

THE ROLE OF MODERN ARTIFICIAL INTELLIGENCE IN DIGITAL IMAGE PROCESSING: A COMPREHENSIVE LITERATURE REVIEW (2021-2026)

¹Pir Amad Ali Shah, ^{*2}Muhammad Sohail, ³Abid Ali,

⁴Syed Rohan Ali Shah

¹Department of Computing and Technology, Sarhad University, Peshawar, 25000, Khyber
Pakhtunkhwa, Pakistan

^{*2}Department of Computing and Technology, Sarhad University, Peshawar, 25000, Khyber
Pakhtunkhwa, Pakistan

³Department of Computing and Technology, Sarhad University, Peshawar, 25000, Khyber
Pakhtunkhwa, Pakistan

⁴Department of Computing and Technology, Sarhad University, Peshawar, 25000, Khyber
Pakhtunkhwa, Pakistan

¹piramadali@gmail.com, ^{*2}sohail.csit@suit.edu.pk, ³abid.ciit@gmail.com

⁴Rohanalisyed@gmail.com

DOI: <https://doi.org/10.5281/zenodo.21805119>

Keywords

Artificial intelligence; Machine learning; digital image processing, deep learning, convolutional neural networks, vision transformers, generative adversarial networks, U-Net, medical imaging, image segmentation, computer vision, lightweight models, model interpretability.

Article History

Received: 03 Jan, 2026

Accepted: 22 Jan, 2026

Published: 25 Jan, 2026

Copyright @Author

Corresponding Author

Muhammad Sohail

Abstract

This review summarizes the recent advances of digital image processing using artificial intelligence methods from 2021 to 2026. It explores the development, performance and applications of convolutional neural networks, Vision Transformers, generative adversarial networks, U-Net variants and hybrid architectures as well as their limitations and future directions. Direct comparison of a few studies that were reviewed to the others is limited with differences in data sets, validation methods, class distributions, and evaluation methods. Several studies reviewed reported accuracy greater than 98% on specific standard data sets, but direct comparison is limited since the data sets, validation methods, class distributions and evaluation methods are different. In addition to healthcare, AI capabilities have been proven to be quite impressive with respect to autonomous driving, remote sensing, industrial quality control, and computer vision. New innovations include the integration of several AI paradigms, attention mechanisms for improved feature extraction, and lightweight models for resource-constrained applications. While there have been great strides, there are still obstacles to overcome in terms of model interpretability, data harmonization, computational efficiency and clinical deployment. It is a review that presents an overview of existing methodologies, comparative performance measures, and emerging trends, offering researchers and practitioners a solid foundation at the intersection of AI and digital image processing.

1. Introduction

The era of modern Artificial Intelligence (AI) and especially Deep Learning (DL) brings a paradigm change in Digital Image Processing. With the development of AI technology, especially in the realm of data-driven methods that automatically learn hierarchical feature representations from raw pixel data, traditional image processing approaches that were dependent upon hand-crafted features and domain-specific algorithms have been replaced. This shift has been driven by a number of factors, such as the availability of large-scale annotated datasets, the increasing power of computational hardware (including GPUs and specialized neural processing units), and breakthrough innovations in neural network AI-enabled image processing has surged forward in this time frame, from 2021 to 2026. Previous decades saw the successful introduction and widespread adoption of the Convolutional Neural Network (CNN), but more complex architectures such as Vision Transformers, attention-based networks, and hybrid networks that blend various architectures have successfully augmented and in some cases replaced CNNs [1], [2], [3]. The improvements have allowed AI systems to handle ever more complex image processing assignments such as medical diagnosis, satellite imagery analysis, industrial quality control, or autonomous navigation, at a level that is often on par with or better than that of a human expert.

The present literature review brings together the results of some peer-reviewed publications from the years 2021 – 2026, covering the use of modern AI methods in different application areas for digital image processing. We review and discuss the development of neural network architectures, the performance metrics of various methodologies, the new trends in the field, and the issues of ongoing research that drive the field forward in this rapidly evolving field [4]-[6].

