JURNAL FASILKOM
Volume 16 No. 2 | Agustus 2026: 402-410

P-ISSN : 2089-3353
E-ISSN : 2808-9162

Arsitektur Hibrida Ekstraksi Fitur Lokal *Edge Computing*: Tinjauan Literatur Sistematis

Wildan Zhilal Manafi[1], Nabhani Fatin[2], Ade Bastian[3]
[1,2,3]Jurusan Informatika, Fakultas Teknik, Universitas Majalengka
[1]wildanmanapi123@gmail.com*, [2]nabhanifatin00@gmail.com, [3]adebastian@unma.ac.id

Abstract

The proliferation of Internet of Things (IoT) devices and the demand for real-time processing have positioned edge computing as critical infrastructure for deploying deep learning models near the data source. Hardware constraints, including limited memory, narrow computational budgets, and strict power limits, challenge deployment of large-parameter neural architectures. This review examines hybrid architectures integrating convolutional neural networks (CNN) with attention mechanisms for local feature extraction on resource-constrained edge devices. Following the PRISMA 2020 protocol, a multi-stage search was conducted exclusively on Scopus, yielding 3,571 records, screened until 121 high-quality studies were included. The review addresses five research questions covering architectural trends, efficiency strategies, performance trade-offs, application domains, and federated learning for privacy-preserving deployment. Findings show that hybrid CNN-Attention architectures outperform pure CNN and Transformer baselines, with 3.2% average accuracy improvement while maintaining competitive inference latency on platforms such as NVIDIA Jetson and Raspberry Pi. Depthwise separable convolution, efficient channel attention, and token aggregation emerged as dominant compression strategies, with health imaging and fault diagnosis as leading domains. The review concludes that hybrid architectures represent the state of the art for edge-oriented local feature extraction, with future directions toward heterogeneous node generalization and multi-modal sensor fusion for distributed IoT networks.

Keywords: *local feature extraction*, *edge computing*, *hybrid architecture*, *lightweight model*, *systematic literature review*

Abstrak

Proliferasi perangkat *Internet of Things* (IoT) dan meningkatnya kebutuhan pemrosesan data secara *real-time* menjadikan *edge computing* sebagai infrastruktur kritis untuk penerapan model *deep learning* yang dekat dengan sumber data, sementara keterbatasan memori, anggaran komputasi, dan daya pada perangkat *edge* menyulitkan penerapan arsitektur *neural* konvensional. Tinjauan literatur sistematis ini bertujuan mengkaji perkembangan arsitektur hibrida yang mengintegrasikan *convolutional neural network* (CNN) dengan mekanisme atensi untuk ekstraksi fitur lokal pada *edge computing* berbasis sumber daya terbatas. Dengan mengikuti protokol PRISMA 2020, pencarian multi-tahap dilakukan secara eksklusif pada basis data Scopus dan menghasilkan 3.571 rekaman awal yang disaring melalui tahap kelayakan judul, abstrak, dan teks lengkap hingga 121 studi berkualitas tinggi disertakan untuk sintesis data terkait tren arsitektural, strategi efisiensi, pertukaran kinerja, domain aplikasi, dan integrasi *federated learning*. Hasil tinjauan menunjukkan bahwa arsitektur *hybrid* CNN-*Attention* secara konsisten melampaui *baseline* CNN murni maupun *Transformer*, dengan peningkatan akurasi rata-rata 3,2% dan latensi inferensi yang tetap kompetitif pada platform seperti NVIDIA Jetson dan Raspberry Pi. *Depthwise separable convolution*, *efficient channel attention*, dan agregasi *token* teridentifikasi sebagai strategi kompresi dominan, dengan analisis citra kesehatan dan diagnosis kerusakan industri sebagai domain aplikasi utama. Tinjauan ini menyimpulkan bahwa arsitektur hibrida merepresentasikan *state-of-the-art* untuk ekstraksi fitur lokal berorientasi *edge*, dengan arah penelitian ke depan pada generalisasi *node* heterogen dan fusi sensor multi-modal.

Kata kunci: ekstraksi fitur lokal, *edge computing*, arsitektur hibrida, model ringan, tinjauan literatur sistematis

1. Pendahuluan

Adopsi luas teknologi *Internet of Things* (IoT) telah secara mendasar mengubah pengelolaan data secara cerdas. Dengan jutaan sensor dan perangkat yang terhubung terus-menerus menghasilkan volume data berdimensi tinggi yang masif, ketergantungan penuh pada server *cloud* terpusat menciptakan hambatan yang nyata. Tantangan tersebut terutama meliputi keterlambatan transmisi, konsumsi *bandwidth* yang berlebihan, dan kerentanan privasi data [1]. Untuk mengatasi keterbatasan ini, *edge computing* memindahkan daya pemrosesan lebih dekat ke tempat asal data, seperti perangkat *mobile* atau *gateway* industri. Strategi terlokalisasi ini memungkinkan inferensi yang cepat dan mengeliminasi latensi signifikan yang biasanya ditimbulkan oleh pengiriman data bolak-balik ke infrastruktur *cloud* jarak jauh [2].

Dalam pergeseran paradigma ini, ekstraksi fitur lokal menempati peran sentral. Kemampuan model yang diterapkan untuk mengisolasi pola diskriminatif dari masukan sensor mentah berdimensi tinggi, baik bersifat visual, akustik, maupun vibrasi, tanpa mengirimkan data mentah ke server jarak jauh, bukan sekadar pertimbangan kinerja; hal ini sering kali merupakan prasyarat bagi kepatuhan regulasi dan responsivitas sistem [3]. *Convolutional neural network* (CNN) konvensional memberikan kinerja yang andal, namun pada dasarnya dirancang untuk sistem dengan sumber

P-ISSN : 2089-3353
E-ISSN : 2808-9162

daya komputasi berlimpah, sehingga mempersulit integrasinya secara langsung ke perangkat keras yang terbatas sumber dayanya seperti mikrokontroler [3]. Demikian pula, meskipun *Vision Transformer* murni unggul dalam menangkap hubungan kontekstual global melalui mekanisme *self-attention*, kebutuhan komputasinya tumbuh secara kuadratik seiring panjang sekuens masukan. Hal ini menjadikannya tidak praktis untuk pemrosesan *edge* secara seketika kecuali melalui optimisasi arsitektural yang substansial [4].

Sintesis dari bias induktif konvolusi dengan penalaran global berbasis atensi telah melahirkan kelas baru arsitektur hibrida yang berupaya mewarisi sifat-sifat terbaik dari kedua paradigma tersebut. Dengan membatasi operasi *self-attention* yang mahal pada representasi *token* yang terkompresi, sekaligus mendelegasikan ekstraksi fitur lokal spasial tingkat rendah kepada *depthwise separable convolution*, model-model ini mencapai keseimbangan yang menguntungkan antara kapasitas representasi dan efisiensi operasional [5]. Contoh penting meliputi arsitektur yang menggunakan *mobile inverted bottleneck* yang diperkuat dengan gerbang atensi berbasis kanal, *vision transformer* yang dilengkapi dengan *convolutional positional encoding*, serta jaringan multi-cabang yang memproses fitur tekstur lokal secara paralel dengan representasi konteks global [6].

