada aplikasi medis ril, model tidak cukup hanya akurat, tetapi juga harus menghasilkan confidence yang dapat dipercaya [7], [8]. Kalibrasi probabilitas menjadi krusial karena skor prediksi kerap digunakan untuk triase, prioritas telah, atau eskalasi pemeriksaan. Dengan demikian, novelty penelitian ini tidak hanya terletak pada peningkatan performa diskriminatif, tetapi pada upaya menilai model secara lebih dekat dengan kebutuhan klinis melalui kombinasi akurasi, reliabilitas probabilitas, dan audit interpretabilitas.

CNN-Transformer hibrida sadar ROI untuk klasifikasi keberadaan batu ginjal pada CT aksial, dengan menggabungkan tulang punggung EfficientNet-B3, encoder Transformer ringan, dan perhatian berbasis CBAM yang membias fitur ke wilayah yang relevan secara diagnostik tanpa anotasi ROI manual [9], [10]. Studi ini membandingkan garis dasar terkontrol (CNN vs hibrida vs hibrida + perhatian) dan melaporkan metrik diskriminasi (F1, ROC-AUC, PR-AUC) bersama inspeksi kalibrasi serta audit ROI berbasis Grad-CAM untuk mengkarakterisasi mode kegagalan dan reliabilitas confidence. Dalam konteks yang lebih luas, pengembangan alat bantu deteksi semacam ini juga selaras dengan upaya menjaga keselamatan jiwa (Hifdz an-Nafs) [11], selama model diposisikan sebagai pendukung keputusan dokter, bukan pengganti penilaian klinis.

II. Metode

A. Kerangka Penelitian

Gambar 1 merangkum alur kerja penelitian end-to-end, mulai dari persiapan data dan pra-pemrosesan standar hingga perbandingan terkontrol dari tiga varian model dan evaluasi berorientasi keandalan.



Gambar 1. Alur Penelitian

B. Persiapan Dataset

Dalam penelitian ini, gambar asli dibagi terlebih dahulu dan augmentasi diterapkan hanya selama pelatihan untuk mengurangi risiko kebocoran serta menjaga validasi/pengujian tetap deterministik [12].

Tabel 1. Ringkasan Dataset dan Protokol Pemisahan Bertingkat Sadar Kebocoran (seed=42) untuk Klasifikasi Batu/Non-Batu.

Aspek	Spesifikasi dalam penelitian ini	Pertimbangan metodologis
Tugas	Klasifikasi biner: Batu vs Non-Batu	Klasifikasi tingkat irisan biner digunakan sebagai target minimal yang relevan secara klinis (kehadiran vs tidak adanya sinyal batu) dan untuk mendukung perbandingan arsitektur terkontrol.
Sumber data	Dataset Pencitraan CT Aksial untuk Deteksi Batu Ginjal	Menyediakan gambar CT aksial yang diatur ke dalam kelas Batu/Non-Batu untuk tugas target.
Korpus yang digunakan	Hanya himpunan data asli	Folder tambahan offline dikecualikan dari pemisahan dan evaluasi untuk mengurangi risiko kebocoran.
Ukuran himpunan data	3.364 gambar CT aksial	1.577 Batu; 1,787 Non-Batu
Sumber label	Struktur folder kelas pada data mentah	Label mengikuti penetapan kelas berbasis folder dan diperbaiki dalam file terpisah.
Strategi pemisahan data	Pemisahan gambar bertingkat, benih = 42	Mempertahankan keseimbangan kelas di seluruh subset.
Pemisahan nominal	70% latihan, 15% validasi, 15% tes	Menyeimbangkan kecukupan data pelatihan dengan validasi/pengujian yang tidak bias.
Distribusi aktual	2.354 latih (train); 505 validasi; 505 uji	Hanya efek pembulatan kecil.
Meningkatkan	Augmentasi online ringan (hanya pelatihan)	Validasi/pengujian tetap deterministik.

C. Pra – Pemrosesan Gambar

Semua gambar diubah ukurannya menjadi 224×224 dan dinormalisasi menggunakan rata-rata/standar deviasi ImageNet agar sesuai dengan statistik yang diharapkan oleh tulang punggung yang telah dilatih sebelumnya ImageNet dan menstabilkan penyetelan halus [13].

Tabel 2. Ringkasan tahapan pra-pemrosesan gambar, termasuk ukuran masukan, normalisasi, augmentasi pelatihan, dan perlakuan validasi/pengujian.

Tahapan	Implementasi dalam penelitian ini	Tujuan metodologis
Ubah ukuran	Ukuran 224×224	Standarkan ukuran input di seluruh model.
Normalisasi	Rata-rata ImageNet [0,485, 0,456, 0,406] dan std [0,229, 0,224, 0,225]	Pertahankan kompatibilitas dengan statistik prapelatihan dan tingkatkan stabilitas pembelajaran
Peningkatan pelatihan	Flip (p=0,5), rotasi (±10°), kecerahan/kontras (p=0,3), noise Gaussian (p=0,2).	Meningkatkan variasi tanpa mendistorsi anatomi.
Pra-pemrosesan validasi/pengujian	Ubah ukuran + normalisasi saja (tanpa augmentasi)	Evaluasi deterministik.