2. Background and Theoretical Foundations

2.1. Evolution from Traditional to AI-Driven Image Processing

In traditional digital image processing, the functions of edge detection, noise removal and feature extraction were accomplished by using mathematical functions and hand-designed filters. These methods worked well for simple problems,

but were difficult to apply to high-dimensional data and problems with multiple variables. This changed with the advent of deep learning where researchers were able to learn features automatically using hierarchical neural network architectures [2], [7].

2.2. Deep Learning Paradigms

There are a number of essential learning paradigms that underpin modern AI-driven image processing. The supervised learning approach is the most popular, which involves training a model from a set of image-annotation pairs to classify, segment or detect patterns in images [2], [8]. Clustering and self-organizing maps are the techniques of unsupervised learning, which allow for discovering patterns that are not labeled [19]. Semi-supervised learning has become popular in medical imaging problems in which unlabeled images are readily available while labeled images are limited in number [2], [4]. Self-supervised learning and contrastive learning have recently become strong methods for learning good feature representations from unlabeled data [5].

2.3. Fundamental Neural Network Architectures

In the majority of image processing systems, Convolutional Neural Networks (CNNs) are used, with the convolutional layers used to learn spatial hierarchies and local patterns [2], [8], [19]. Modern models are still inspired by landmark architectures like VGG, ResNet and Inception, which have introduced the idea of skip connections, batch normalization, and multi-scale feature extraction, respectively [11]-[19].

U-Net [2] is an encoder-decoder architecture with skip connections that preserve spatial information and enable precise pixel-level localization. It is widely used in biomedical image segmentation. Recent variants, including Swin-UNet [3] and dual-path adversarial U-Net [13], integrate transformer blocks, dense feature fusion, and adversarial learning to improve global-context modelling and segmentation performance.

The paradigm shift that we are introducing is Vision Transformers (ViTs), which are based on using self-attention mechanisms, first introduced to NLP, for images [12], [17]. Traditional CNNs have difficulty with long-range dependencies, while transformers, which model images as sequences of

patches, can better model long-range dependencies but may require more computational resources [5], [29].

Generative Adversarial Networks (GANs), which are composed of generator and discriminator networks that are trained in an adversarial manner,

have been used in image synthesis, data augmentation and domain adaptation [4], [5], [25]. Self-attention GANs and their variants have been shown to have extraordinary skill at creating photo-realistic images [25].