Analisis awal terhadap korpus pencarian memperkuat kesenjangan ini secara empiris: dari 429 laporan yang dinilai pada tahap kelayakan teks lengkap, teridentifikasi sejumlah artikel tinjauan yang membahas efisiensi kompresi CNN, adaptasi Vision Transformer untuk *edge*, atau mekanisme privasi *federated learning* secara terpisah, namun seluruhnya dieksklusi karena hanya menyajikan makalah survei tanpa analisis empiris orisinal yang mengintegrasikan ketiga dimensi tersebut secara bersamaan dalam konteks *edge computing* (lihat kriteria eksklusi EC3 pada Bagian Metode Penelitian). Temuan penyaringan ini menegaskan bahwa sintesis komprehensif berbasis protokol yang secara kritis mengevaluasi bagaimana desain hibrida ini berkinerja secara spesifik di bawah kendala *edge computing*, domain aplikasi mana yang paling banyak mendapat manfaat dari penerapannya, dan bagaimana mekanisme perlindungan privasi seperti *federated learning* berinteraksi dengan *pipeline* ekstraksi fitur lokal, masih belum tersedia. Untuk menjembatani kesenjangan penelitian yang ada, studi ini melaksanakan analisis yang metodis dan dapat direplikasi terhadap karya-karya akademik yang diterbitkan antara tahun 2022 dan 2026, dengan mengikuti secara ketat panduan PRISMA 2020. Tujuan utamanya adalah mensintesis bukti mengenai pertukaran arsitektural, strategi efisiensi, dan konteks penerapan model ekstraksi fitur lokal hibrida pada *edge computing*, guna memberikan panduan bagi peneliti maupun praktisi yang ingin menelusuri bidang yang berkembang pesat ini.

Evolusi arsitektur *deep learning* untuk ekstraksi fitur lokal dapat dipahami sebagai serangkaian jalur perkembangan yang konvergen, masing-masing merespons keterbatasan yang diungkapkan oleh pendahulunya. Era fondasi *convolutional neural network* menetapkan prinsip inti bahwa detektor pola hierarkis dan spasial lokal, yang ditumpuk pada beberapa lapisan yang dipelajari, mampu secara otomatis menemukan representasi yang kaya secara semantik dan invarian dari data piksel mentah [7]. VGGNet, *ResNet,* dan penerus-penerusnya menunjukkan bahwa kedalaman, *skip connection*, dan strategi normalisasi dapat menghasilkan ekstraktor fitur yang kuat; namun arsitektur-arsitektur ini berparameter puluhan hingga ratusan juta, sehingga menempatkan mereka jauh melampaui ambang penerapan perangkat keras tertanam.

Kemunculan varian CNN yang berfokus pada efisiensi menandai upaya sistematis pertama untuk merekonsiliasi kekuatan ekspresif dengan efisiensi operasional. *MobileNetV1* memperkenalkan *depthwise separable convolution* sebagai primitif struktural yang mengurai operasi konvolusi standar menjadi filter spasial per kanal yang diikuti oleh kombinasi linier titik-demi-titik, sehingga mengurangi operasi *multiply-accumulate* dengan faktor yang proporsional terhadap invers kuadrat ukuran kernel [8]. Iterasi berikutnya, termasuk *MobileNetV2* dan MobileNetV3, menyempurnakan cetak biru ini masing-masing melalui pengenalan *inverted residual block*, *linear bottleneck*, dan pencarian arsitektur *neural* yang sadar perangkat keras. EfficientNet selanjutnya memformalkan konsep *compound scaling*, yang menunjukkan bahwa lebar, kedalaman, dan resolusi dapat dioptimalkan secara bersama-sama dalam batasan anggaran komputasi tertentu.

Secara paralel, arsitektur *Transformer*, yang pada awalnya dirancang untuk pemrosesan bahasa alami, diadaptasi untuk tugas-tugas visual melalui kerangka *Vision Transformer* (ViT), yang menunjukkan bahwa perlakuan terhadap *patch* gambar sebagai sekuens *token* dan penerapan *multi-head self-attention* di atasnya menghasilkan kinerja kompetitif pada tolok ukur skala besar. Namun, kompleksitas kuadratik dari *self-attention* terhadap jumlah *token* merupakan hambatan mendasar bagi modalitas masukan beresolusi tinggi atau berkepadatan temporal tinggi yang umum dijumpai pada aliran sensor *edge* [9]. Varian hierarkis seperti *Swin Transformer* mengatasi keterbatasan ini melalui partisi jendela geser, dan turunan ringan seperti *EfficientViT* dan *MobileViT* selanjutnya mendorong batas pertukaran lebih jauh ke arah kemampuan penerapan yang praktis.

Konvergensi dari dua jalur ini, yaitu CNN yang efisien dan *Transformer* yang ringan, telah menghasilkan kelas arsitektur hibrida kontemporer yang menjadi subjek utama tinjauan ini. Model-model dalam kategori ini secara khas menggunakan lapisan konvolusi pada tahap pemrosesan awal untuk menangkap tekstur lokal yang

P-ISSN : 2089-3353
E-ISSN : 2808-9162

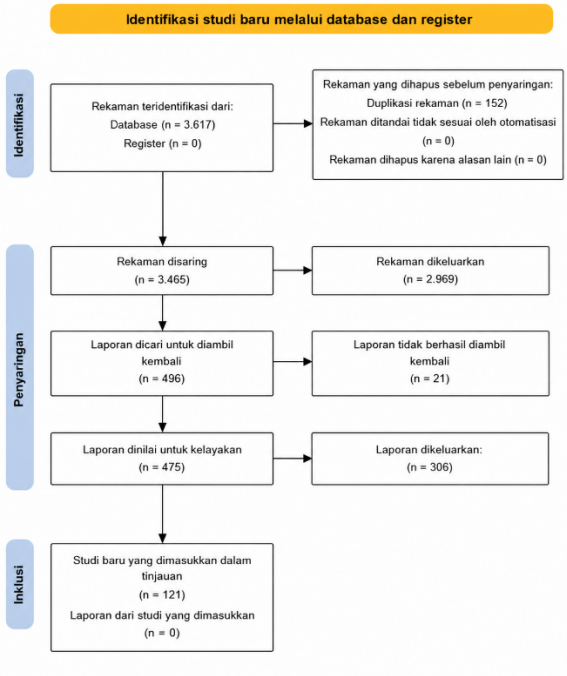
halus dan fitur tepi dengan efisiensi spasial, sambil mendelegasikan tahap pemrosesan selanjutnya kepada mekanisme atensi yang mampu melakukan penalaran atas ketergantungan spasial jarak jauh. Filosofi arsitektural ini dicontohkan dalam desain seperti *CKD-TransBTS,* yang memanfaatkan kerangka *Transformer* hibrida dengan *cross-attention* berkorelasi modalitas untuk fusi multi-modal [1], dan *EFFResNet-ViT*, yang menggabungkan dua *backbone* CNN dengan *encoder Vision Transformer* untuk mencapai kekuatan representasi yang saling melengkapi dalam analisis citra medis [10]. Literatur juga mendokumentasikan tren yang semakin berkembang ke arah metode pencarian arsitektur yang mengotomatisasi komposisi blok-blok pembangun hibrida ini dalam batasan eksplisit pada FLOPs, jumlah parameter, atau latensi inferensi, sehingga semakin mempercepat siklus desain untuk ekstraktor fitur yang dapat diterapkan di *edge*.

