

Perkembangan Metode Image Quality Assessment Berbasis Deep Learning: A Systematic Literature Review terhadap Tren dan Evaluasi Performa

Amelia Nurul Islamiyah¹, Regita Nuralvianti Pratiwi²

^{1,2}Program Studi Informatika, Universitas Majalengka

Jl. KH. Abdul Halim No.103, Majalengka, Kabupaten Majalengka telp: (0233) 281496,

E-mail: [1amelianurul706@gmail.com](mailto:amelianurul706@gmail.com), [2nrlvianti18@gmail.com](mailto:nrlvianti18@gmail.com)

ARTICLE INFO

Article history:

Received March 2026

Revised April 2026

Accepted June 2026

Available online July 2026

ABSTRACT

This study presents a Systematic Literature Review (SLR) based exclusively on Scopus-indexed publications related to Deep Learning-based Image Quality Assessment (IQA). The review aims to identify dominant deep learning architectures, application domains, research trends, performance evaluation approaches, and research gaps in IQA studies. The literature selection process followed the PRISMA 2020 framework. An initial search retrieved 1,599 articles from the Scopus database, which were systematically screened using predefined inclusion and exclusion criteria, resulting in 36 eligible studies for detailed analysis. The findings indicate that Convolutional Neural Networks (CNNs) remain the most widely adopted architecture due to their efficiency and adaptability, while Transformer and Hybrid models demonstrate superior performance in handling complex image distortions. Furthermore, No-Reference Image Quality Assessment (NR-IQA) dominates recent research, whereas Reduced-Reference IQA (RR-IQA) remains relatively underexplored. The review highlights future opportunities in adaptive RR-IQA, multimodal artificial intelligence, and broader application domains.

Keywords: Deep Learning, Image Quality Assessment, Convolutional Neural Network, Transformer, No-Reference Image Quality Assessment, Systematic Literature Review

1. Pendahuluan

Perkembangan teknologi pengolahan citra digital mendorong meningkatnya kebutuhan terhadap metode Image Quality Assessment (IQA) yang mampu mengevaluasi kualitas visual citra secara akurat dan otomatis. IQA telah menjadi komponen kritis dalam berbagai bidang, mulai dari pencitraan medis, industri manufaktur, hingga penginderaan jauh. Kajian sistematis terbaru menegaskan bahwa kualitas citra secara langsung memengaruhi akurasi diagnosis dan hasil klinis [1], sementara evaluasi kualitas pada sistem tenaga listrik [2] dan citra ISAR berbasis visual attention model [3] menunjukkan luasnya cakupan aplikasi IQA. Perkembangan ini mendorong pergeseran dari metode evaluasi konvensional menuju pendekatan berbasis deep learning yang lebih adaptif dan otomatis.

Secara umum, metode IQA terbagi menjadi tiga kategori yaitu *Full-Reference (FR-IQA)*, *Reduced-Reference (RR-IQA)* dan *No-Reference (NR-IQA)*. Metrik tradisional seperti PSNR dan SSIM terbukti tidak selalu berkorelasi dengan persepsi kualitas manusia pada kondisi distorsi kompleks [4], [5]. Pendekatan yang lebih maju seperti GMSD [4] dan FSIM [6] memperbaiki korelasi tersebut, namun tetap terbatas pada fitur *hand-crafted*. *Deep neural network* kemudian terbukti mampu mempelajari representasi fitur perseptual secara otomatis untuk NR-IQA maupun FR-IQA sekaligus [7], dan deep features dari jaringan pra-latih terbukti menjadi metrik perseptual yang sangat efektif [8]. Arsitektur Transformer yang diterapkan pada blind IQA mampu menangkap dependensi jarak jauh antar patch citra secara lebih efektif dibandingkan CNN konvensional, sehingga menghasilkan korelasi yang lebih tinggi dengan penilaian kualitas subjektif manusia pada berbagai dataset benchmark [9]. Tinjauan terhadap metode RR-IQA mengonfirmasi bahwa kategori ini merupakan area yang paling kurang berkembang dibanding dua kategori lainnya [9].

Urgensi IQA semakin meningkat pada berbagai domain aplikasi. Pada domain medis, *explainable IQA* diperlukan agar klinisi dapat memverifikasi keluaran model secara langsung [10], *semantic segmentation* berbasis *deep learning* meningkatkan objektivitas penilaian kolonoskopi [11], dan *semi-supervised multi-task learning* dikembangkan untuk evaluasi kualitas citra fundus dengan kebutuhan anotasi yang lebih efisien [12]. Pada domain industri, arsitektur *dual-pipeline* memisahkan tahap *grading* dan deteksi cacat secara bersamaan [13], sedangkan algoritma berbasis *rule* multi-kriteria lebih sesuai untuk perangkat IIoT berbiaya rendah [14]. Pada penginderaan jauh, evaluasi resolusi optis kamera UAV telah dilakukan sebagai dasar penjaminan kualitas citra udara [15].

Berbagai pendekatan terbaru terus bermunculan untuk menjawab tantangan IQA yang semakin kompleks. *Multiscale attention mechanism* yang dikombinasikan dengan statistik *natural scene* (MAMIQA) menghasilkan korelasi lebih tinggi dengan penilaian manusia [16], dan integrasi fitur global-lokal terbukti meningkatkan akurasi NR-IQA secara signifikan [17]. Perkembangan terkini menunjukkan bahwa *Multimodal Large Language Models* (MLLM) mulai dimanfaatkan untuk NR-IQA pada distorsi autentik, dengan evaluasi terhadap model GPT-4V dan Claude.AI menunjukkan potensi sekaligus keterbatasan dalam konsistensi prediksi [18]. Selain itu, metrik IQA kini tidak hanya digunakan sebagai alat evaluasi, tetapi juga sebagai fungsi loss dalam pelatihan model restorasi citra, penggunaan metrik IQA perseptual sebagai loss function pada model dehazing menghasilkan output yang lebih selaras dengan persepsi visual manusia dibandingkan loss berbasis piksel konvensional seperti MSE [20]. Namun demikian, belum banyak penelitian yang secara sistematis menganalisis perkembangan metode IQA secara lintas domain, melakukan sintesis kritis komparatif antar pendekatan, dan mengkaji profil bibliometrik penelitian terkini. Kondisi ini menegaskan adanya *research gap* berupa kurangnya kajian literatur menyeluruh mengenai perkembangan metode IQA berbasis *deep learning* dari perspektif yang holistik [19].