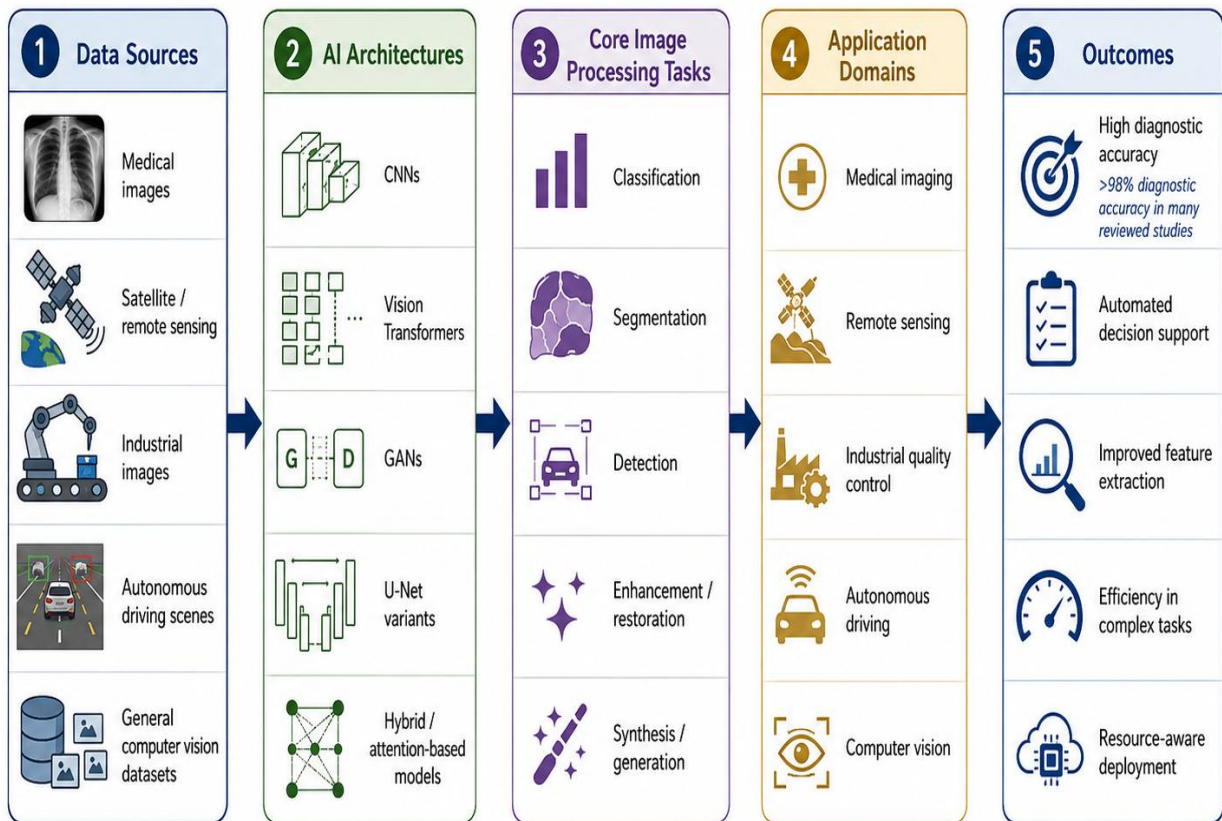


Figure 1 Conceptual framework of modern AI in Digital Image Processing

3. Core AI Architectures and Techniques

3.1. Convolutional Neural Networks and Variants

CNNs are still the workhorse of image processing application because of their parameter efficiency and translation invariance properties. In recent years, emphasis has been placed on enhancements to depth, width and computational speed. Very deep architectures (100+ layers) can be trained without degradation using deep residual networks with skip connections [19]. However, depth wise separable convolutions decrease the computational complexity without compromising representational power, allowing CNN to be implemented on mobile and embedded devices [16].

Specialized CNN architectures have proven to be beneficial for medical imaging applications. Researchers have found that the VGG-19 and

InceptionV3 models outperform others in the field of brain tumor detection when fine-tuned through transfer learning, achieving 100% validation accuracy in certain datasets [11]. When the classification of Alzheimer's disease is conducted based on MRI images, shallow CNN has achieved around 98% accuracy, much better than Support Vector Machines and Naive Bayes, which are traditional machine learning methods [20].

3.2. U-Net and Encoder-Decoder Architectures

Because of their capability to supply fine-grained predictions at pixel level and maintain spatial context, U-Net architectures have emerged as the cornerstone for medical image segmentation tasks [2], [4]. The symmetric encoder-decoder architecture with skip connections facilitates the

flow of information from high-resolution input layers to the output map of segmentation.

UNet has been improved in recent years with the introduction of transformer components. Swin-UNet is a U-Net based structure with Swin Transformer blocks, which allows for capturing local and global context information [3]. Recently, the accuracy for biomedical image classification using a hybrid model based on U-Net and Gated Recurrent Units (GRUs) has reached 94.51% on ISIC 2017 and 95.38% on ISIC 2020 [3]. The UNet model is enhanced with multiple modules of receptive fields and dense feature fusion, which have been shown to achieve better noise removal and detail preservation in benchmark datasets [13] for dual-path adversarial deep denoising networks.

3.3. Vision Transformers and Attention Mechanisms

Vision Transformers have been shown to be strong alternatives to CNNs, especially in the context of tasks that need to make use of the global context. ViTs are capable of capturing long-range dependence between features by dividing images into patches and applying self-attention layers, which outperforms convolutions with small receptive fields [5], [12], [17].

Multi-manifold attention mechanisms are a generalization of the standard self-attention mechanism, and they aim at modeling input spaces defined over multiple manifolds (Euclidean, Symmetric Positive Definite, and Grassmann) whose statistics and geometry are different to generate more descriptive attention maps [17]. It helps the Vision Transformers to pay attention to critical appearance, color, and texture information for better classification and segmentation performances.