D. Arsitektur Model Yang Diusulakan

Penelitian ini menerapkan tiga varian untuk mengisolasi kontribusi (i) lokalitas konvolusional, (ii) konteks global berbasis transformator, dan (iii) perhatian sadar ROI implisit untuk klasifikasi kehadiran Batu vs Non - Batu. Baseline menggunakan tulang punggung EfficientNet-B3 yang telah dilatih sebelumnya oleh ImageNet diikuti dengan kepala klasifikasi. Hibrida menggantikan klasifikasi langsung dengan encoder Transformer ringan yang beroperasi pada token yang berasal dari peta fitur CNN tahap akhir untuk menyuntikkan konteks jarak jauh [14], [15]. Hybrid+Attention menyisipkan CBAM antara tulang punggung CNN dan tokenisasi untuk menimbang ulang fitur (saluran dan spasial) ke wilayah informatif sebelum penalaran global [16].

$$\hat{y} = Softmax(f_{EB3}(x)) \quad (1)$$

Dalam varian Hibrida, peta fitur EfficientNet diproyeksikan ke dimensi penyematan 256 dan dibentuk ulang menjadi urutan token dengan penyematan posisi yang dapat dipelajari. Encoder Transformer 2 lapis digunakan dengan 8 kepala perhatian, dimensi umpan ke depan 1024, dan aktivasi GELU; representasi global diperoleh dengan pengumpulan rata-rata atas token (tanpa token CLS) dan diteruskan ke kepala MLP untuk klasifikasi biner.

CBAM menerapkan saluran berurutan dan perhatian spasial ke peta fitur CNN menggunakan rasio reduksi 16, menghasilkan representasi fitur berbobot ulang yang kemudian ditokenisasi dan diproses oleh encoder Transformer yang sama seperti pada varian Hybrid. Perlu ditekankan bahwa "ROI-aware" di sini bersifat implisit (dipandu perhatian) dan tidak bergantung pada ROI manual, kotak pembatas, atau masker segmentasi.

Secara keseluruhan, arsitektur mengoperasionalkan fusi lokal-global dengan bias induktif berbasis perhatian tambahan terhadap anatomi informatif, selaras dengan arah terbaru yang memasangkan diskriminasi dengan relevansi wilayah dan auditabilitas.

E. Konfigurasi Pelatihan

Pengaturan pelatihan identik di semua varian (Tabel 3) untuk mendukung perbandingan arsitektur terkontrol [17].

Tabel 3. Konfigurasi Pelatihan yang Dijaga Konstan Pada Seluruh Varian Model Untuk Mendukung Perbandingan Arsitektur Terkontrol

Komponen	Konfigurasi
Fungsi kerugian	CrossEntropyLoss
Pengoptimal	AdamW
Tingkat belajar	1×10^{-4}
Weight decay (peluruhan bobot)	1×10^{-4}
Ukuran batch	32
Epoch maksimum	30
Penjadwal	ReduceLROnPlateau (faktor 0,5, kesabaran 3)
Berhenti lebih awal	Kesabaran 7; pilih yang terbaik dengan validasi F1
Reprodusibilitas	Benih 42; backend deterministik; Setelan identik di seluruh varian

F. Metrik Evaluasi

Seluruh model dievaluasi pada set pengujian yang ditahan menggunakan metrik diskriminasi dan analisis berorientasi keandalan. Skor F1 untuk kelas Stone digunakan sebagai metrik utama; selain itu dilaporkan akurasi, presisi, ingatan/sensitivitas, spesifisitas, ROC-AUC, dan PR-AUC (presisi rata-rata) yang dihitung dari probabilitas Stone yang diprediksi [14].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$Recall = Sensitivity = \frac{TP}{TP + FN} \tag{4}$$

$$Specificity = \frac{TN}{TN + FP} \tag{5}$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{6}$$

TP, TN, FP, dan FN menunjukkan positif/negatif yang benar/palsu, dengan label diturunkan dari softmax argmax. Terkait komponen "calibrated" pada judul, penelitian ini memaknai kalibrasi sebagai evaluasi reliabilitas probabilitas keluaran model, bukan sebagai penerapan teknik kalibrasi pasca-hoc yang secara eksplisit mengoreksi probabilitas [18]. Dalam penelitian ini, tidak diterapkan teknik kalibrasi pasca-hoc seperti temperature scaling atau isotonic regression; dengan demikian, istilah "calibrated" merujuk pada evaluasi reliabilitas probabilitas, bukan pada koreksi probabilitas model. Oleh karena itu, diagram keandalan dan histogram probabilitas digunakan untuk mengaudit kecenderungan over-confidence atau under-confidence pada rentang probabilitas tertentu, sedangkan Grad-CAM dan tinjauan kesalahan yang ditargetkan digunakan sebagai audit kualitatif [19]. Dengan kerangka ini, naskah menilai apakah confidence model cukup layak dipercaya untuk konteks klinis, sembari menahan klaim bahwa probabilitas telah terkalibrasi secara optimal tanpa tahap koreksi tambahan seperti temperature scaling atau isotonic regression.

Selain inspeksi visual, reliabilitas probabilitas dapat diringkas secara numerik menggunakan Brier score, yaitu galat kuadrat rata-rata antara probabilitas prediksi kelas Stone dan label biner: , dengan dan [20]. Nilai yang lebih kecil menunjukkan kesalahan probabilistik yang lebih rendah (kalibrasi yang lebih baik), sehingga berguna untuk melengkapi diagram keandalan ketika jumlah sampel per-bin terbatas.

G. Rincian Implementasi dan Kerangka Evaluasi Komparatif

Eksperimen dirancang sebagai perbandingan arsitektur terkontrol di seluruh Baseline, Hibrida, dan Hibrida+Perhatian, dengan mempertahankan kriteria pemisahan, prapemrosesan, pengoptimal, penjadwal, dan penghentian. Implementasi menggunakan Python 3.10 dengan PyTorch/torchvision/timm, Albumentations, scikit-learn, dan OpenCV di bawah pengaturan deterministik (seed=42). Analisis interpretabilitas dan keandalan dilakukan melalui Grad-CAM dan plot kalibrasi untuk mendukung perbandingan transparan dan berorientasi audit di luar metrik judul [21].