Berdasarkan *Research Gap* tersebut, penelitian ini melakukan *Systematic Literature Review* (SLR) terhadap tiga puluh enam artikel terindeks Scopus untuk menganalisis perkembangan metode IQA secara komprehensif. Kebaruan penelitian ini terletak pada analisis terintegrasi yang tidak hanya mengkaji metode yang digunakan, tetapi juga tren perkembangan teknologi, sintesis kritis keunggulan dan keterbatasan lintas pendekatan, serta profil bibliometrik sebagai dasar identifikasi arah penelitian yang masih kurang terjamah.

Penelitian ini menjawab lima pertanyaan penelitian yang dirumuskan sebagai berikut:

- RQ1 : Metode Deep Learning apa yang paling dominan digunakan dalam penelitian Image Quality Assessment terkini, dan bagaimana perbandingan kinerja antara pendekatan CNN, Transformer dan Hybrid dalam berbagai kondisi distorsi?
- RQ2 : Bagaimana peran domain aplikasi dalam menentukan pilihan metode IQA dan apakah terdapat pola sistematis antara karakteristik domain dengan arsitektur deep learning yang digunakan?
- RQ3 : Bagaimana tren perkembangan metode IQA dari tahun ke tahun berdasarkan literatur terkini dan ke arah mana penelitian IQA berbasis deep learning berkembang?
- RQ4 : Apa keunggulan dan keterbatasan kritis metode IQA berbasis deep learning di bandingkan dengan metode tradisional, ditinjau dari aspek akurasi, skalabilitas, dan kemampuan generalisasi lintas domain?

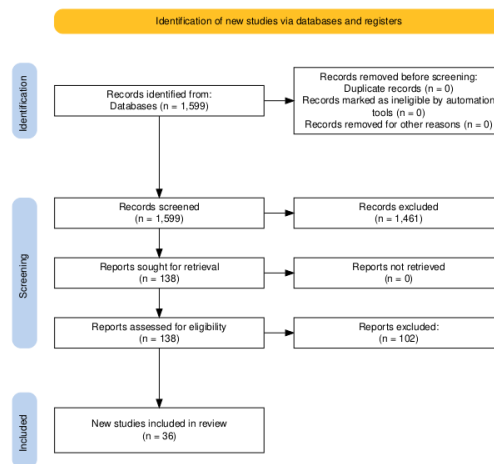
RQ5 : Bagaimana profil bibliometric penelitian IQA terkini ditinjau dari distribusi domain aplikasi, jenis pendekatan (FR/RR/NR), dan sumber publikasi, serta apa implikasinya terhadap arah penelitian yang masih kurang terjamah?

Tinjauan literatur sistematis ini diharapkan dapat memberikan gambaran menyeluruh mengenai perkembangan terkini metode IQA berbasis deep learning, mengidentifikasi celah penelitian yang belum terjawab, serta menjadi referensi bagi peneliti dalam memilih pendekatan IQA yang paling sesuai dengan kebutuhan domain aplikasinya.

2. Metode Penelitian

Penelitian ini menggunakan pendekatan *Systematic Literature Review* (SLR) untuk mengkaji perkembangan metode *Image Quality Assessment* (IQA) berbasis *deep learning* secara sistematis dan komprehensif. SLR dipilih karena memungkinkan identifikasi, evaluasi, dan sintesis literatur yang relevan secara terstruktur sehingga menghasilkan temuan yang dapat dipertanggungjawabkan secara ilmiah [1]. Proses penelitian mengikuti pedoman *Preferred Reporting Items for Systematic Reviews and Meta-Analyses* (PRISMA) 2020 yang merupakan standar internasional dalam pelaporan *systematic review* dan terbukti meningkatkan transparansi serta kualitas dokumentasi proses seleksi literatur [20].

Pencarian literatur dilakukan secara eksklusif pada database Scopus menggunakan kombinasi kata kunci: "*Image Quality Assessment*", "*Deep Learning*", "*CNN*", "*Transformer*", dan "*No-Reference IQA*". Scopus dipilih sebagai satu-satunya sumber pencarian karena merupakan database multidisiplin terbesar yang mencakup jurnal teknik dan ilmu komputer bereputasi tinggi, serta menjamin seluruh artikel yang terindeks telah melalui proses *peer review* yang ketat [1], [21]. Dengan demikian, seluruh artikel yang diidentifikasi, diseleksi, hingga dianalisis dalam penelitian ini bersumber eksklusif dari Scopus. Pencarian awal menghasilkan 1.599 artikel sebagai kandidat awal yang selanjutnya diproses melalui empat tahap seleksi sesuai alur PRISMA 2020 sebagaimana ditunjukkan pada Gambar 2.1.



Gambar 2.1 Diagram Prisma

Tabel 2.1 Hasil Screening

Tahapan	Jumlah
Records Identified	1599
Records After Duplicates Removed	138
Records Screened	138
Records Excluded	102
Records Assessed for Eligibility	138
Reports Excluded	102
Studies Included	36

Pada tahap identifikasi, seluruh 1.599 artikel diteruskan ke tahap berikutnya karena tidak ditemukan duplikasi maupun artikel yang ditandai tidak memenuhi syarat oleh *automation tools*. Pada tahap *screening* berdasarkan judul dan abstrak, sebanyak 1.461 artikel dieliminasi karena tidak relevan dengan topik IQA atau berada di luar bidang pengolahan citra digital, sehingga diperoleh 138 artikel yang seluruhnya berhasil diperoleh full-text-nya. Pada tahap *eligibility*, analisis full-text dilakukan terhadap 138 artikel menggunakan kriteria inklusi: (1) membahas metode IQA secara eksplisit sebagai fokus utama; (2) menggunakan pendekatan berbasis deep learning, CNN, transformer, atau metode evaluasi kualitas citra terukur lainnya; (3) memiliki hasil evaluasi performa yang jelas menggunakan metrik standar; serta (4) menggunakan dataset atau skenario pengujian yang dijelaskan secara eksplisit dan dapat direplikasi [12], [21]. Sebanyak 102 artikel dieksklusikan karena tidak memenuhi satu atau lebih kriteria tersebut, sehingga pada tahap inklusi diperoleh 36 artikel terindeks Scopus yang memenuhi seluruh kriteria dan digunakan sebagai korpus analisis dalam penelitian ini.

Data yang diekstraksi dari setiap artikel meliputi metode IQA beserta arsitektur modelnya, domain aplikasi dan karakteristik dataset, strategi evaluasi performa termasuk metrik yang digunakan, serta keunggulan dan keterbatasan metode yang diterapkan. Data kemudian dianalisis secara kualitatif menggunakan pendekatan sintesis naratif untuk mengidentifikasi pola perkembangan metode, tren teknologi, dan hubungan antara metode IQA dengan domain aplikasinya [23], [21]. Pendekatan ini dipilih mengingat heterogenitas yang tinggi antar studi dalam hal domain aplikasi, arsitektur model, dan metrik evaluasi yang digunakan. Hasil analisis digunakan untuk menjawab lima *Research Questions* (RQ1–RQ5) yang telah dirumuskan.