A hybrid of Vision Transformers and CNNs has been particularly promising. In one study, the use of Vision Transformers enabled fast extraction of high level features with kernel-based CNNs capturing very detailed patterns, and Multi-Class Support Vector Machines were used as a final layer to improve the classification [12]. Adapted from the Segment Anything Model for medical imaging, the SAM-Med2D model consists of transformer-based encoders, attention modules, and demonstrates state-of-the-art performance in

segmentation and classification across medical imaging modalities [10].

3.4. Generative Adversarial Networks

The GANs have transformed the image synthesis, enhancement, and data augmentation problem. The adversarial training paradigm, which trains a generator network to generate realistic images and a discriminator to classify them into real or fake images, has shown outstanding success in a variety of applications [4, 5, 25].

Self-attention GANs add the attention mechanism to the generator and the discriminator, which allow for focusing on relevant parts of images and learning long-range dependencies [25]. Self-attention GANs have also successfully applied to synthesize training images, which help to enhance the performance of the classifier, highlighting the usefulness of GANs as tools for data augmentation for art style classification [25].

With the rise of contrastive learning and Transformer architectures, the CTNGAN framework combines the three to achieve robust medical image analysis [5]. The tripartite integration combines GAN-based augmentation, contrastive learning to boost feature robustness, and transformer to model long-range spatial patterns to tackle data scarcity and class imbalance. It successfully got a maximum accuracy of 98.5% and an F1 score of 0.968 on various benchmark medical imaging datasets [5].

3.5. Hybrid and Multi-Modal Architectures

The trend in modern image processing is the use of hybrid architectures incorporating several AI paradigms. These strategies combine the best attributes of various neural network architectures leading to better performance [1], [3], [6] and [11].

An interesting hybrid is the use of CNNs with Recurrent Neural Networks (RNNs) in brain tumor analysis, where CNNs are used for extracting spatial features, while RNNs are used to process sequential data [6]. This integration is much more accurate at demarcating the tumor and classifies the images far more accurately than do models built in a single architecture [6].

The other hybrid one combines the traditional image processing methods with deep learning processes. In this study, the image enhancement technique Contrast-Limited Adaptive Histogram

Equalization (CLAHE) was integrated with CNN, GAN, and Transfer Learning with pre-trained Inception V3 models to improve the accuracy of liver tumor detection, with an accuracy of 93.3%, sensitivity of 92.2%, specificity of 94.5%, and F1-

score of 92.8% [8]. The hybrid approach was found to be significantly superior to the standalone deep learning techniques, highlighting the relevance of classical image processing techniques when combined with modern AI techniques [8], [11].