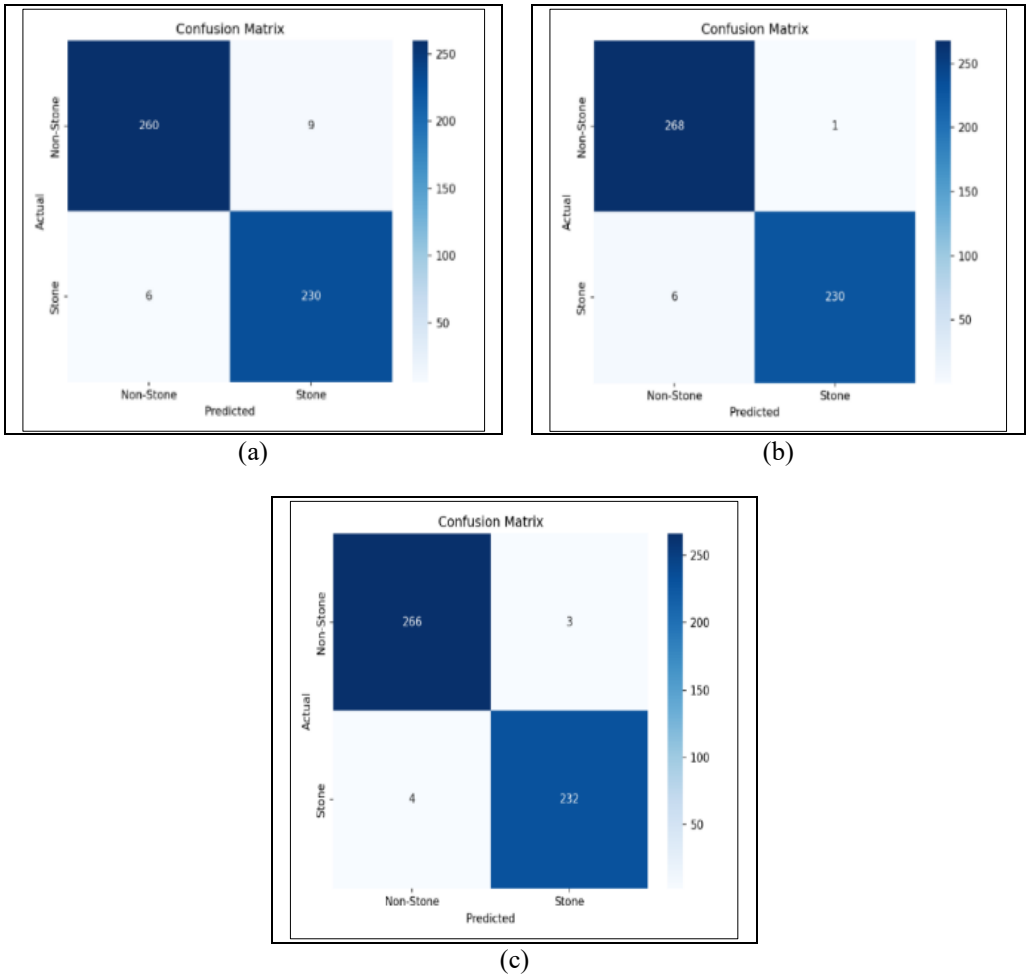
III.Hasil dan Pembahasan

Untuk menilai efek penambahan encoder Transformer dan perhatian CBAM pada klasifikasi keberadaan batu, tiga varian model dibandingkan pada set uji yang ditahan. Ringkasan hasil utama dilaporkan pada Tabel 4.

Tabel 4. Kinerja klasifikasi pada set uji untuk model Baseline, Hybrid, dan Hybrid+Attention.

Model	Akurasi	Presisi	Recall (Sensitivitas)	Skor F1	ROC-AUC	Kekhususan	PR-AUC
Baseline (EfficientNet-B3)	0.9703	0.9623	0.9746	0.9684	0.9966	0.9665	0.9965
Hibrida (EfficientNet-B3 + Encoder Transformator)	0.9861	0.9957	0.9746	0.9850	0.9964	0.9963	0.9973
Hibrida + Perhatian (EfficientNet-B3 + CBAM + Encoder Transformator)	0.9861	0.9872	0.9831	0.9851	0.9967	0.9888	0.9972

Tabel 4 menunjukkan bahwa keuntungan utama berasal dari penambahan Transformer (Baseline→Hybrid), meningkatkan F1 dan secara tajam mengurangi positif palsu (9→1). Menambahkan CBAM menggeser profil kesalahan ke arah lebih sedikit batu yang terlewatkan (FN: 6→4; ingatan: 0,9746→0,9831) dengan peningkatan positif palsu sederhana (1→3), yaitu, trade-off yang mendukung sensitivitas yang mungkin lebih disukai ketika meminimalkan batu yang terlewatkan diprioritaskan.