3. Hasil dan Pembahasan

Berdasarkan proses seleksi literatur menggunakan metode PRISMA, diperoleh sebanyak 36 artikel yang memenuhi kriteria inklusi dan digunakan dalam tahap analisis. Artikel-artikel tersebut dipublikasikan pada rentang tahun 2011 hingga 2026 dan membahas berbagai aspek Image Quality Assessment (IQA), mulai dari pengembangan metrik kualitas citra, evaluasi kualitas citra medis, hingga penerapan deep learning pada berbagai domain aplikasi [24], [25].

Secara umum, penelitian IQA mengalami perkembangan yang signifikan dalam beberapa tahun terakhir. Perkembangan tersebut ditandai dengan meningkatnya penggunaan pendekatan deep learning, khususnya Convolutional Neural Network (CNN), Transformer, dan model hybrid yang mampu meningkatkan kemampuan sistem dalam mengevaluasi kualitas citra secara otomatis [26], [27], [23], [28].

Berdasarkan domain aplikasinya, penelitian IQA yang dianalisis dapat dikelompokkan ke dalam beberapa kategori utama, yaitu domain medis, industri dan kontrol kualitas, penginderaan jauh, lingkungan dan kualitas udara, serta penelitian umum yang berfokus pada pengembangan metode IQA. Domain medis menjadi salah satu area penelitian yang paling aktif karena kebutuhan akan evaluasi kualitas citra yang akurat untuk mendukung diagnosis dan pengambilan keputusan klinis [29], [30], [19], [31]. Sementara itu,

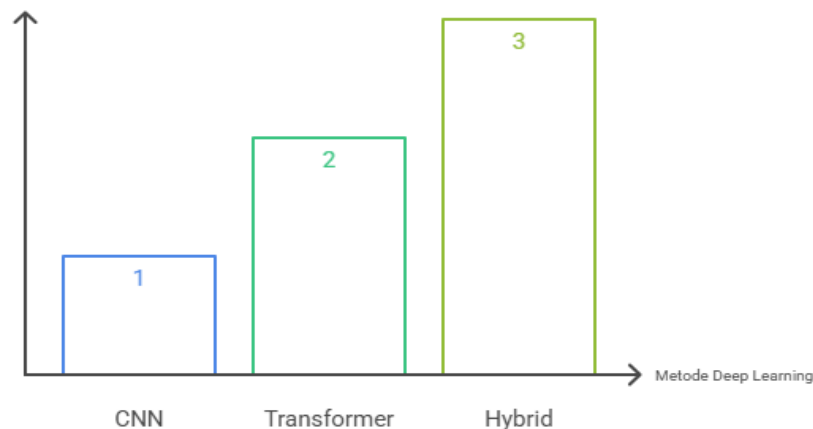
domain industri dan kontrol kualitas memanfaatkan IQA untuk mendukung otomatisasi inspeksi dan pemantauan kualitas produk secara real-time [2], [32], [3].

Selain itu, berdasarkan jenis pendekatannya, penelitian IQA dapat diklasifikasikan menjadi Full-Reference Image Quality Assessment (FR-IQA), Reduced-Reference Image Quality Assessment (RR-IQA), dan No-Reference Image Quality Assessment (NR-IQA). Penelitian terkini menunjukkan bahwa pendekatan NR-IQA semakin mendominasi karena lebih sesuai dengan kondisi dunia nyata yang umumnya tidak memiliki citra referensi sebagai pembanding [26], [27], [33].

Untuk memberikan pemahaman yang lebih komprehensif mengenai perkembangan penelitian IQA, analisis selanjutnya disusun berdasarkan lima research question (RQ) yang mencakup identifikasi metode deep learning yang dominan, hubungan antara domain aplikasi dan metode yang digunakan, tren perkembangan penelitian, keunggulan dan keterbatasan pendekatan yang ada, serta profil bibliometrik dan peluang penelitian di masa mendatang.

3.1. RQ1 : Metode Deep Learning apa yang paling dominan digunakan dalam penelitian Image Quality Assessment terkini, dan bagaimana perbandingan kinerja antara pendekatan CNN, Transformer dan Hybrid dalam berbagai kondisi distorsi?

RQ1 bertujuan untuk mengidentifikasi metode deep learning yang paling dominan digunakan dalam penelitian Image Quality Assessment (IQA) serta membandingkan karakteristik dan performa pendekatan CNN, Transformer, dan Hybrid berdasarkan hasil sintesis dari 36 artikel yang dianalisis. Analisis ini dilakukan untuk memahami perkembangan pendekatan deep learning yang digunakan dalam menilai kualitas citra serta mengevaluasi keunggulan dan keterbatasan masing-masing metode pada berbagai kondisi distorsi.



Gambar 3.1 Perbandingan Kinerja Metode Deep Learning dalam IQA

Tabel 3.1 Tabel Perbandingan Karakteristik CNN, Transformer dan Hybrid dalam IQA

Pendekatan	Dominasi Penggunaan	Keunggulan	Keterbatasan	Kondisi Unggul
CNN	Tinggi	Efisien, mudah dilatih, cocok untuk dataset terbatas	Sulit menangkap hubungan global	Real-time dan dataset kecil

Transformer	Sedang	Mampu menangkap hubungan global antar fitur	Membutuhkan komputasi tinggi dan dataset besar	Citra kompleks dengan distorsi beragam
Hybrid	Rendah-Sedang	Menggabungkan keunggulan CNN dan transformer, akurasi tinggi	Arsitektur lebih kompleks	Distorsi autentik dan benchmark multi-dataset