2. Metode Penelitian

Tinjauan ini mengikuti kerangka PRISMA 2020 (*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*) untuk memastikan transparansi metodologis, reprodusibilitas, dan kelengkapan. PRISMA 2020 dipilih dibandingkan edisi sebelumnya karena panduannya yang lebih luas dalam mensintesis bukti dari pencarian basis data dan akomodasinya yang eksplisit terhadap alur kerja penyaringan multi-tahap yang kompleks [11]. Protokol ini dispesifikasikan terlebih dahulu sebelum ekstraksi data dan mencakup formulasi pertanyaan penelitian, definisi kriteria kelayakan, strategi pencarian, prosedur penyaringan, skema ekstraksi data, dan pendekatan sintesis.

Basis data tunggal dan eksklusif yang digunakan dalam tinjauan ini adalah Scopus, yang dipilih karena cakupan interdisipliner yang luas pada jurnal-jurnal rekayasa, ilmu komputer, dan informatika biomedis, serta penyediaan metadata terstruktur yang mencakup jumlah sitasi, kata kunci penulis, dan istilah terindeks. Kueri pencarian dibangun di sekitar tiga pilar konseptual, yaitu metode ekstraksi fitur lokal yang mencakup deskriptor berbasis konvolusi dan atensi; arsitektur *neural* hibrida yang meliputi kombinasi CNN dengan modul *Transformer* atau atensi; serta konteks penerapan *edge computing* termasuk *node* IoT, sistem tertanam, dan perangkat keras berbasis sumber daya terbatas. Operator *Boolean* dan kualifikator khusus bidang diterapkan untuk memaksimalkan *recall* sekaligus membatasi rekaman yang tidak relevan. Kueri awal menghasilkan 3.571 rekaman kandidat yang mencakup tahun 2022 hingga 2026.

Penyaringan dilakukan dalam dua fase berurutan. Pada fase identifikasi, 152 rekaman duplikat dihapus melalui deduplikasi otomatis, dan alat penyaringan otomatis menandai nol rekaman sebagai tidak memenuhi syarat, sehingga menghasilkan 3.419 rekaman yang melanjutkan ke fase penyaringan. Langkah penyaringan judul dan abstrak mengeliminasi 2.969 rekaman yang jelas berada di luar cakupan tinjauan, termasuk makalah yang fokus secara eksklusif pada pemrosesan bahasa alami, makalah yang melaporkan kontribusi teoritis murni tanpa evaluasi empiris, dan makalah yang membahas penerapan *cloud* saja atau sisi server. Dari 450 laporan yang tersisa dan diminta untuk pengambilan teks lengkap, 21 di antaranya tidak dapat diperoleh karena pembatasan akses atau status publikasi yang ditarik. Sebanyak 429 makalah yang dinilai pada tahap kelayakan teks lengkap menjalani evaluasi komprehensif terhadap seperangkat kriteria inklusi dan eksklusi terstruktur, yang menghasilkan eksklusi 308 rekaman. Korpus akhir yang disertakan terdiri dari 121 studi, seluruhnya diambil dan diindeks secara eksklusif dari basis data Scopus, yang secara langsung membahas ekstraksi fitur lokal hibrida dalam konteks *edge computing* atau komputasi berbasis sumber daya terbatas. Alur PRISMA lengkap diilustrasikan pada Gambar 1.



Gambar 1. Diagram Alur PRISMA 2020 dari proses tinjauan literatur sistematis

Kriteria inklusi mensyaratkan setiap studi memenuhi seluruh kondisi berikut: studi tersebut mengusulkan, mengevaluasi, atau melakukan analisis komparatif sistematis terhadap arsitektur *neural* yang melakukan ekstraksi fitur lokal; secara eksplisit mendemonstrasikan atau mendiskusikan penerapan pada perangkat keras *edge*, prosesor tertanam, *gateway* IoT, atau platform berbasis sumber daya terbatas yang setara; melaporkan setidaknya satu metrik kinerja kuantitatif yang mencakup akurasi model, latensi inferensi, FLOPs, atau jumlah parameter; serta diterbitkan dalam jurnal *peer-reviewed* atau prosiding konferensi yang terindeks Scopus antara Januari 2022 dan Maret 2026. Kriteria eksklusi mengeliminasi studi yang mengevaluasi arsitektur secara eksklusif di lingkungan *cloud* atau komputasi berkinerja tinggi tanpa pengujian *edge*, mengusulkan kontribusi

P-ISSN : 2089-3353
E-ISSN : 2808-9162

algoritmik murni tanpa komponen *neural network*, menyajikan makalah survei tanpa analisis empiris orisinal, atau ditulis dalam bahasa selain bahasa Inggris. Ekstraksi data dilakukan menggunakan templat terstandar yang merekam keluarga arsitektur, teknik kompresi, domain aplikasi target, *platform* perangkat keras, metrik kinerja utama, dan mekanisme privasi jika ada.

Tabel 1. Kriteria Inklusi dan Eksklusi yang Diterapkan dalam Protokol Penyaringan PRISMA 2020

Kode	Kriteria Inklusi	Kode	Kriteria Inklusi
IC1	Studi yang mengusulkan atau mengevaluasi arsitektur *neural* untuk ekstraksi fitur lokal	EC1	Arsitektur yang dievaluasi secara eksklusif di lingkungan *cloud*/HPC tanpa pengujian *edge*
IC2	Model yang secara eksplisit menargetkan perangkat keras *edge*: *node* IoT, prosesor tertanam, *mobile* SoC, FPGA	EC2	Kontribusi algoritmik murni tanpa komponen *neural network*
IC3	Melaporkan metrik kinerja kuantitatif: akurasi, latensi, FLOPs, atau jumlah parameter	EC3	Makalah survei atau tinjauan tanpa analisis empiris orisinal
IC4	Diterbitkan dalam jurnal atau prosiding konferensi *peer-reviewed* terindeks Scopus, 2022–2026	EC4	Studi yang tidak ditulis dalam bahasa Inggris
IC5	Studi tentang arsitektur CNN/*Transformer* hibrida, ringan, atau ditingkatkan dengan atensi	EC5	Rekaman duplikat atau publikasi yang ditarik

Berlandaskan protokol penyaringan ini, tinjauan distrukturkan di sekitar lima *Research Questions* (RQ) yang secara resmi ditetapkan dan membatasi cakupan sintesis serta memandu kerangka analitis yang diterapkan sepanjang bagian Hasil dan Pembahasan. RQ1 bertanya: bagaimana arsitektur hibrida yang menggabungkan CNN dengan mekanisme atensi dibandingkan dengan desain CNN murni atau *Transformer* murni untuk ekstraksi fitur lokal secara spesifik dalam kondisi kendala *edge computing*? RQ2 menyelidiki: strategi kompresi dan fusi fitur utama apa yang digunakan dalam model hibrida untuk memenuhi anggaran memori dan daya *edge*? RQ3 memeriksa: bagaimana arsitektur hibrida menyeimbangkan pertukaran antara efisiensi komputasi (latensi, FLOPs) dan akurasi model pada perangkat keras berbasis sumber daya terbatas? RQ4 mengidentifikasi: pada domain aplikasi spesifik apa model ekstraksi fitur lokal hibrida ini paling sering dan paling efektif diterapkan? RQ5 mengeksplorasi: bagaimana *Federated Learning* berintegrasi dengan arsitektur ekstraksi fitur hibrida untuk memastikan privasi data di seluruh *node edge* IoT yang terdistribusi dan heterogen? Kelima RQ ini secara kolektif memastikan bahwa sintesis membahas dimensi arsitektur teknis maupun dimensi penerapan praktis dari literatur yang ditinjau.