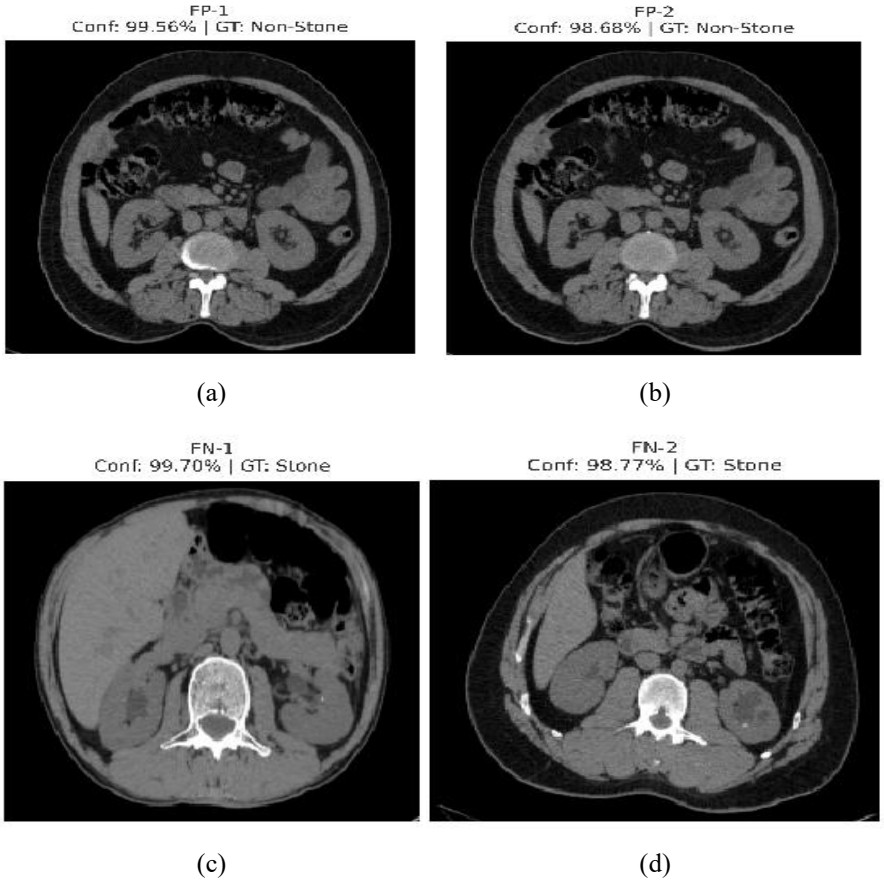


Gambar 2. Confusion matrix pada set uji yang ditahan untuk (a) Baseline, (b) Hybrid, dan (c) Hybrid+Attention.

Pada Gambar 2, panel (a) menunjukkan confusion matrix model Baseline, panel (b) menunjukkan model Hybrid, dan panel (c) menunjukkan model Hybrid+Attention. Secara umum, perubahan arsitektur tidak hanya menaikkan skor agregat, tetapi juga menggeser distribusi jenis kesalahan. Model Hibrida terutama menekan false positive (9→1) tanpa menurunkan false negative (tetap 6), yang mengindikasikan bahwa integrasi konteks global melalui Transformer membantu membedakan kalkulus sejati dari hiperattenuasi non-batu yang secara lokal tampak serupa. Sebaliknya, Hybrid+Attention menurunkan false negative (6→4) dengan kenaikan false

positive yang kecil (1→3), menunjukkan bahwa CBAM meningkatkan sensitivitas terhadap sinyal batu yang halus, tetapi pada saat yang sama masih dapat tertarik pada struktur hiperdens lain yang berada di sekitar lintasan traktus urinarius. Dari sudut pandang klinis, pergeseran ini bukan sekadar perubahan numerik: false negative berpotensi lebih bermakna karena dapat menunda perhatian terhadap pasien yang benar-benar memiliki batu, sedangkan sejumlah kecil false positive tambahan umumnya masih dapat dieliminasi melalui telaah radiologis lanjutan. Namun, trade-off tersebut tetap perlu dibaca secara hati-hati pada data heterogen, karena variasi protokol akuisisi-misalnya perbedaan dosis radiasi, ketebalan irisan, atau parameter rekonstruksi-dapat mengubah tekstur, noise, dan ketajaman tepi, sehingga memengaruhi bagaimana model memisahkan batu kecil dari peniru hiperattenuasi.

Sebagai pendalaman analisis salah deteksi, Gambar 3 menyajikan visualisasi Grad-CAM pada kasus false positive dan false negative yang representatif pada model Hybrid+Attention. Visualisasi ini digunakan untuk menilai apakah kesalahan model dipicu oleh perhatian yang bergeser ke struktur non-target, atau tetap berada di sekitar area yang secara anatomi masuk akal namun belum cukup diskriminatif untuk membedakan batu dari mimickers pada citra CT yang heterogen.

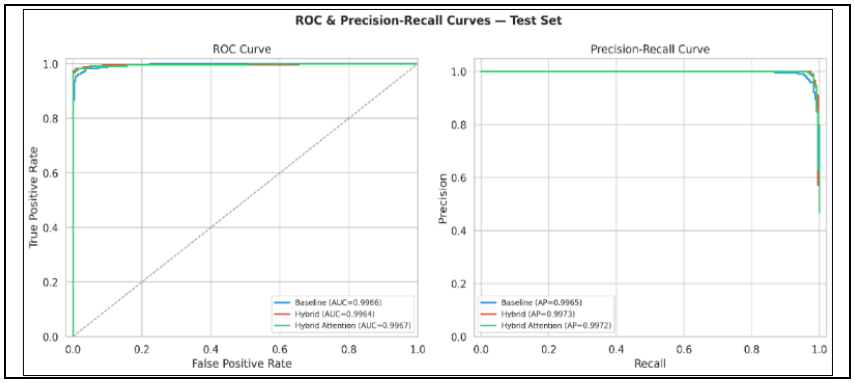


Gambar 3. Grad-CAM pada empat kasus salah deteksi representatif model Hybrid+Attention: (a) dan (b) false positive dan (c) dan (d) false negative

Gambar 3 menunjukkan bahwa pada sebagian false positive, peta aktivasi tetap terkonsentrasi di sekitar area hiperdens yang menyerupai kalkulus, sehingga kesalahan lebih mungkin dipicu oleh kemiripan morfologis dengan struktur non-batu daripada perhatian yang sepenuhnya acak. Pada false negative, aktivasi cenderung lebih lemah atau lebih menyebar ke jaringan sekitar, menunjukkan bahwa sinyal lesi belum cukup dominan untuk dipisahkan secara andal, terutama pada batu kecil, berkontras rendah, atau citra yang dipengaruhi heterogenitas akuisisi. Secara keseluruhan, visualisasi ini menegaskan bahwa perhatian implisit berbasis CBAM meningkatkan sensitivitas, tetapi belum sepenuhnya menghilangkan pengaruh mimickers dan variasi kualitas citra terhadap mode kegagalan model.