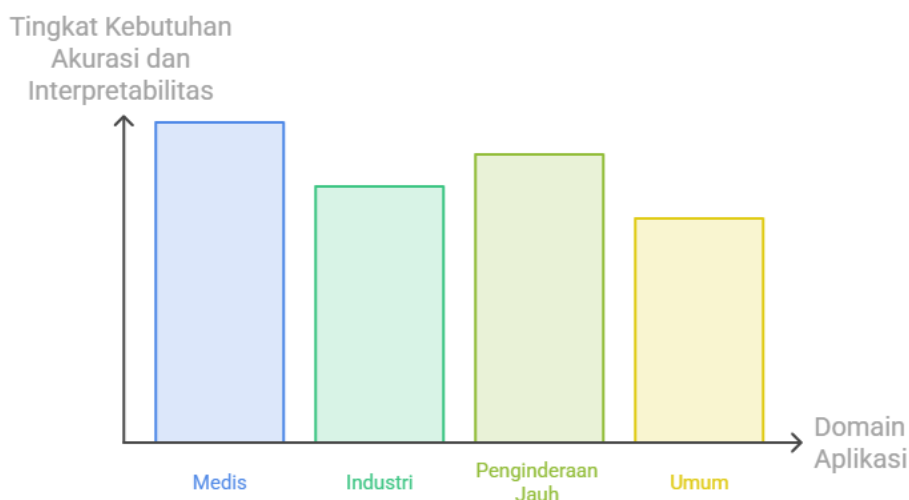
Berdasarkan analisis terhadap literatur terkini, CNN tetap menjadi arsitektur yang paling dominan dalam penelitian IQA, namun transformer dan arsitektur hybrid semakin menunjukkan keunggulan pada kondisi tertentu. Jaringan berbasis hyper network yang bersifat self-adaptive mampu melakukan blind IQA secara akurat pada citra liar (in-the-wild) tanpa memerlukan citra referensi, dengan performa yang melampaui metode NR-IQA berbasis CNN konvensional pada dataset KonIQ-10k dan SPAQ [32].

Perbandingan empiris antara CNN dan arsitektur Transformer pada klasifikasi fase pembungaan tanaman menunjukkan bahwa Transformer mencapai akurasi lebih tinggi, namun CNN tetap unggul dalam efisiensi komputasi pada implementasi dengan sumber daya terbatas [35]. Pendekatan NR-IQA berbasis fusi fitur diferensial mampu meningkatkan akurasi penilaian kualitas citra planar tanpa memerlukan citra referensi, memperkuat relevansi metode fusion dalam ekosistem NR-IQA terkini [36].

Arsitektur Transformer terbukti mengungguli CNN dalam menangkap hubungan fitur global pada klasifikasi kualitas citra, meskipun CNN tetap lebih efisien secara komputasi untuk implementasi real-time [30]. Model berbasis attention mechanism multiskala yang dikombinasikan dengan statistik natural scene (MAMIQA) menghasilkan korelasi lebih tinggi dengan penilaian manusia dibanding CNN murni pada empat benchmark standar [16]. Pendekatan fusion fitur global-lokal menunjukkan bahwa integrasi konteks lokal dan global secara bersamaan merupakan kunci peningkatan akurasi NR-IQA pada distorsi autentik [32]. Secara kritis, CNN unggul dalam efisiensi dan interpretabilitas pada dataset terbatas [18] transformer unggul pada pemahaman konteks global, sedangkan pendekatan hybrid menunjukkan performa terbaik secara keseluruhan namun dengan kompleksitas implementasi tertinggi.

RQ2 : Bagaimana peran domain aplikasi dalam menentukan pilihan metode IQA dan apakah terdapat pola sistematis antara karakteristik domain dengan arsitektur deep learning yang digunakan?

RQ2 bertujuan untuk menganalisis hubungan antara domain aplikasi dan metode IQA yang digunakan dalam penelitian terkini. Analisis ini dilakukan untuk mengidentifikasi apakah terdapat pola tertentu dalam pemilihan arsitektur deep learning berdasarkan karakteristik domain aplikasi, seperti medis, industri, penginderaan jauh, dan lingkungan. Dengan memahami hubungan tersebut, dapat diketahui faktor-faktor yang memengaruhi pemilihan metode IQA pada berbagai konteks penggunaan.



Gambar 3.2 Karakteristik Kebutuhan Domain Aplikasi pada Penelitian IQA

Tabel 3.2 Karakteristik Domain dan Metode IQA yang Digunakan

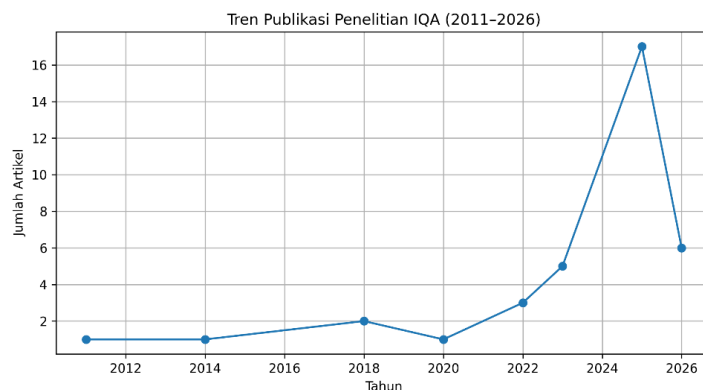
Domain	Metode Domain	Alasan Pemilihan
Medis	CNN, Transformer, Attention	Akurasi dan interprebilitas tinggi
Industri	CNN, Dual-Pipeline	Efisiensi dan real-time
Penginderaan Jauh	Attention, Transformer	Menangani distorsi kompleks
Umum	Hybrid dan NR-IQA	Fleksibilitas berbagai kondisi

Analisis terhadap literatur terkini menunjukkan pola yang cukup sistematis antara karakteristik domain dan pilihan arsitektur IQA. Sistem IQA pada domain medis dituntut untuk mampu memberikan penjelasan yang konsisten agar dapat diterima oleh klinisi, sehingga aspek explainability menjadi pertimbangan utama di atas sekadar akurasi numerik [11]. Dalam konteks kolonoskopi, pendekatan semantic segmentation berbasis deep learning dipilih karena mampu memberikan penilaian objektif yang selaras dengan standar prosedur klinis [12].

Pada domain industri, prioritasnya bergeser ke efisiensi dan kemampuan real-time; algoritma visual berbasis rule multi-kriteria lebih cocok untuk perangkat IIoT berbiaya rendah dibanding model deep learning berat [15], arsitektur dual-pipeline diterapkan untuk memisahkan tahap grading dan deteksi cacat agar beban komputasi dapat dikelola secara lebih efisien [14]. Kompleksitas distorsi atmosferik dan geometris mendorong penggunaan attention mechanism pada domain penginderaan jauh, sebagaimana diterapkan pada penilaian kualitas citra ISAR [3], dan evaluasi resolusi optis kamera UAV. [16]. Adapun pada domain umum lintas aplikasi, ekstraksi fitur distorsi multidimensi lebih tepat digunakan pada citra infrastruktur jaringan listrik yang memiliki karakteristik noise sangat spesifik [2]. Sistem analisis drywall otomatis berbasis computer vision untuk pemantauan progres dan kontrol kualitas dalam konstruksi bangunan menunjukkan bahwa kebutuhan domain konstruksi mendorong penggunaan pendekatan berbasis deep learning yang ringan dan terintegrasi dengan sistem inspeksi lapangan [38]. Pada domain lingkungan, metode transformasi data tabular ke citra membuka peluang baru untuk analisis kualitas udara dalam ruangan, sebuah pendekatan lintas domain yang masih sangat jarang dieksplorasi dalam literatur IQA [29]. Pola yang teridentifikasi adalah: semakin tinggi tuntutan akurasi dan interpretabilitas suatu domain, semakin kompleks arsitektur yang dipilih, dengan konsekuensi langsung pada kebutuhan komputasi dan ukuran dataset pelatihan.