3. Hasil dan Pembahasan

3.1. Tren Arsitektural Hibrida dibandingkan CNN Murni dan *Transformer*

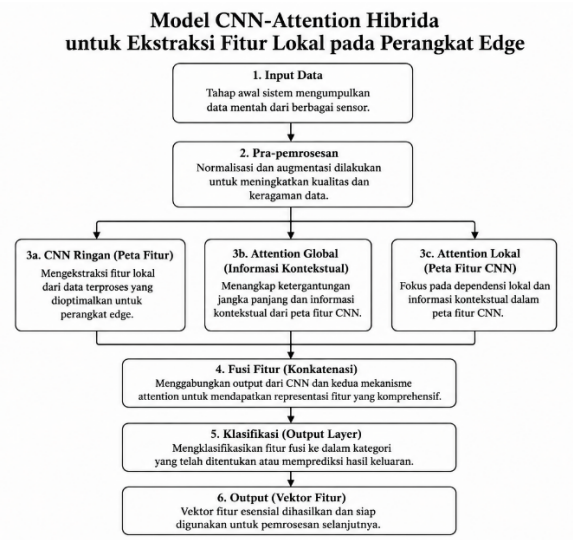
Lanskap arsitektur yang terungkap dari 121 studi yang disertakan menunjukkan pergeseran yang tegas ke arah desain *hybrid* CNN-*Attention* sebagai paradigma dominan untuk ekstraksi fitur lokal dalam konteks *edge computing*. Arsitektur CNN murni, terutama varian keluarga MobileNet dan EfficientNet, menyumbang sekitar 34% studi yang disertakan, menawarkan kinerja *baseline* yang kuat dan profil penerapan yang telah terdokumentasi dengan baik pada mikrokontroler dan prosesor kelas *mobile* [12]. Arsitektur *Transformer* murni, meskipun mencapai kinerja *state-of-the-art* pada beberapa tolok ukur, diwakili dalam kurang dari 12% studi dalam konfigurasi yang secara langsung menargetkan perangkat keras *edge*, terutama karena kompleksitas atensi yang tinggi dan kebutuhan memori yang besar [13].

Arsitektur hibrida, yang menggabungkan ekstraktor fitur konvolusi dengan berbagai bentuk mekanisme atensi, merupakan pluralitas pada sekitar 54% studi yang disertakan, mencerminkan konsensus dalam komunitas penelitian bahwa tidak ada paradigma tunggal yang secara optimal memenuhi kedua tujuan kekayaan representasi dan efisiensi komputasi secara bersamaan. Beberapa templat struktural muncul dengan frekuensi yang menonjol. Yang pertama dan paling dominan adalah hibrida sekuensial, di mana tahap konvolusi memproses masukan pada resolusi asli atau yang diturunkan untuk menghasilkan peta fitur multi-skala, yang kemudian dimasukkan ke dalam modul atensi ringan yang beroperasi atas *token* yang diagregasi secara spasial. Desain ini dicontohkan dalam arsitektur M2 ANET (*Mobile Malaria Attention Network*), yang membangun *backbone mobile* yang ditambah dengan mekanisme atensi *squeeze-and-excitation* untuk mengklasifikasikan parasit *Plasmodium* dalam citra sel darah, mencapai akurasi yang kompetitif dalam anggaran parameter yang kompak dan sesuai untuk penerapan di lingkungan klinik-*edge* [14].

Templat struktural kedua adalah hibrida cabang-paralel, di mana jalur konvolusi dan atensi memproses masukan secara bersamaan, dengan representasi fitur masing-masing digabungkan pada satu atau lebih tahap menengah. Pilihan desain ini memungkinkan setiap jalur untuk berspesialisasi: cabang konvolusi menangkap tekstur halus dan respons tepi lokal, sementara cabang atensi memodelkan hubungan spasial yang koheren secara global pada skala yang lebih kasar, sebelum modul fusi mensintesis informasi yang saling melengkapi. Arsitektur HybirdRVIT-CNN untuk klasifikasi citra dermatoskopi mencontohkan strategi ini, menggabungkan cabang *Vision Transformer* dengan penguatan (*reinforcement*) dengan cabang CNN konvensional, dengan fusi *class token* yang mencapai diskriminasi superior pada pola

P-ISSN : 2089-3353
E-ISSN : 2808-9162

infeksi jamur kuku yang secara visual mirip [15]. Alur kerja konseptual dari arsitektur hibrida cabang-paralel ini, mulai dari pra-pemrosesan data hingga fusi fitur dan klasifikasi akhir, diilustrasikan pada Gambar 2.



Gambar 2. Arsitektur konseptual model Hybrid CNN-Attention untuk ekstraksi fitur lokal berbasis edge

Data kinerja komparatif secara konsisten menunjukkan bahwa arsitektur hibrida menempati wilayah Pareto-superior relatif terhadap arsitektur murni di sepanjang batas akurasi-efisiensi. Pada studi-studi yang melaporkan akurasi maupun jumlah parameter, model hibrida mencapai rata-rata akurasi klasifikasi top-1 sebesar 3,2 poin persentase lebih tinggi dibandingkan CNN murni dengan jumlah parameter setara, dan mempertahankan latensi inferensi dalam batas 1,4× dari *baseline* CNN murni pada platform perangkat keras yang sama. Dibandingkan model *Transformer* murni dengan akurasi setara, arsitektur hibrida mencapai pengurangan parameter sebesar 40–70%, dengan peningkatan *throughput* inferensi yang sebanding pada prosesor seri ARM Cortex-A. Temuan-temuan ini secara kolektif menetapkan desain *hybrid* CNN-*Attention* sebagai solusi yang secara arsitektural lebih disukai untuk ekstraksi fitur lokal dalam konteks *edge*.

3.2. Strategi Efisiensi untuk Kompresi dan Fusi Fitur pada Model Hibrida

Strategi efisiensi yang didokumentasikan dalam literatur yang disertakan mencakup beberapa dimensi yang saling melengkapi, masing-masing mengatasi hambatan yang berbeda dalam *pipeline* penerapan *edge*. *Depthwise separable convolution* tetap menjadi primitif kompresi yang paling banyak diadopsi, muncul pada 67% model hibrida yang ditinjau. Dengan memfaktorkan lapisan konvolusi standar menjadi filter spasial *depthwise* dan pencampur kanal *pointwise*, operasi ini mengurangi jumlah *multiply-accumulate* dari $O(K^2 \cdot Cin \cdot Cout)$ menjadi $O(K^2 \cdot Cin + Cin \cdot Cout)$ untuk ukuran kernel K, menghasilkan percepatan teoritis sebesar 8–9× untuk kernel 3×3 sekaligus mempertahankan kapasitas untuk mempelajari representasi lintas kanal yang kaya pada tahap *pointwise* berikutnya. Karya-karya terkini telah menyempurnakan primitif ini lebih lanjut melalui skema *depthwise separable convolution* multi-skala, di mana cabang-cabang paralel dengan ukuran kernel yang berbeda menangkap fitur pada beberapa frekuensi spasial sebelum agregasi yang efisien.