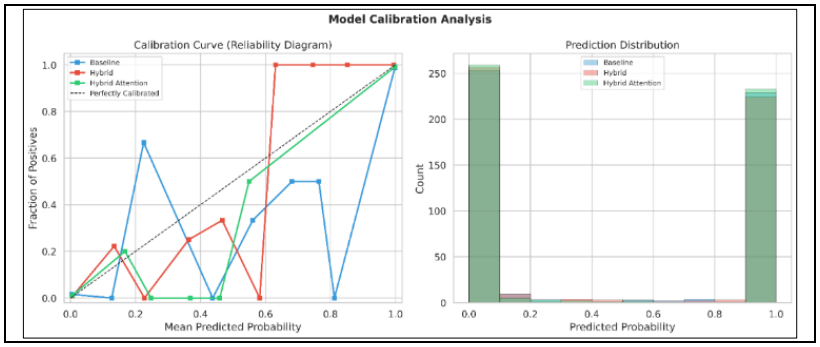
Untuk melengkapi metrik skalar pada Tabel 4, analisis berbasis kurva digunakan untuk menilai perilaku diskriminatif model di seluruh rentang ambang keputusan. Dalam konteks ini, kurva ROC dan precision-recall (PR) pada Gambar 4 memberikan gambaran apakah keunggulan masing-masing varian bersifat stabil secara global atau terutama muncul pada titik operasi tertentu. Karena data uji relatif seimbang tetapi konsekuensi

klinis false negative tetap penting, kombinasi ROC dan PR dipakai secara komplementer sebelum pembahasan dialihkan ke inspeksi kalibrasi pada Gambar 5.



Gambar 4. Kurva ROC dan PR untuk tiga varian model pada set uji yang ditahan.

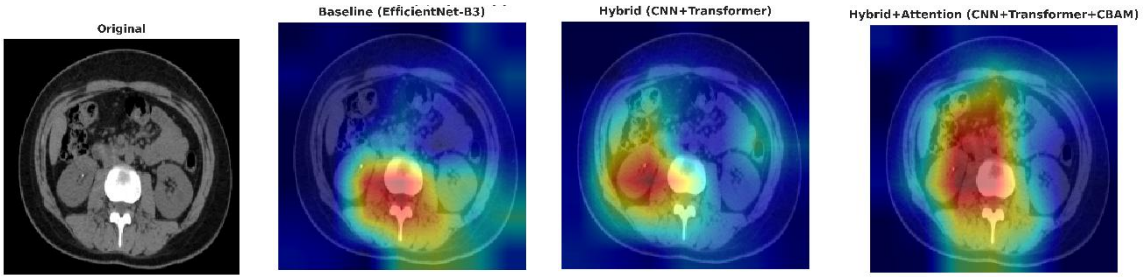
Gambar 4 menunjukkan bahwa seluruh varian memiliki kurva ROC dan PR yang nyaris jenuh, konsisten dengan nilai ROC-AUC (0,9964-0,9967) dan PR-AUC (0,9965-0,9973) yang sangat berdekatan. Temuan ini menandakan bahwa pada level peringkat global, ketiga model sama-sama sangat kuat sehingga kurva ROC/PR kurang sensitif untuk membedakan keunggulan praktis antarvarian. Oleh karena itu, pembacaan performa yang lebih bermakna perlu dipindahkan ke titik operasi tetap serta distribusi jenis kesalahan pada Tabel 4, Gambar 2, dan Gambar 3, tempat trade-off antara false positive dan false negative tampak lebih jelas. Dengan kata lain, Gambar 4 terutama menegaskan bahwa perbedaan antarmodel tidak terletak pada kegagalan diskriminasi global, melainkan pada pergeseran profil kesalahan saat dihadapkan pada data CT yang heterogen.

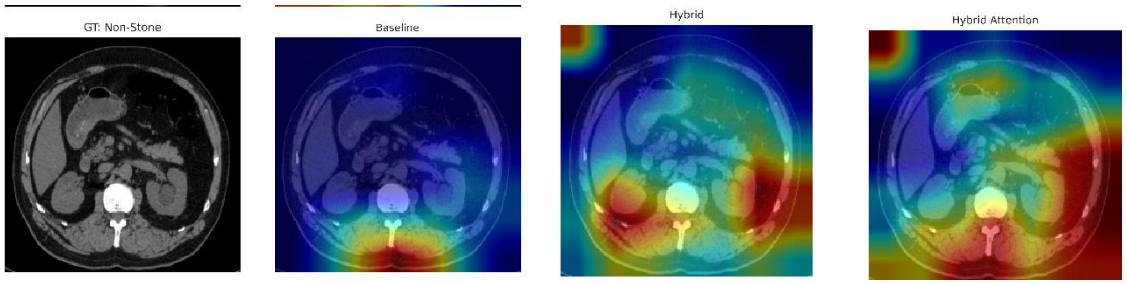


Gambar 5. Diagram keandalan dan histogram probabilitas untuk evaluasi kalibrasi tiga varian model.

Gambar 5 memberikan pemeriksaan kalibrasi kualitatif dan harus ditafsirkan dengan hati-hati karena diagram keandalan dapat tidak stabil dengan sampel terbatas per tempat sampah. Garis dasar menunjukkan penyimpangan yang lebih tidak teratur dari diagonal, sedangkan varian hibrida tampak sedikit lebih dekat dengan kalibrasi ideal di tempat sampah probabilitas lebih tinggi; namun, sisa over/under-confidence tetap ada, dan semua model memusatkan massa probabilitas mendekati 0 dan 1. Oleh karena itu, kalibrasi tidak boleh diklaim sebagai "terpecahkan" di sini; sebaliknya, plot ini paling baik dilihat sebagai sinyal audit yang memotivasi pekerjaan masa depan dengan kalibrasi pasca-hoc dan skor kuantitatif, idealnya di bawah validasi eksternal eksplisit dan pergeseran akuisisi.