3.3. RQ3 : Bagaimana tren perkembangan metode IQA dari tahun ke tahun berdasarkan literatur terkini dan ke arah mana penelitian IQA berbasis deep learning berkembang?

RQ3 bertujuan untuk mengidentifikasi tren perkembangan penelitian Image Quality Assessment (IQA) berdasarkan literatur yang dianalisis. Analisis ini dilakukan untuk memahami perubahan pendekatan yang digunakan dalam penelitian IQA dari waktu ke waktu, mulai dari metode berbasis fitur tradisional hingga penerapan deep learning modern seperti CNN, attention mechanism, transformer, dan multimodal large language models (MLLM). Selain itu, analisis ini juga bertujuan untuk mengidentifikasi arah perkembangan penelitian IQA di masa mendatang berdasarkan pola publikasi dan evolusi metode yang ditemukan dalam 36 artikel yang ditinjau.

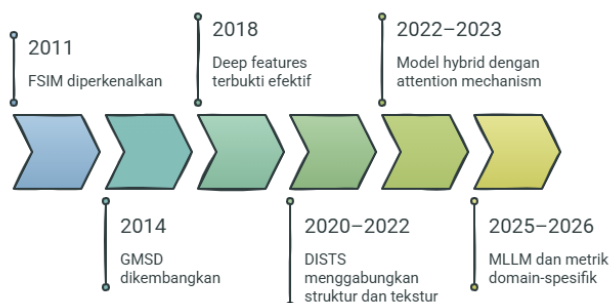


Gambar 3.3 Tren Publikasi Penelitian IQA (2011-2026)

Berdasarkan Gambar 3.3, jumlah publikasi terkait Image Quality Assessment (IQA) menunjukkan tren peningkatan yang cukup signifikan dalam beberapa tahun terakhir. Pada periode 2011–2020, jumlah penelitian masih relatif terbatas dan didominasi oleh pengembangan metrik kualitas citra berbasis fitur tradisional. Namun, mulai tahun 2022 terjadi peningkatan jumlah publikasi yang lebih konsisten, yang mencapai puncaknya pada periode 2025–2026.

Peningkatan tersebut menunjukkan semakin besarnya perhatian peneliti terhadap pengembangan metode IQA, khususnya yang memanfaatkan teknologi deep learning. Selain dipengaruhi oleh perkembangan arsitektur seperti CNN, Transformer, dan Hybrid, peningkatan publikasi juga didorong oleh semakin luasnya penerapan IQA pada berbagai domain, seperti medis, industri, penginderaan jauh, dan sistem cerdas berbasis kecerdasan buatan. Temuan ini mengindikasikan bahwa IQA telah berkembang dari sekadar evaluasi kualitas citra menjadi komponen penting dalam berbagai sistem pengambilan keputusan berbasis visual.

Evolusi Metode IQA: Dari Sinyal ke Deep Learning



Gambar 3.4 Evolusi Metode IQA

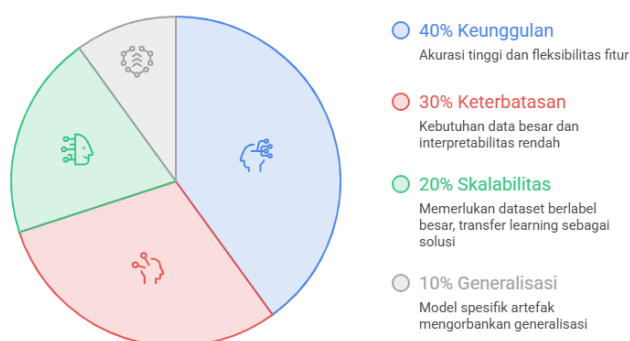
Tren perkembangan metode IQA menunjukkan evolusi yang sangat jelas dari metrik berbasis sinyal tradisional menuju pendekatan berbasis representasi pembelajaran mendalam. Metrik klasik seperti FSIM

[6] dan GMSD [4] yang bertumpu pada fitur hand-crafted menjadi titik awal yang kemudian digantikan secara bertahap oleh model berbasis deep features. Peralihan paradigma ini ditandai dengan pembuktian bahwa fitur dalam dari jaringan deep learning pra-latih memberikan representasi perseptual yang lebih efektif dibanding metrik yang dirancang khusus [8]. Pada periode 2020–2022, tren bergerak ke arah model yang menyatukan kemiripan struktur dan tekstur secara simultan melalui pendekatan DISTS, yang terbukti lebih robust terhadap distorsi geometris dibanding SSIM [5]. Pada rentang 2022–2023, muncul model hybrid yang mengombinasikan attention mechanism dengan statistik berbasis domain, seperti MAMIQA [17] dan model fusion optimasi [39].

Memasuki 2025–2026, tren paling menonjol adalah penggunaan Multimodal Large Language Models (MLLM) untuk NR-IQA, sebagaimana dievaluasi oleh model GPT-4V dan Claude.AI [19], serta pengembangan metrik domain-spesifik seperti MalScore untuk dataset visual malware [40] dan kerangka evaluasi kualitas citra dermoskopi. [41]. Tinjauan komprehensif mengenai color IQA juga mengonfirmasi bahwa penelitian terbaru bergerak menuju model yang sensitif terhadap persepsi warna manusia, melampaui sekadar evaluasi struktur [42]. Secara keseluruhan, tren ini mengarah pada tiga karakteristik model IQA masa depan: berbasis semantik, lintas-modal, dan dapat beroperasi tanpa referensi dalam kondisi distorsi autentik yang kompleks.

3.4. RQ4 : Apa keunggulan dan keterbatasan kritis metode IQA berbasis deep learning di bandingkan dengan metode tradisional, ditinjau dari aspek akurasi, skalabilitas, dan kemampuan generalisasi lintas domain?