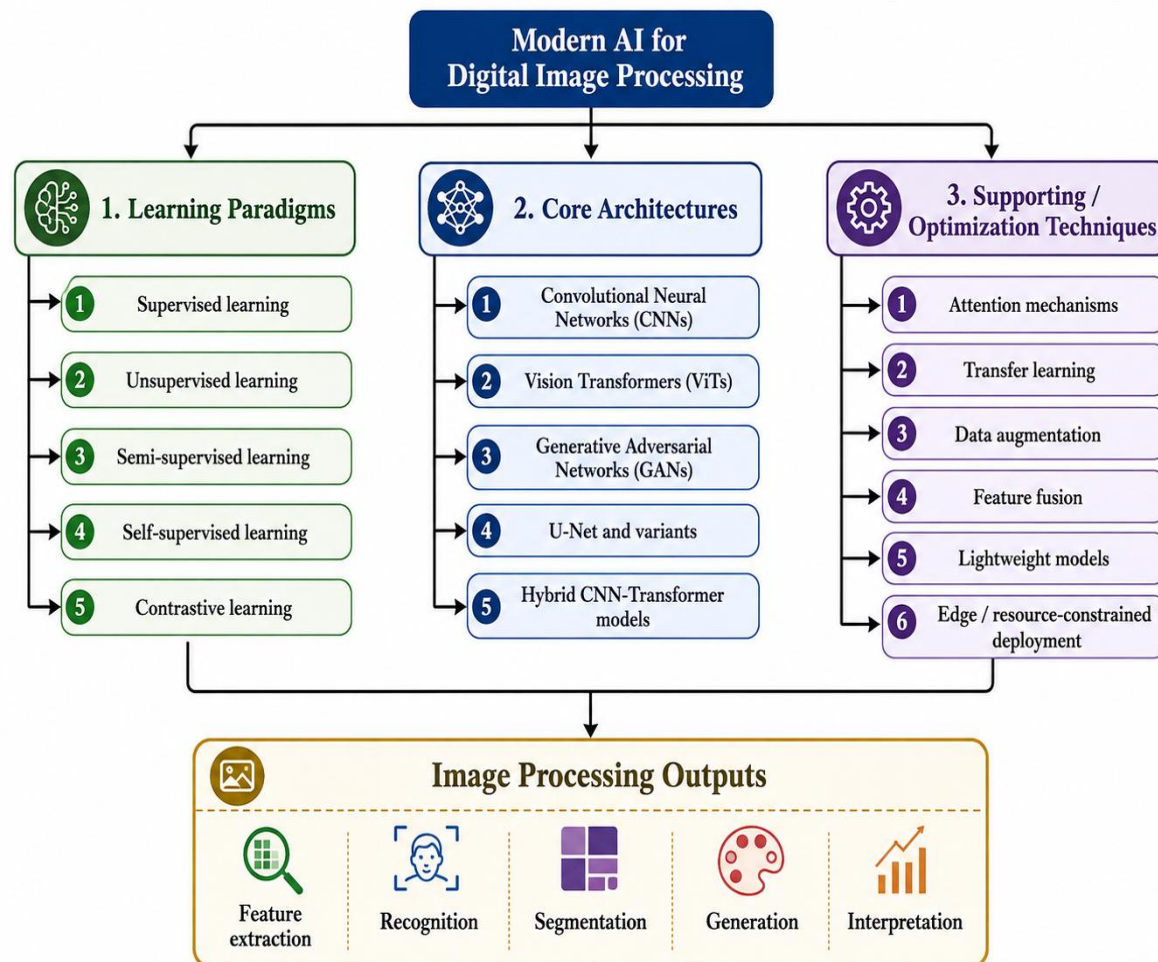


Figure 2 Taxonomy of AI methods in Digital Image Processing

4. Application Domains and Use Cases

4.1. Medical Imaging and Healthcare

Medical imaging is the most widely explored application domain for AI-based image processing, ranging from diagnosis to treatment planning and monitoring of disease across various imaging modalities [1], [2], [4], [5], [6], [7], [8], [10], [11], [12], [20], [26], [28], [29].

4.1.1. Disease Detection and Classification

AI systems have been found to be of excellent performance in identifying and categorizing diseases in medical images. In the diagnosis of Alzheimer's disease, the deep CNN model based

on MRI image successfully classified the patient into a non-demented, very mild demented, mild demented, and moderate demented category with the accuracy rate of 98% and a training rate of 0.9990, a validation rate of 0.9756 [20].

Advanced CNN architectures have transformed malaria blood smear image detection, achieving high accuracy up to 99.68% in the classification of malaria-infected and uninfected samples with high sensitivity and specificity in one study [22]. The performance is particularly interesting in resource-limited areas where experienced microscopists might not be available, and it is for this reason that

AI-based malaria diagnosis is very promising in such areas [22].

In the case of traumatic brain injury (TBI), deep learning-based algorithms like CNN and transformers have been used to identify and classify acute imaging abnormalities including intracranial hemorrhage, skull fracture, intracranial mass effect, and stroke [29]. These algorithms can reorder the radiological reading queue based on the suspicion of critical findings, and may help shorten the interpretation time, thus possibly enhancing patient outcome [29].

4.1.2. Medical Image Segmentation

The precise delineation of anatomical structures and pathological regions (segmentation) is a fundamental aspect of medical image analysis. U-Net architectures and its variants are the most widely used method in this field [2], [4], [10], [12]. Brain tumor segmentation has received much research, and hybrid deep learning models, such as hybrid CNN – RNN models, have been shown to outperform in detecting complex tumor boundaries [6]. We introduced the SAM-AutoMed model, which is an upgraded version of SAM-Med2D, by using an enhanced MobileNet v3 backbone in place of the prompt encoder, to get the state-of-the-art automatic segmentation results on a variety of medical imaging datasets [10].