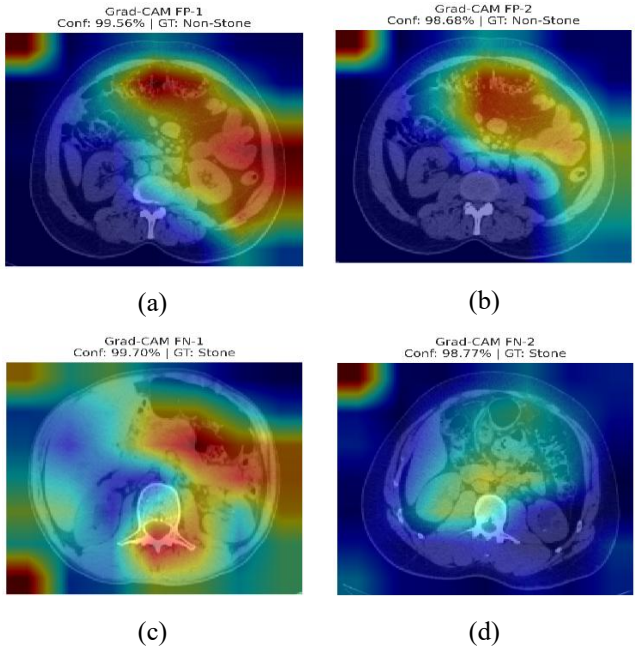
Selanjutnya, diperiksa apakah keuntungan arsitektur selaras dengan penggunaan bukti yang lebih masuk akal, menggunakan Grad-CAM sebagai audit kualitatif umum dan t-SNE sebagai pandangan eksploratif pemisahan fitur; keduanya tidak diperlakukan sebagai lokalisasi kebenaran dasar. Dalam susunan ini, Gambar 6 menyajikan pola atensi umum pada kasus Stone dan Non-Stone yang representative.





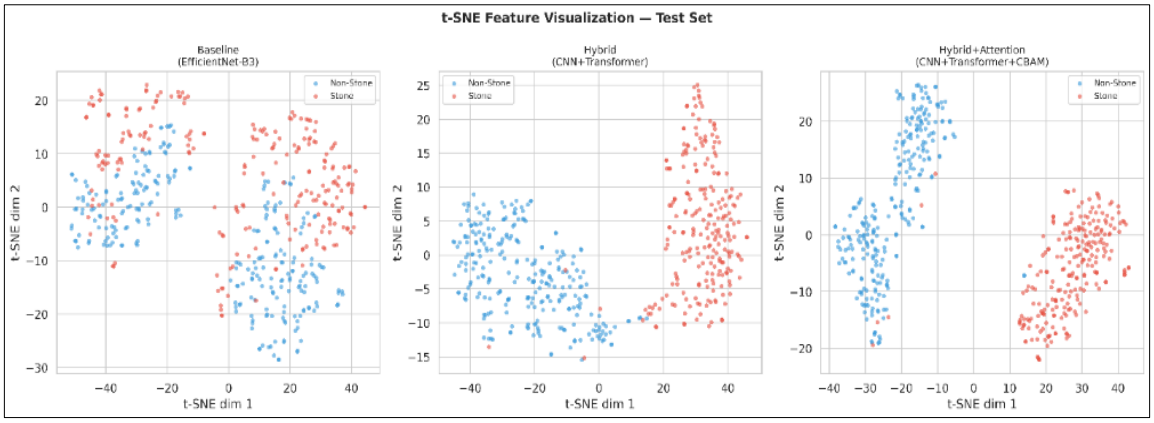
Gambar 6. Grad-CAM pada kasus Stone (baris atas) dan Non-Stone (baris bawah) untuk model Garis Dasar, Hibrida, dan Hibrida + Attention (kolom kiri ke kanan).

Gambar 6 menunjukkan pergeseran kualitatif yang konsisten dalam cara model menggunakan bukti visual. Baseline cenderung menghasilkan aktivasi yang lebih difus dan lebih mudah menyebar ke latar belakang, sedangkan model Hibrida mulai memperlihatkan fokus yang lebih terarah seiring hadirnya konteks global. Pada Hybrid+Attention, peta aktivasi tampak paling terlokalisasi dengan respons latar yang lebih rendah, yang secara konseptual sejalan dengan fungsi CBAM sebagai mekanisme pembobotan ulang fitur menuju wilayah yang lebih informatif secara diagnostik. Nilai utama temuan ini bukan sekadar bahwa peta terlihat lebih baik secara visual, melainkan bahwa visualisasi tersebut menyediakan jendela audit untuk menilai apakah peningkatan sensitivitas memang disertai pergeseran perhatian ke area yang lebih masuk akal secara anatomi. Dengan demikian, Grad-CAM di sini berfungsi sebagai alat untuk menilai kewajaran mekanisme keputusan model pada data CT yang heterogen, bukan sebagai bukti final bahwa model telah melokalisasi lesi secara presisi. Oleh sebab itu, interpretasi hasil visual tetap harus dibatasi: tanpa anotasi lesi, pengukuran overlap, atau verifikasi radiologis, Grad-CAM paling tepat digunakan untuk memahami mode kegagalan, menilai konsistensi fokus model, dan mengidentifikasi apakah kesalahan muncul karena perhatian yang sepenuhnya melenceng atau karena perhatian sudah berada di sekitar target namun belum cukup diskriminatif.



Gambar 7. Visualisasi Grad-CAM pada empat kasus salah deteksi representatif model Hybrid+Attention: (a) false positive 1, (b) false positive 2, (c) false negative 1, dan (d) false negative 2.