Keunggulan dan Keterbatasan Deep Learning dalam IQA



Gambar 3.5 Metode IQA

Evaluasi kritis terhadap literatur terkini mengungkap bahwa keunggulan deep learning dalam IQA bersifat kondisional, bukan mutlak. Dari sisi akurasi, metode berbasis deep learning secara konsisten mengungguli metrik tradisional: Metode DISTS terbukti mencapai korelasi PLCC lebih dari 0,90 pada dataset TID2013, melampaui SSIM yang hanya sekitar 0,78 [5], Metrik analitik Perceptual Error Logarithm (PEL) terbukti mampu bersaing dengan beberapa model deep learning pada kondisi distorsi standar, sekaligus menawarkan efisiensi komputasi yang signifikan karena tidak memerlukan proses pelatihan model [43]. Dari sisi skalabilitas, model deep learning memerlukan dataset berlabel besar yang sulit diperoleh pada domain spesifik. Transfer learning terbukti menjadi solusi efektif pada dataset bedah yang sangat terbatas, meskipun dengan risiko bias terhadap domain sumber [44]. Studi LPIPS membuktikan bahwa fitur deep dari jaringan pra-latih meski bukan dirancang khusus untuk IQA terbukti menjadi metrik perseptual yang sangat efektif, bahkan melampaui metrik tradisional hand-crafted pada berbagai jenis distorsi autentik [8].

Keterbatasan generalisasi menjadi isu paling kritis. Model IQA MRI yang dirancang artifact-specific dan content-specific terbukti sengaja mengorbankan generalisasi demi akurasi pada artefak tertentu [24],

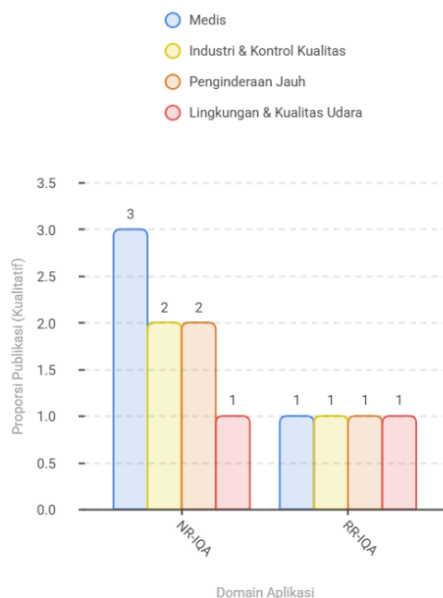
Metode kompensasi noise pada sensor laser terbukti tidak dapat langsung ditransfer ke jenis sensor yang berbeda, mengonfirmasi keterbatasan generalisasi lintas perangkat [32]. Pendekatan berbasis neutrosophic similarity score terbukti mampu menangani ketidakpastian citra medis yang tidak dapat ditangani oleh metode deterministic tradisional [31]. Metode berbasis Variational Auto-Encoder dengan contrastive attention terbukti mampu mengatasi kompleksitas dehazing yang gagal ditangani oleh PSNR/SSIM konvensional [25]. Secara kritis, deep learning unggul dalam akurasi dan fleksibilitas fitur tetapi lemah dalam interpretabilitas, kebutuhan data, dan generalisasi; sementara metode tradisional unggul dalam efisiensi dan transparansi tetapi tidak mampu menangkap distorsi kompleks yang bersifat semantic.

3.5. RQ5 : Bagaimana profil bibliometric penelitian IQA terkini ditinjau dari distribusi domain aplikasi, jenis pendekatan (FR/RR/NR), dan sumber publikasi, serta apa implikasinya terhadap arah penelitian yang masih kurang terjamah?

RQ5 bertujuan untuk menganalisis profil bibliometrik penelitian IQA berdasarkan distribusi domain aplikasi, jenis pendekatan IQA, serta sumber publikasi yang digunakan dalam 36 artikel yang dianalisis. Selain itu, analisis ini dilakukan untuk mengidentifikasi area penelitian yang masih kurang dieksplorasi dan berpotensi menjadi research gap pada penelitian mendatang.

Distribusi Pendekatan Penelitian IQA Berdasarkan Domain Aplikasi

NR-IQA mendominasi di sebagian besar domain aplikasi, sementara RR-IQA sangat terbatas di semua domain.



Gambar 3.6 Evaluasi Performa dalam IQA

Tabel 3.3 Distribusi Domain dan Pendekatan IQA

Domain	NR-IQA	RR-IQA	Temuan
Medis	Dominan	Terbatas	Fokus pada akurasi klinis
Industri	Dominan	Terbatas	Fokus real-time
Penginderaan Jauh	Dominan	Terbatas	Distorsi kompleks
Lingkungan	Sangat Terbatas	Sangat Terbatas	Masih minim penelitian

Tabel 3.4 Sumber Publikasi Dominan

Sumber Publikasi	Karakteristik
Sensors	Publikasi terbanyak pada domain aplikasi IQA
Electronics	Fokus pada implementasi system dan deep learning
Journal of Imaging	Banyak mempublikasi metode dan evaluasi IQA
IEEE/ACM Proceedings	Fokus pada inovasi algoritma dan benchmark

Analisis bibliometrik terhadap 36 literatur yang dikaji mengungkap pola distribusi yang signifikan dan memiliki implikasi strategis bagi arah penelitian ke depan. Dari sisi domain aplikasi, penelitian IQA terkonsentrasi pada tiga kluster utama: domain medis (mencakup kolonoskopi [12], embrio [26], citra fundus, jahitan bedah [2], [44], dan MRI [24]), domain industri dan kontrol kualitas (mencakup mangga [14], IIoT [15], dan jaringan listrik [2]), serta domain penginderaan jauh (mencakup ISAR [3] dan UAV [16]). Domain yang relatif kurang terwakili dalam literatur terkini adalah lingkungan hidup dan kualitas udara, sebagaimana ditunjukkan dalam pendekatan data-to-image untuk analisis kualitas udara dalam ruangan menjadi salah satu penelitian pionir pada domain lingkungan yang masih sangat kurang terwakili dalam literatur IQA [29]. Selain itu, domain konstruksi mulai muncul sebagai area aplikasi IQA yang baru, yang direpresentasikan oleh sistem inspeksi drywall otomatis berbasis computer vision, meskipun jumlah penelitian pada domain ini masih sangat terbatas dibandingkan domain medis maupun industri manufaktur [38]. Dari sisi jenis pendekatan, NR-IQA mendominasi publikasi terbaru karena relevansinya yang lebih tinggi di kondisi nyata, sementara Reduced-Reference IQA (RR-IQA) masih sangat terbatas; Tinjauan sistematis mengonfirmasi bahwa RR-IQA merupakan area yang paling kurang berkembang di antara ketiga kategori IQA [23].