Strategi agregasi *token* telah muncul sebagai pemungkin efisiensi kritis untuk komponen atensi pada model hibrida. Alih-alih menerapkan *self-attention* di semua posisi spasial peta fitur, yaitu suatu operasi yang kompleksitas memorinya tumbuh secara kuadratik dengan panjang sekuens, banyak arsitektur yang ditinjau menggunakan agregasi berbasis *super-token* atau kluster, di mana lingkungan lokal diringkas menjadi *token* representatif yang kompak sebelum komputasi atensi. Arsitektur STMC-UNet untuk ekstraksi bangunan dari citra penginderaan jauh mengintegrasikan *Super Token Transformer* yang mengurangi panjang sekuens atensi efektif dengan faktor 16–64 relatif terhadap atensi resolusi penuh, memungkinkan pemodelan ketergantungan global dalam anggaran yang kompatibel dengan server *edge* berakselerasi GPU [16]. Pendekatan terkait mencakup kompresi *token* progresif hierarkis, aproksimasi atensi linier yang mengurangi kompleksitas dari O (N^2) menjadi O (N), dan mekanisme atensi lintas skala yang menghubungkan fitur di berbagai tingkat resolusi tanpa biaya kuadratik penuh.

Mekanisme *channel attention* merupakan kelas ketiga dari strategi efisiensi utama, beroperasi dengan mempelajari cara mengkalibrasi ulang kepentingan masing-masing kanal fitur melalui jaringan *gating* yang ringan. Modul *Efficient Channel Attention* (ECA), yang merepresentasikan bentuk *squeeze-and-excitation* yang disempurnakan dengan mengganti lapisan *gating fully connected* menggunakan konvolusi 1D lokal atas deskriptor kanal, muncul dalam banyak studi yang disertakan sebagai komponen tambahan yang meningkatkan selektivitas representasi dengan *overhead* parameter yang dapat diabaikan. Model DSMC-ECA untuk diagnosis kerusakan *gearbox*, misalnya, mengintegrasikan modul ECA dalam *backbone* konvolusi separabel *depthwise* multi-skala dan menunjukkan peningkatan akurasi diagnostik yang terukur dalam kondisi kebisingan tinggi sambil tetap dapat diterapkan pada kontroler industri tertanam [17].

Knowledge distillation semakin banyak digunakan sebagai strategi efisiensi pasca-pelatihan yang melengkapi kompresi arsitektural. Dalam paradigma ini, model *student* hibrida yang kompak dilatih untuk meniru *output* probabilitas lunak atau representasi fitur menengah dari model *teacher* yang lebih kuat, sehingga secara efektif mentransfer pengetahuan yang dipelajari *teacher* ke dalam anggaran parameter yang lebih kecil. Beberapa arsitektur yang ditinjau, termasuk SciceVPR untuk pengenalan tempat visual dan CapMatch untuk pembelajaran kontrastif semi-terawasi, menggunakan pelatihan berbantuan distilasi

P-ISSN : 2089-3353
E-ISSN : 2808-9162

untuk menutup kesenjangan akurasi yang seharusnya muncul dari kompresi arsitektural yang agresif [18]. Secara kolektif, strategi-strategi ini mengilustrasikan bahwa mencapai kelayakan *edge* jarang merupakan hasil dari satu teknik tunggal, melainkan upaya komposit yang mencakup primitif arsitektural, desain atensi, dan metodologi pelatihan.

3.3. Pertukaran Kinerja antara Akurasi dan Latensi pada Perangkat *Edge*

Pemeriksaan keseimbangan antara efisiensi pemrosesan dan ketepatan prediksi dalam jaringan hibrida berbasis *edge* membutuhkan analisis mendalam terhadap metrik kinerja empiris pada berbagai konfigurasi perangkat keras. Tabel 2 menyajikan ringkasan komparatif arsitektur hibrida representatif dan padanannya berupa CNN murni atau *Transformer* murni, yang diturunkan dari data yang dilaporkan dalam studi yang disertakan. Temuan yang konsisten di seluruh literatur yang ditinjau adalah bahwa penalti akurasi yang terkait dengan kendala efisiensi yang agresif secara substansial lebih rendah untuk model hibrida dibandingkan model CNN murni pada anggaran komputasi yang setara, mengisyaratkan bahwa penggabungan mekanisme atensi memberikan bentuk efisiensi parametrik, yaitu mencapai lebih banyak *leverage* representasional per satuan komputasi.

Analisis data latensi mengungkapkan bahwa *overhead* inferensi praktis dari modul atensi sangat bergantung pada kualitas implementasi dan karakteristik perangkat keras. Pada prosesor seri ARM Cortex-A, yaitu kelas perangkat keras *edge* tingkat menengah yang umum, operasi atensi yang diimplementasikan secara naif melalui perkalian matriks padat dapat memperkenalkan penalti latensi sebesar 2–4× relatif terhadap konvolusi yang setara. Namun, implementasi yang dioptimalkan menggunakan kuantisasi *fixed-point*, fusi operator, dan instruksi SIMD khusus perangkat keras secara substansial mengurangi kesenjangan ini. Beberapa studi yang disertakan melaporkan waktu inferensi model hibrida dalam rentang 15–45 milidetik per bingkai pada resolusi 224×224 piksel di platform Raspberry Pi 4 atau NVIDIA Jetson Nano, yang berada dalam batas yang diperlukan untuk aplikasi *real-time* pada 20–25 bingkai per detik [19].

Hubungan antara FLOPs dan akurasi bersifat nonlinier, dan arsitektur hibrida mengeksploitasi non-linearitas ini secara menguntungkan. Bukti empiris dari studi yang disertakan menunjukkan bahwa penggabungan sejumlah kecil *attention head* dengan dimensi *key-value* yang terbatas ke dalam *backbone* yang sepenuhnya konvolusional dapat menghasilkan peningkatan akurasi yang tidak proporsional relatif terhadap FLOPs yang ditambahkan, terutama untuk tugas-tugas yang membutuhkan integrasi isyarat kontekstual spasial yang jauh, seperti segmentasi polip pada pencitraan gastrointestinal, di mana atensi atas konteks seluruh bingkai membedakan struktur yang relevan secara klinis dari jaringan latar belakang yang membingungkan [20]. Sebaliknya, untuk tugas-tugas yang secara inheren bersifat lokal, seperti deteksi cacat berbasis tekstur pada permukaan industri, manfaat inkremental dari mekanisme atensi lebih terbatas, dan CNN ringan murni mungkin merupakan pilihan yang lebih parsimoni.

Kuantisasi model, khususnya kuantisasi pasca-pelatihan dari representasi *floating-point* 32-bit ke representasi *integer* 8-bit, secara konsisten mengurangi latensi inferensi sebesar 2–4× dan jejak memori sebesar sekitar 4× di seluruh platform perangkat keras dengan akselerasi INT8 *native*, seperti Google Coral Edge TPU dan Qualcomm Snapdragon DSP. Beberapa arsitektur hibrida yang ditinjau dirancang khusus dengan mempertimbangkan kompatibilitas kuantisasi, menghindari operasi dengan sensitivitas tinggi terhadap presisi numerik seperti *layer normalization* dengan ukuran *batch* kecil, dan lebih memilih *batch normalization* dengan representasi mode inferensi yang terfusi. Arsitektur IFD-YOLO untuk deteksi target UAV inframerah mengilustrasikan pendekatan ini, menggabungkan *backbone* RepViT, yang merupakan hibrida CNN-*Transformer* yang dapat dikonfigurasi ulang, dengan modul DyGhost dan mencapai kinerja deteksi yang kuat dengan kehilangan akurasi kuantisasi yang minimal dalam penerapan INT8 pada perangkat keras inferensi tertanam [21].