Gambar 7 memperluas interpretasi pada Gambar 6 dengan secara khusus menyoroti kasus salah deteksi pada model Hybrid+Attention. Pada false positive, peta aktivasi tetap terkonsentrasi di sekitar area hiperdens yang menyerupai kalkulus, menunjukkan bahwa kesalahan lebih mungkin dipicu oleh kemiripan morfologis dengan struktur non-batu daripada perhatian yang sepenuhnya acak. Pada false negative, aktivasi cenderung lebih lemah atau lebih menyebar ke jaringan sekitar, yang mengindikasikan bahwa sinyal lesi belum cukup dominan untuk dipisahkan secara andal, terutama pada batu kecil, berkontras rendah, atau citra yang dipengaruhi heterogenitas akuisisi. Temuan ini menegaskan bahwa perhatian implisit berbasis CBAM meningkatkan sensitivitas, tetapi belum sepenuhnya menghilangkan pengaruh mimickers dan variasi kualitas citra terhadap mode kegagalan model.



Gambar 8. Proyeksi t-SNE dari fitur lapisan kedua untuk ketiga varian sebagai visualisasi eksploratif pemisahan kelas.

Gambar 8 (t-SNE) secara kualitatif menunjukkan peningkatan pemisahan kelas dari Baseline ke Hybrid dan Hybrid+Attention. Sebagai proyeksi eksploratif yang sensitif terhadap hiperparameter, visualisasi ini tidak digunakan sebagai bukti representasi yang lebih baik secara mandiri; visualisasi tersebut dilaporkan semata-mata untuk melengkapi hasil kuantitatif.

Hybrid CNN-Transformer memberikan peningkatan utama dibandingkan baseline CNN, mendukung premis bahwa fusi lokal-global bermanfaat bagi klasifikasi kehadiran batu CT aksial. CBAM menghasilkan pergeseran yang lebih kecil tetapi relevan secara klinis menuju lebih sedikit batu yang terlewatkan (penarikan lebih tinggi; lebih sedikit negatif palsu) dengan peningkatan positif palsu yang sederhana, menunjukkan profil operasi yang mendukung sensitivitas. Audit kualitatif (Grad-CAM dan t-SNE) konsisten dengan aktivasi yang lebih terfokus dan peningkatan pemisahan fitur, tetapi bukan bukti lokalisasi. Secara praktis, model ini diposisikan lebih baik sebagai alat bantu penyaringan/triase daripada sistem yang berdiri sendiri, dan klaim penerapan yang lebih kuat memerlukan pemisahan berdasarkan pasien, validasi eksternal eksplisit di bawah pergeseran akuisisi, dan metrik kalibrasi kuantitatif.

Penelitian ini memiliki keterbatasan yang membatasi klaim eksternal. Pertama, pemisahan data bersifat gambar daripada pasien karena pengidentifikasi pasien tidak tersedia, sehingga kebocoran tingkat pasien sisa tidak dapat dikesampingkan. Kedua, ketahanan di bawah akuisisi heterogen hanya dinilai secara tidak langsung; bukti yang lebih kuat memerlukan validasi eksternal eksplisit. Ketiga, analisis Grad-CAM dan kalibrasi bersifat kualitatif dalam bentuk saat ini; bukti yang lebih kuat akan memerlukan anotasi tingkat lesi, tinjauan ahli, dan metrik kalibrasi numerik. Poin-poin ini harus ditangani sebelum menegaskan kesiapan klinis.

IV. Kesimpulan dan saran

Studi ini mengusulkan Hybrid CNN-Transformer yang sadar ROI untuk klasifikasi keberadaan batu ginjal pada CT aksial. Arsitekturnya menggabungkan tulang punggung EfficientNet-B3, encoder Transformer yang ringan untuk konteks global, dan CBAM untuk membias fitur ke wilayah yang relevan secara diagnostik. Studi ini juga melaporkan plot kalibrasi dan Grad-CAM sebagai alat audit, sehingga evaluasi melampaui diskriminasi menuju reliabilitas probabilitas dan kewajaran perhatian model. Secara eksperimental, kedua varian hibrida mengungguli garis dasar CNN, dengan Hybrid+Attention mencapai keseimbangan keseluruhan terbaik (Akurasi 0,9861; F1 0,9851; ROC-AUC 0,9967) dan recall kelas Stone tertinggi. Temuan ini mendukung bahwa fusi lokal-global dan pembobotan ulang fitur berbasis perhatian dapat meningkatkan klasifikasi batu pada CT aksial, sementara audit kualitatif menunjukkan pola aktivasi yang lebih terfokus. Secara praktis, model ini berpotensi membantu dokter sebagai alat bantu skrining, triase, dan pembacaan awal yang lebih konsisten pada data dengan protokol pencitraan yang beragam. Dengan tetap menempatkannya sebagai pendukung, bukan pengganti, keputusan klinis, pengembangan teknologi ini juga selaras dengan upaya menjaga keselamatan jiwa (Hifdz an-Nafs) melalui deteksi yang lebih cepat dan andal. Namun, klaim penerapan yang lebih kuat tetap memerlukan pemisahan berbasis pasien, validasi eksternal eksplisit di bawah pergeseran akuisisi, dan metrik kalibrasi kuantitatif yang lebih lengkap

Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada seluruh pihak yang telah memberikan kontribusi dalam pelaksanaan penelitian ini, baik melalui penyediaan data, diskusi teknis, maupun dukungan fasilitas penelitian.

Daftar Pustaka

[1] M. Montatore *et al.*, "Current Status on New Technique and Protocol in Urinary Stone Disease," *Curr. Radiol. Rep.*, vol. 11, pp. 161–176, 2023, doi: 10.1007/s40134-023-00420-5.

[2] N. J. Rao, H. Girish, M. C. Gowrishankar, S. Kumar, and N. Kumar, "A two-stage deep learning framework for kidney stone detection and clinical severity grading in CT imaging," *Inform. Med. Unlocked*, vol. 59, Jan. 2025, doi: 10.1016/j.imu.2025.101704.