Dari sisi sumber publikasi, jurnal multidisiplin berbasis open access seperti Sensors, Electronics, dan Journal of Imaging (MDPI) mendominasi literatur IQA terkini, mencerminkan aksesibilitas yang tinggi namun juga mengisyaratkan perlunya validasi lebih ketat pada venue bereputasi tinggi seperti IEEE Transactions. Metrik evaluasi untuk segmentasi citra yang merupakan komponen kritis dalam banyak pipeline IQA masih memerlukan standarisasi lebih lanjut [45]. Evaluasi berbasis saliency visual manusia masih sangat jarang diintegrasikan ke dalam framework IQA standar, padahal relevansinya tinggi untuk aplikasi yang berorientasi pada persepsi pengguna [27]. Secara keseluruhan, implikasi bibliometrik ini menunjukkan tiga research gap utama: pertama, kurangnya penelitian RR-IQA pada domain spesifik; kedua, minimnya studi IQA pada domain lingkungan dan kualitas udara; dan ketiga, belum adanya kerangka evaluasi terpadu yang mengintegrasikan saliency visual dengan metrik objektif.

4. Kesimpulan

Penelitian ini berhasil menganalisis perkembangan metode *Image Quality Assessment* (IQA) berbasis *deep learning* melalui *Systematic Literature Review* terhadap 36 artikel terindeks Scopus. Hasil kajian menunjukkan bahwa CNN masih menjadi metode yang paling dominan digunakan karena efisiensi dan kemudahan implementasinya, sementara Transformer dan pendekatan Hybrid menawarkan performa yang lebih baik dalam menangani distorsi kompleks melalui pemanfaatan konteks global dan mekanisme perhatian (*attention mechanism*). Pemilihan metode IQA terbukti dipengaruhi oleh karakteristik domain aplikasi, di mana domain medis menekankan akurasi dan interpretabilitas, domain industri berfokus pada efisiensi dan kemampuan real-time, serta domain penginderaan jauh membutuhkan metode yang mampu menangani distorsi yang kompleks. Selain itu, tren penelitian menunjukkan pergeseran dari metrik tradisional berbasis fitur menuju pendekatan *deep learning*, Transformer, dan pemanfaatan kecerdasan buatan multimodal. Meskipun metode *deep learning* memberikan akurasi yang lebih tinggi dibandingkan metode tradisional, masih terdapat tantangan pada aspek interpretabilitas, kebutuhan data yang besar, dan kemampuan generalisasi lintas domain. Analisis bibliometrik juga menunjukkan dominasi pendekatan NR-IQA dalam berbagai domain aplikasi, sedangkan penelitian RR-IQA dan penerapan IQA pada bidang lingkungan masih relatif terbatas. Temuan ini mengindikasikan bahwa pengembangan metode RR-IQA yang lebih adaptif, integrasi saliency visual dengan metrik objektif, serta perluasan penerapan IQA pada domain yang masih kurang dieksplorasi menjadi peluang penelitian yang menjanjikan di masa mendatang.

References

- [1] Herath HMSS, Herath HMKKMB, Madusanka N, et al. A Systematic Review of Medical Image Quality Assessment. *J Imaging*. 2025;11(4):1–32.
- [2] Chen Z, Du J, Li J, et al. MDFN: Enhancing Power Grid Image Quality Assessment via Multi-Dimension Distortion Feature. *Sensors*. 2025;25(11):18.
- [3] Zhang J, Zhao Z, Tian X. ISAR Image Quality Assessment Based on Visual Attention Model. *Appl Sci*. 2025;15(4).
- [4] Xue W, Zhang L, Mou X, et al. Gradient magnitude similarity deviation: A highly efficient perceptual image quality index. *IEEE Trans Image Process*. 2014;23(2):684–695.
- [5] Ding K, Ma K, Wang S, et al. Image Quality Assessment: Unifying Structure and Texture Similarity. *IEEE Trans Pattern Anal Mach Intell*. 2022;44(5):2567–2581.
- [6] Zhang L, Zhang L, Mou X, et al. FSIM: A Feature Similarity Index for Image Quality Assessment. *IEEE Trans Image Process*. 2011;20(8):2378–2386.
- [7] Bosse S, Maniry D, Müller KR, et al. Deep Neural Networks for No-Reference and Full-Reference Image Quality Assessment. *IEEE Trans Image Process*. 2018;27(1):206–219.
- [8] Severo D, Theis L, Ballé J. the Unreasonable Effectiveness of Linear Prediction As a Perceptual Metric. *12th Int Conf Learn Represent ICLR 2024*. 2024;(1).
- [9] Li L xin, Chen C zi, Zhao N xuan. Blind image quality assessment based on transformer. <https://doi.org/101117/122643493> [Internet]. 2022 [cited 2026 Jun 10];12342:855.
- [10] Wang H, Li L, Wang W, et al. Development and validation of a deep learning image quality feedback system for infant fundus photography. *Sci Rep*. 2025;15(1):1–12.
- [11] Ozer C, Guler A, Turkvatan Cansever A, et al. Consistent explainable image quality assessment for medical imaging. *Heal Inf Sci Syst*. 2026;14(1).
- [12] Vulpoi RA, Ciobanu A, Drug VL, et al. Deep Learning-Based Semantic Segmentation for Objective Colonoscopy Quality Assessment. *J Imaging*. 2025;11(3):1–20.
- [13] Telesco LG, Nejamkin D, Mata E, et al. Semi-supervised multi-task learning for interpretable quality assessment of fundus images. *Biomed Signal Process Control* [Internet]. 2026 [cited 2026 Jun 8];113.
- [14] Lin S, Chiu H. From Quality Grading to Defect Recognition: A Dual-Pipeline Deep Learning Approach for Automated Mango Assessment. *Electron*. 2026;15(3).
- [15] Lech P. Multi-Criteria Visual Quality Control Algorithm for Selected Technological Processes Designed for Budget IIoT Edge Devices. *Electron*. 2025;14(16):20.
- [16] Zachar P, Wilk Ł, Pilarska-mazurek M, et al. Assessment of UAV image quality in terms of optical resolution. *Int Arch Photogramm Remote Sens Spat Inf Sci Vol XLVIII-1/W4-2025 EuroCOW 2025 – Eur Work Calibration Orientat Remote Sensing*, 16–18 June 2025, Warsaw, Pol Assess. 2025;XLVIII(June):16–18.
- [17] Yu L, Li J, Pakdaman F, et al. MAMIQA: No-Reference Image Quality Assessment Based on Multiscale Attention Mechanism With Natural Scene Statistics. *IEEE Signal Process Lett*. 2023;30:588–592.
- [18] Yang Y, Idris NB, Wahab AWA, et al. An image quality assessment algorithm based on ‘global + local’ feature fusion. *PeerJ Comput Sci*. 2025;11:1–25.
- [19] Varga D. Comparative Evaluation of Multimodal Large Language Models for No-Reference Image Quality Assessment with Authentic Distortions: A Study of OpenAI and Claude.AI Models. *Big Data Cogn Comput*. 2025;9(5).
- [20] Dobre-Baron R, Savu-Jivanov A, Ancuți C. Evaluating Image Quality Metrics as Loss Functions for Image Dehazing †. *Sensors*. 2025;25(15):1–14.
- [21] Felicetti F, Lamonaca F, Carni DL, et al. Image Quality Standardization in Radiomics: A Systematic Review of Artifacts, Variability, and Feature Stability. *Sensors*. 2026;26(3):1–22.
- [22] Haddaway NR, Page MJ, Pritchard CC, et al. PRISMA2020: An R package and Shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and Open Synthesis. *Campbell Syst Rev*. 2022;18(2):1–12.
- [23] Dost S, Saud F, Shabbir M, et al. Reduced reference image and video quality assessments: review of methods. *Eurasip J Image Video Process* [Internet]. 2022;2022(1):1–31.
- [24] Lei K, Syed AB, Zhu X, et al. Artifact- and content-specific quality assessment for MRI with image rulers. *Med Image Anal* [Internet]. 2022 [cited 2026 Jun 8];77:102344.
- [25] Vishwakarma S, Pillai A, Punj D. An Enhancement in Single-Image Dehazing Employing