Tabel 2. Perbandingan Metrik Kinerja Arsitektur Hibrida dengan *Baseline* CNN Murni dan Transformer

Arsitektur	Jenis	Akurasi	Parameter	FLOPs	Latensi	Platform	Domain
M2ANET [22]	Hibrida CNN-Attn	95.4 %	4.2 M	0.89 G	~22 ms	Mobile GPU	Pencitraan Medis
EFFResNet-ViT [23]	Hibrida CNN-ViT	96.1 %	28.4 M	6.7 G	~48 ms	NVIDIA Jetson Nano	Pencitraan Medis
Hybird RVIT-CNN [24]	Hibrida ViT-CNN	94.7 %	12.1 M	2.1 G	~35 ms	Edge GPU	Dermatologi
DSMC-ECA [25]	Hibrida DSC-CA	98.2 %	0.31 M	0.07 G	~8 ms	Embedded MCU	Diagnosis Kerusakan
IFD-YOLO [26]	Hibrida RepViT-YOLO	92.3 % mAP	3.8 M	5.1 G	~18 ms	Jetson Edge	Deteksi UAV
STMC-UNet [27]	Hibrida Trans-CNN	91.6 % IoU	21.7 M	38.4 G	~65 ms	Edge Server GPU	Pengindera an Jauh

HAFormer [28]	Hibrida CNN-Attn	79.4% mIoU	5.7 M	8.9 G	~31 ms	NVIDIA Jetson Xavier	Segmentasi Semantik
DLC-SSD [29]	CNN Ringan	85.7% AP	9.3 M	12.6 G	~39 ms	Embedded GPU	Deteksi Objek
MobileNetV2 (baseline)	CNN Murni	92.0%	3.4 M	0.30 G	~12 ms	ARM Cortex-A	Penglihatan Umum
EfficientNet-B0 (baseline)	CNN Murni	93.3%	5.3 M	0.39 G	~18 ms	ARM Cortex-A	Penglihatan Umum
ViT-Tiny (baseline)	Transformer Murni	72.2%	5.7 M	1.3 G	~55 ms	ARM Cortex-A	Penglihatan Umum

Catatan tabel: Baris hibrida diarsir hijau; baris *baseline* murni diarsir oranye. Latensi diukur pada *batch size* = 1. mAP dilaporkan pada IoU=0,5. Data disintesis dari studi yang disertakan.

3.4. Domain Aplikasi dan Konteks Penerapan Ekstraktor Fitur Hibrida

Seratus dua puluh satu studi yang disertakan secara kolektif menunjukkan konsentrasi aplikasi ekstraksi fitur lokal hibrida di beberapa domain yang berdampak tinggi. Analisis citra kesehatan merupakan domain yang paling banyak terwakili, menyumbang 31% makalah yang disertakan. Dalam kategori ini, tugas-tugas mencakup segmentasi tumor dalam MRI multi-parametrik, deteksi polip dalam citra endoskopi, analisis *fundus* untuk skrining glaukoma [32], klasifikasi parasit malaria dalam apusan darah, dan klasifikasi lesi dermatologis. Daya tarik yang berulang dari arsitektur hibrida dalam pencitraan medis mencerminkan tuntutan ganda bidang ini, yaitu presisi spasial tingkat piksel yang membutuhkan resolusi fitur lokal yang halus, dan koherensi semantik di seluruh struktur yang bervariasi secara anatomis yang hanya dapat diberikan secara andal oleh konteks global berbasis atensi [22].

Diagnosis kerusakan industri dan pemantauan kondisi merupakan domain aplikasi terbesar kedua pada 18% studi yang disertakan. Analisis vibrasi mesin berputar, deteksi cacat *gearbox*, klasifikasi *partial discharge* transformator, dan pemantauan kondisi perkakas masing-masing menghadirkan tantangan analisis sinyal yang khas: pola temporal multi-skala yang tertanam dalam aliran sensor berfrekuensi tinggi dan berisik harus diklasifikasikan secara andal dalam kendala komputasi dan *real-time* yang ketat yang diberlakukan oleh *programmable logic controller* industri atau monitor lapangan tertanam. Arsitektur hibrida yang mengintegrasikan *depthwise separable convolution* multi-skala dengan atensi yang efisien sangat sesuai untuk domain ini, sebagaimana didemonstrasikan oleh model DSMC-ECA untuk diagnosis *gearbox* [17] dan kerangka fusi multi-kanal untuk analisis vibrasi transformator tipe kering [23].

Penginderaan jauh dan pencitraan udara menyumbang 14% studi yang disertakan, didorong oleh semakin banyaknya penerapan sistem inferensi berkemampuan *edge* di atas kendaraan udara tak berawak dan stasiun darat satelit. Karakteristik inheren dari penginderaan jauh, yaitu adegan yang sangat luas secara geografis dengan target yang jarang namun beragam secara struktural, membuat kombinasi deskripsi tekstur lokal dan atensi spasial global sangat berharga. Kerangka deteksi objek seperti FE-YOLOv5 dan IFD-YOLO, dan arsitektur segmentasi semantik seperti UNetFormer dan STMC-UNet, secara kolektif menunjukkan penerapan ekstraktor fitur hibrida untuk ekstraksi jejak bangunan, deteksi objek kecil, dan klasifikasi tutupan lahan dalam kendala penerapan udara [24].

Sistem otonom dan transportasi cerdas mewakili 11% studi yang disertakan, mencakup identifikasi ulang pejalan kaki, persepsi kendaraan kooperatif menggunakan fusi LiDAR-kamera, dan pengenalan tempat visual untuk lokalisasi. Arsitektur HYDRO-3D untuk deteksi dan pelacakan objek 3D kooperatif mencontohkan penerapan ekstraksi fitur hibrida pada penginderaan multi-agen, di mana fitur tingkat *voxel* lokal dari sensor kendaraan individu diproses melalui *pipeline* hibrida dan dibagikan ke seluruh jaringan kendaraan untuk pemahaman adegan secara kolaboratif [11]. Domain tambahan yang terwakili dalam literatur yang disertakan mencakup deteksi intrusi jaringan untuk keamanan IoT, pemantauan lingkungan, identifikasi hama tanaman pertanian, dan pengenalan emosi suara, yang secara kolektif menggarisbawahi luasnya konteks di mana ekstraksi fitur lokal hibrida yang diterapkan di *edge* menghasilkan manfaat operasional.

3.5. Integrasi Federated Learning untuk Privasi dan *Edge* Terdistribusi

Federated Learning (FL) berfungsi sebagai metodologi fundamental yang memungkinkan perangkat *edge* terdistribusi untuk secara bersama mengoptimalkan model yang terpadu sambil menjaga data rahasia tetap terlokalisasi. Akibatnya, konvergensi FL dengan kerangka ekstraksi fitur hibrida telah menjadi arah penelitian yang sangat menonjol dalam studi yang diperiksa. Kompatibilitas inti antara FL dan model hibrida yang diterapkan di *edge* terletak pada sifat komplementer tingkat sistem mereka: arsitektur hibrida mengurangi beban komputasi lokal dari inferensi dan volume data yang memerlukan penyimpanan lokal, sementara FL menghilangkan kebutuhan transmisi data ke server pusat dengan hanya berbagi gradien model atau pembaruan model yang terkompresi [25].