[3] A. A. Hafiz, D. Vericho, V. J. Carter, D. C. Thio, M. Isnan, and B. Pardamean, "Vision Transformer and CNNs in Kidney Stone Classification: A Comparative Study," in *Procedia Computer Science*, Elsevier B.V., 2025, pp. 1466–1473. doi: 10.1016/j.procs.2025.09.088.

[4] M. Ali, Y. Saleem, S. Hina, and G. A. Shah, "DDoSViT: IoT DDoS attack detection for fortifying firmware Over-The-Air (OTA) updates using vision transformer," *Internet of Things (The Netherlands)*, vol. 30, Mar. 2025, doi: 10.1016/j.iot.2025.101527.

[5] H. Iwata, T. Shibayama, M. Watanabe, and H. Shimohiro, "Toward clinical reliability: Visualizing and interpreting AI-based classification in peripheral blood smear analysis," *Machine Learning with Applications*, vol. 22, p. 100780, Dec. 2025, doi: 10.1016/j.mlwa.2025.100780.

[6] T. Li, Z. Zhang, M. Zhu, Z. Cui, and D. Wei, "Combining transformer global and local feature extraction for object detection," *Complex & Intelligent Systems*, vol. 10, no. 4, pp. 4897–4920, Aug. 2024, doi: 10.1007/s40747-024-01409-z.

[7] C. Singh, A. Singh, and S. Dhelim, "Neuro-symbolic AI for rice disease diagnosis with calibrated attention and rule-aware explanations," *Information Processing in Agriculture*, 2026, doi: 10.1016/j.inpa.2026.02.006.

[8] N. Ullah, H. Sultan, J. S. Hong, S. G. Kim, R. Akram, and K. R. Park, "Convolutional self-attention with adaptive channel-attention network for obstructive sleep apnea detection using limited training data," *Eng. Appl. Artif. Intell.*, vol. 156, Sep. 2025, doi: 10.1016/j.engappai.2025.111154.

[9] R. Bhuvanya *et al.*, "Deep learning-based nail disease diagnosis leveraging the DERMANet architecture with ConvNeXt and CBAM," *Array*, vol. 30, Jul. 2026, doi: 10.1016/j.array.2026.100750.

[10] T. Mahmood, A. Wahid, J. S. Hong, S. G. Kim, and K. R. Park, "A novel convolution transformer-based network for histopathology-image classification using adaptive convolution and dynamic attention," *Eng. Appl. Artif. Intell.*, vol. 135, Sep. 2024, doi: 10.1016/j.engappai.2024.108824.

[11] N. Ullah, F. Guzmán-Aroca, F. Martínez-Álvarez, I. De Falco, and G. Sannino, "A novel explainable AI framework for medical image classification integrating statistical, visual, and rule-based methods," *Med. Image Anal.*, vol. 105, Oct. 2025, doi: 10.1016/j.media.2025.103665.

[12] P. A. Abdalla, B. S. Mahmood, and N. R. Hama, "MyKidney: A Web-based AI tool for automated kidney stone detection from CT imaging," *Invention Disclosure*, vol. 5, Dec. 2025, doi: 10.1016/j.inv.2025.100046.

[13] M. Vergin Raja Sarobin, S. Gupta, and A. A. Aziz, "Advancing brain tumor classification through pre-trained transformer and transfer learning models," *Franklin Open*, vol. 14, Mar. 2026, doi: 10.1016/j.fraope.2026.100493.

[14] H. Alhichri, A. Alswayed, Y. Bazi, N. Ammour, and N. Alajlan, "Classification of Remote Sensing Images Using EfficientNet-B3 CNN Model With Attention," *IEEE Access*, vol. 9, pp. 14078–14094, 2021, doi: 10.1109/access.2021.3051085.

[15] Y. Wang, Y. Qiu, P. Cheng, and J. Zhang, "Hybrid CNN-Transformer Features for Visual Place Recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 33, pp. 1109–1122, 2023, doi: 10.1109/tcsvt.2022.3212434.

[16] M. A. Rahman, "HyFormer-Net: A Synergistic CNN-Transformer with Interpretable Multi-Scale Fusion for Breast Lesion Segmentation and Classification in Ultrasound Images," *ArXiv*, vol. abs/2511.01013, p., 2025, doi: 10.48550/arxiv.2511.01013.

[17] C. Tabosa, M. Salgado, D. Leite, and A. Cunha, "ScienceDirect Comparative Analysis of CNNs and Vision Transformers for Lesion Classification in Capsule Endoscopy," *Procedia Comput. Sci.*, vol. 278, pp. 1186–1193, 2026, [Online]. Available: www.sciencedirect.com

[18] T. Dimitriadis, L. Duembgen, A. Henzi, M. Puke, and J. Ziegel, "Honest calibration assessment for binary outcome predictions," *Biometrika*, p., 2022, doi: 10.1093/biomet/asac068.

[19] S. Ahmed, D. Dera, S. U. Hassan, N. Bouaynaya, and G. Rasool, "Failure Detection in Deep Neural Networks for Medical Imaging," *Front. Med. Technol.*, vol. 4, p., 2022, doi: 10.3389/fmedt.2022.919046.

[20] Simran, V. Kukreja, V. Ahuja, S. Mehta, and A. Banal, "AI-driven model for knee cartilage degeneration using SAM, Swin, Grad-CAM, and CapsNet," *Franklin Open*, vol. 14, Mar. 2026, doi: 10.1016/j.fraope.2025.100472.

[21] S. Moreno-Montes, C. Delgado-Torres, E. Duzenli, N. Pérez-Zanón, R. Marcos-Matamoros, and A. Soret, "Comparative analysis of statistical downscaling methods for multi-model decadal climate predictions over Western Europe," *Clim. Serv.*, vol. 42, Apr. 2026, doi: 10.1016/j.cliser.2026.100639.