- Contrastive Attention over Variational Auto-Encoder (CA-VAE) Method. *Int J Math Eng Manag Sci.* 2023;8(4):738–754.
- [26] Yanai A, Horie A, Sakurai A, et al. Innovative AI models for clinical decision-making : predicting blastocyst formation and quality from time-lapse embryo images up to embryonic. *Comput Biol Med [Internet].* 2025;195(February):110637.
- [27] Zhang Y, Song J, Gu S, et al. Saliency-Bench: A Comprehensive Benchmark for Evaluating Visual Explanations. *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. Association for Computing Machinery;* 2025.
- [28] Mohammadi S, Jenadeleh M, Sneyers J, et al. Evaluation of Objective Image Quality Metrics for High-Fidelity Image Compression. *IEEE Access [Internet].* 2025;1–19.
- [29] Isaac A, Barrera P, Rodríguez LM, et al. Tabular Data-to-Image Methods for Improved Indoor Air Quality Analysis and Risk Detection. 8th Int. Conf. Syst. Eng. - Cybersecurity & AI Build. a Reliab. Digit. Futur. (CIIS 2025), Oct. 0103, 2025, Lima, Peru. Association for Computing Machinery; 2025.
- [30] Testolina M, Lazzarotto D, Rodrigues R, et al. On the Performance of Subjective Visual Quality Assessment Protocols for Nearly Visually Lossless Image Compression. *MM 2023 - Proc. 31st ACM Int. Conf. Multimed. Association for Computing Machinery;* 2023.
- [31] Rachel J, Devarasan E. Multi-Scale Filter Analysis with a Neutrosophic Similarity Score-Based Enhancement Framework. *Int J Comput Intell Syst.* 2026;19(1).
- [32] Rotter P, Knapik D, Klemiato M, et al. Compensation of Speckle Noise in 2D Images from Triangulation Laser Profile Sensors Using Local Column Median Vectors with an Application in a Quality Control System. *Sensors.* 2025;25(11):1–20.
- [33] Zhang Y, Fang H, Ng HP. End-to-end Viewpoint Optimization for Inspection based on Defect Detectability Assessment. 2025 5th Int. Conf. Artif. Intell. Appl. Technol. AIAT 2025. Association for Computing Machinery; 2026.
- [34] Su S, Yan Q, Zhu Y, et al. Su_Bindly_Assess_Image_Quality_in_the_Wild_Guided_by_a_CVPR_2020_paper. *Proc IEEE/CVF Conf Comput Vis Pattern Recognit.* 2020;3667–3676.
- [35] Arct B, Świdarski B, Róžańska MA, et al. Comparative Evaluation of CNN and Transformer Architectures for Flowering Phase Classification of *Tilia cordata* Mill. with Automated Image Quality Filtering. *Sensors.* 2025;25(17):1–37.
- [36] Xie Y, Zhao Y, Xia T, et al. No-Reference Plane Image Quality Assessment Based on Differential Feature Fusion. 7th Int. Conf. Video, Signal Image Process. (VSIP 2025), Novemb. 1416, 2025, Kunming, China. Association for Computing Machinery; 2025.
- [37] Yang Y, Liu C, Wu H, et al. A quality assessment algorithm for no-reference images based on transfer learning. *PeerJ Comput Sci.* 2025;11(2024).
- [38] Trzeciakiewicz M, Barreiro AC, Gard N, et al. Automatic Drywall Analysis for Progress Tracking and Quality Control in Construction. *Proc Int Jt Conf Comput Vision, Imaging Comput Graph Theory Appl.* 2025;3(Visigrapp):707–716.
- [39] Varga D. An Optimization-Based Family of Predictive, Fusion-Based Models for Full-Reference Image Quality Assessment. *J Imaging.* 2023;9(6).
- [40] Czaplicki J, Rahouti M, Chehri A, et al. MalScore: A Quality Assessment Framework for Visual Malware Datasets Using No-Reference Image Quality Metrics. *Futur Internet.* 2025;17(12):1–18.
- [41] Ferraris A, Branciforti F, Meiburger KM, et al. No-Reference Quality Assessment of Dermoscopic Images Using Minimal Expert Supervision. *Appl Sci [Internet].* 2026 [cited 2026 Jun 8];16(4).
- [42] Gao M, Song C, Zhang Q, et al. Research Progress on Color Image Quality Assessment. *J Imaging.* 2025;11(9):1–41.
- [43] Bezerra SAC, Bezerra Júnior SAC, De S. Pio JL, et al. Perceptual Error Logarithm: An Efficient and Effective Analytical Method for Full-Reference Image Quality Assessment. *IEEE Access.* 2025;13:68587–68606.
- [44] Ishchenko R, Solopov M, Popandopulo A, et al. Evaluation of Transfer Learning Efficacy for Surgical Suture Quality Classification on Limited Datasets. *J Imaging.* 2025;11(8):1–14.
- [45] Rottmayer N, Redenbach C. A Novel Distance-Based Metric for Quality Assessment in Image Segmentation. *IET Image Process.* 2026;20(1):1–9.