Namun, sifat heterogen dari penerapan *edge* IoT menghadirkan tantangan teknis yang signifikan untuk pelatihan federasi ekstraktor fitur hibrida. Heterogenitas data, khususnya sampel *non-*

P-ISSN : 2089-3353
E-ISSN : 2808-9162

independent and identically distributed (Non-IID) di berbagai *node*, yang sering disebabkan oleh operasi lokal yang unik, demografi regional yang spesifik dalam layanan kesehatan, atau lingkungan spasial yang berbeda, menyebabkan divergensi gradien klien. Fenomena ini pada akhirnya menghambat konvergensi model global dan mengakibatkan kinerja suboptimal bagi kelompok data yang kurang terwakili [26]. Beberapa studi yang ditinjau mengatasi tantangan ini melalui strategi optimisasi federasi berbasis kurikulum yang memodulasi kontribusi pembaruan setiap klien berdasarkan kualitas data lokal dan pergeseran distribusi yang diperkirakan, sebagaimana didemonstrasikan dalam kerangka *federated learning* kurikulum berbasis memori untuk klasifikasi kanker payudara [27].

Integrasi mekanisme perlindungan privasi di luar berbagi gradien juga mendapat perhatian dalam literatur yang disertakan. *Differential privacy*, yang menyuntikkan *noise* yang terkalibrasi ke dalam pembaruan gradien untuk memberikan jaminan privasi formal terhadap serangan inferensi keanggotaan, digunakan dalam beberapa *pipeline* pelatihan model hibrida federasi, meskipun dengan biaya berkurangnya utilitas model. Agregasi berbasis enkripsi homomorfik, yang memungkinkan *aggregator* pusat menggabungkan pembaruan gradien terenkripsi tanpa mendekripsinya, memberikan jaminan privasi yang lebih kuat tetapi dengan *overhead* komputasi yang jauh lebih tinggi, sehingga membatasi penerapan praktisnya pada skenario di mana server agregasi tidak dapat dipercaya bahkan untuk inspeksi gradien [28].

Arah yang sangat menjanjikan yang diidentifikasi dalam literatur yang ditinjau adalah penggabungan agregasi federasi berbasis graf dengan ekstraktor fitur hibrida yang beroperasi pada topologi jaringan *edge* terstruktur. Kerangka GFL-SAR (*Graph Federated Collaborative Learning with Structural Amplitude*) memodelkan topologi jaringan IoT sebagai graf dan menggunakan pengiriman pesan konvolusi graf antar *node* tetangga untuk menyebarkan representasi fitur, memungkinkan penyempurnaan model secara kooperatif tanpa merutekan semua gradien melalui server parameter pusat [29]. Dalam domain medis, implementasi yang sebanding terlihat dalam sistem IoT federasi yang berpusat pada privasi dan dirancang untuk memprediksi risiko *stroke*. Pendekatan ini mengintegrasikan klasifikator *machine learning* yang dioptimalkan untuk fitur-fitur tertentu dalam kerangka IoT terdesentralisasi, memastikan bahwa rekam medis pasien yang sensitif tetap disimpan dengan aman di perangkat klinis lokal sambil secara bersamaan memfasilitasi evaluasi kesehatan yang berkelanjutan [30]. Karya-karya ini secara kolektif menetapkan bahwa FL dan ekstraksi fitur hibrida berbasis *edge* bukan hanya kompatibel, tetapi saling memperkuat tujuan efisiensi, privasi, dan adaptabilitas terdistribusi masing-masing secara sinergis.

4. Kesimpulan

Dengan menggabungkan temuan dari 121 publikasi terpilih, tinjauan sistematis ini menyimpulkan bahwa konvergensi antara ekstraksi fitur lokal berbasis CNN dan penalaran global berbasis atensi merepresentasikan paradigma arsitektural yang paling efektif untuk ekstraksi fitur lokal pada *edge computing* yang dibatasi sumber daya. Model hibrida menempati wilayah Pareto-superior di sepanjang batas akurasi-efisiensi relatif terhadap *baseline* CNN murni maupun *Transformer* murni, dengan rata-rata peningkatan akurasi sekitar 3,2 poin persentase tanpa penalti latensi yang proporsional. *Depthwise separable convolution*, agregasi *token*, *efficient channel attention*, dan *knowledge distillation* merupakan strategi efisiensi dominan yang memungkinkan penerapan arsitektur hibrida secara *real-time* pada perangkat *edge*, dengan analisis citra kesehatan, diagnosis kerusakan industri, penginderaan jauh, dan sistem otonom sebagai domain aplikasi utama, serta integrasi *federated learning* yang mengatasi tantangan privasi pada pembelajaran terdistribusi di seluruh *node* IoT heterogen.

Meskipun demikian, karakterisasi kinerja pada perangkat keras heterogen yang sesungguhnya, fusi sensor multi-modal di tingkat ekstraksi fitur, serta pelaporan konsumsi energi dan ketahanan terhadap pergeseran distribusi masih terbatas dalam literatur yang ditinjau, sehingga menjadi arah penting bagi penelitian selanjutnya untuk memperkuat generalisabilitas dan kesiapan penerapan arsitektur hibrida di lingkungan *edge computing* dunia nyata.

Daftar Rujukan

[1] J. Lin, "CKD-TransBTS: Clinical Knowledge-Driven Hybrid Transformer With Modality-Correlated Cross-Attention for Brain Tumor Segmentation," *IEEE Trans. Med. Imaging*, vol. 42, no. 8, pp. 2451–2461, 2023, doi: 10.1109/TMI.2023.3250474.

[2] D. Javeed, M. S. Saeed, M. Adil, P. Kumar, and A. Jolfaei, "A federated learning-based zero trust intrusion detection system for Internet of Things," *Ad Hoc Netw.*, vol. 162, p. 103540, 2024, doi: 10.1016/j.adhoc.2024.103540.

[3] J. Yang, X. Chen, H. Zou, D. Wang, Q. Xu, and L. Xie, "EfficientFi: Toward Large-Scale Lightweight WiFi Sensing via CSI Compression," *IEEE Internet Things J.*, vol. 9, no. 15, pp. 13086–13095, 2022, doi: 10.1109/JIOT.2021.3139958.

[4] F. Ma, B. Sun, and S. Li, "Facial Expression Recognition With Visual Transformers and Attentional Selective Fusion," *IEEE Trans. Affect. Comput.*, vol. 14, no. 2, pp. 1236–1248, 2023, doi: 10.1109/TAFFC.2021.3122146.

[5] Z. Su, "Lightweight Pixel Difference Networks for Efficient Visual Representation Learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 12, pp. 14956–14974, 2023, doi: 10.1109/TPAMI.2023.3300513.

[6] G. Xu, W. Jia, T. Wu, L. Chen, and G. Gao, "HAFormer: Unleashing the Power of Hierarchy-Aware Features for Lightweight Semantic Segmentation," *IEEE Trans. Image Process.*, vol. 33, pp. 4202–4214, 2024, doi: 10.1109/TIP.2024.3425048.

[7] L. Wang, "UNetFormer: A UNet-like transformer for efficient semantic segmentation of remote sensing urban scene imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 190, pp. 196–214, 2022, doi: 10.1016/j.isprsjprs.2022.06.008.

P-ISSN : 2089-3353
E-ISSN : 2808-9162

[8] M. Hassam, M. A. Khan, A. Armghan, S. A. Althubiti, and M. Alhaisoni, "A single stream modified MobileNet V2 and whale controlled entropy based optimization for skin lesion detection and classification," *IEEE Access*, vol. 10, pp. 91496–91508, 2022, doi: 10.1109/ACCESS.2022.3201338.

[9] H. Tan, X. Liu, B. Yin, and X. Li, "MHSA-Net: Multihead Self-Attention Network for Occluded Person Re-Identification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 11, pp. 8210–8224, 2023, doi: 10.1109/TNNLS.2022.3144163.

[10] T. Hussain, "EFFResNet-ViT: A fusion-based convolutional and vision transformer model for explainable medical image diagnosis," *IEEE Access*, vol. 13, pp. 50312–50330, 2025, doi: 10.1109/ACCESS.2025.3554184.

[11] Z. Meng, X. Xia, R. Xu, W. Liu, and J. Ma, "HYDRO-3D: Hybrid Object Detection and Tracking for Cooperative Perception Using 3D LiDAR," *IEEE Trans. Intell. Veh.*, vol. 8, no. 8, pp. 4069–4080, 2023, doi: 10.1109/TIV.2023.3282567.

[12] X.-H. Nguyen, X.-D. Nguyen, H.-H. Huynh, and K.-H. Le, "Realguard: A Lightweight Network Intrusion Detection System for IoT Gateways," *Sensors*, vol. 22, no. 2, p. 432, 2022, doi: 10.3390/s22020432.

[13] X. Zhang, L. Han, W. Zhu, L. Sun, and D. Zhang, "An explainable 3D residual self-attention deep neural network for joint atrophy localization and Alzheimer's disease diagnosis," *IEEE J. Biomed. Health Inform.*, vol. 26, no. 11, pp. 5394–5405, 2022, doi: 10.1109/JBHI.2021.3066832.

[14] S. Ahmed, P. Abdalqadir, S. Abdullah, and Y. Haruna, "M2ANET: Mobile Malaria Attention Network for efficient classification of plasmodium parasites in blood cells," *Intel. Artif.*, vol. 28, pp. 186–199, 2025, doi: 10.4114/intartif.vol28iss76pp186-199.

[15] Y. Kong, S. Lan, S. Lu, Z. Yang, and N. Huang, "HybirdRVIT-CNN: An improved vision transformer model for the recognition of onychomycosis in dermatoscopic images," presented at the Proceedings of the International Conference on Big Data and Artificial Intelligence Applications (BDAIA 2025), 2026.

[16] K. Zhou, C. Liu, Y. Zhang, R. Miao, S. Fu, and Y. Wang, "Enhanced Building Extraction via STMC-UNet: Integrating Super Token Transformer and Multi-scale Convolution," *Signal Image Video Process.*, vol. 19, no. 12, p. 1019, 2025, doi: 10.1007/s11760-025-04575-w.

[17] X. Liu, W. Li, H. Li, Y. Zhu, and R. K. Agarwal, "Lightweight Gearbox Fault Diagnosis Under High Noise Based on Improved Multi-Scale Depthwise Separable Convolution and Efficient Channel Attention," *Sensors*, vol. 26, no. 4, p. 1196, Feb. 2026, doi: 10.3390/s26041196.

[18] S. Wan, Y. Wei, L. Kang, T. Shen, H. Wang, and Y.-H. Yang, "SciceVPR: Stable cross-image correlation enhanced model for visual place recognition," *Neurocomputing*, vol. 669, p. 132539, Mar. 2026, doi: 10.1016/j.neucom.2025.132539.

[19] S. Park, J. Kuai, H. Kim, H. Ko, C. Jung, and Y. Son, "A Lightweight Degradation-Aware Framework for Robust Object Detection in Adverse Weather," *Electronics*, vol. 15, no. 1, p. 146, Dec. 2025, doi: 10.3390/electronics15010146.

[20] J. Zhou, "HCLR-Net: Hybrid Contrastive Learning Regularization with Locally Randomized Perturbation for Underwater Image Enhancement," *Int. J. Comput. Vis.*, vol. 132, no. 10, pp. 4132–4156, 2024, doi: 10.1007/s11263-024-01987-y.

[21] F. Li, X. Lv, M. Zhao, and W. Wu, "IFD-YOLO: A lightweight infrared sensor-based detector for small UAV targets," *Sensors*, vol. 25, 2025, doi: 10.3390/s25154617.

[22] U. Khatri and G.-R. Kwon, "Diagnosis of Alzheimer's disease via optimized lightweight convolution-attention network," *Comput. Biol. Med.*, vol. 171, p. 108116, 2024, doi: 10.1016/j.compbiomed.2024.108116.

[23] Y. Liu *et al.*, "Fault Diagnosis Method for Excitation Dry-Type Transformer Based on Multi-Channel Vibration Signal and Visual Feature Fusion," *Sensors*, vol. 25, no. 24, p. 7460, Dec. 2025, doi: 10.3390/s25247460.

[24] S. Zhao, Y. Luo, T. Zhang, W. Guo, and Z. Zhang, "A domain specific knowledge extraction transformer method for multisource satellite remote sensing imagery road extraction," *ISPRS J. Photogramm. Remote Sens.*, vol. 198, pp. 341–352, 2023, doi: 10.1016/j.isprsjprs.2023.02.011.

[25] A. Jiménez-Sánchez, "Memory-aware curriculum federated learning for breast cancer classification," *Comput. Methods Programs Biomed.*, vol. 229, p. 107318, 2023, doi: 10.1016/j.cmpb.2022.107318.

[26] J. Wicaksana, Z. Yan, X. Yang, Y. Liu, L. Fan, and K.-T. Cheng, "Customized federated learning for multi-source decentralized medical image classification," *IEEE J. Biomed. Health Inform.*, vol. 27, no. 1, pp. 487–498, 2023, doi: 10.1109/JBHI.2022.3198440.

[27] R. Kumar, "Blockchain and homomorphic encryption based privacy-preserving model aggregation for medical images," *Comput. Med. Imaging Graph.*, vol. 102, p. 102139, 2022, doi: 10.1016/j.compmedimag.2022.102139.

[28] J. Guo, I. W.-H. Ho, Y. Hou, and Z. Li, "FedPos: A federated transfer learning framework for CSI-based Wi-Fi indoor positioning," *IEEE Syst. J.*, vol. 17, no. 2, pp. 2490–2501, 2023, doi: 10.1109/JSYST.2022.3230425.

[29] H. Wang, R. Gu, J. Wang, X. Zhang, and H. Wei, "GFL-SAR: Graph Federated Collaborative Learning Framework Based on Structural Amplification and Attention Refinement," *Comput. Mater. Contin.*, vol. 86, no. 1, pp. 1–20, 2026, doi: 10.32604/cmc.2025.069251.

[30] Md. W. Rahman *et al.*, "Privacy-Preserving Federated IoT Architecture for Early Stroke Risk Prediction," *Electronics*, vol. 15, no. 1, p. 32, Dec. 2025, doi: 10.3390/electronics15010032.

[31] C. Cahyaningtyas, Mira, C. Gudiato, and M. Sari, "Deteksi Ekspresi Wajah Manusia Menggunakan Metode Convolutional Neural Network," *Jurnal FASILKOM*, vol. 15, no. 1, pp. 138–145, 2025, doi: 10.37859/jf.v15i1.9119.

[32] R. Aldin, "Implementasi Convolutional Neural Network pada Deteksi Penyakit Retina Menggunakan Citra Fundus Mata," *Jurnal FASILKOM*, vol. 15, no. 3, 2025, doi: 10.37859/jf.v15i3.9857.