Hybrid methods combining image enhancement (CLAHE) with deep learning segmentation for the detection of liver tumor have demonstrated high accuracy and precision for the tumor boundary delineation on both MRI and CT images [8]. Combining imaging data with clinical and genomic data (multi-modal) allows for better risk-stratification and personalized recommendations for targeted therapy [1].

4.1.3. Image Enhancement and Super-Resolution

The quality of medical images affects the accuracy of diagnosis. Improvement Strategies using AI overcome spatial resolution, contrast and noise issues. Deep learning-based MRI super-resolution uses CNNs, GANs, transformers, diffusion models and implicit neural representations to improve the spatial resolution, which offers valuable insights for clinical and neuroscientific evaluation [7].

AI-integrated multi-modality fusion imaging provides all-round diagnostic information. Graph

neural networks and hybrid fusion architectures enable the linkage of structural imaging (CT/MR) and functional imaging (PET) data with clinical data, which are used to develop multimodal digital twins to assist personalized care in healthcare [1].

4.2. Autonomous Driving and Intelligent Transportation

The identification of objects, understanding of the scene, and navigation are critical capabilities of autonomous vehicles that depend on real-time image processing. The use of AI-related image enhancement is especially important in low-light situations, which can pose a risk of crashes [23].

CNN architectures have been employed for night vision enhancement for autonomous driving. The LE-net (Light Enhancement Net) model, which is optimized to enhance very low-light road images via transformation of daytime images, achieved better performance than baseline models for enhancing extremely low-light road images [23]. This is a crucial ability for the safe operation of autonomous vehicles in rural areas where there are no street lights [23].

The detection of craters for lunar exploration is a special autonomous navigation problem. A Vision-Transformer-based model, fine-tuned using Low-Rank Adaptation and optimized with both Complete Intersection over Union (CIoU) and contrastive loss functions, was used to detect craters on challenging lunar images with a recall of 94.0% and a precision of 73.1% [24].

4.3. Remote Sensing and Geospatial Analysis

AI's applications in remote sensing involve analyzing satellite and aerial images to glean insights into the Earth's surface [9], [14].

Fast feature extraction networks are helpful for aerial scene categorization, which extract the interactivity between images in different channels and use adaptive strategies to determine the optimal convolutional operations [9]. The CIFNet (Cross-scale Interaction Network) for salient object detection in optical remote sensing images tackles the problems of scale variety and complicated edge details, and achieves inference speed of 58 fps for 40.40M parameters, and improves the performance by 3.45% on MAE and by 1.38% on F-measure compared with the best methods [14].

In remote sensing image, salient object detection process is dealing with various scales, shapes and uncertain positions of objects. Advanced architectures that integrate furcate skip connection modules, dilated convolutions, global leading attention modules and expansion-integration modules can be used to effectively capture multi-scale contextual information and refine features through attention [14].

4.4. Industrial Applications and Quality Control

Industrial image processing systems require high accuracy, real-time operation and operating on the resource limited hardware [21].

For industrial applications, the solution to single-image super-resolution has been proposed by the use of light-weight architectures which trade off image reconstruction quality and computational efficiency. The Progressive Interaction-Learning Network (PILN) uses hierarchical feature-integration in temporal and spatial directions, plus improved interaction-learning units, and pixel wise learning using residual cells to get state-of-the-art results with a computation budget that is appropriate for portable industrial devices [21]. Deep learning models specifically engineered for mobile Neural Processing Units (NPU) have launched a revolution in mobile image signal processing (ISP) for smartphone cameras. For Full HD photos, it has been shown that near real-time processing (60-100 milliseconds) is possible with high fidelity of the image (PSNR of 23.73 and SSIM of 0.8487) for compact architectures such as Small-net [16] and Channel Spatial Attention Networks.

4.5. Computer Vision and General Image Processing

In addition to the applications listed above, recent works have established the state-of-the-art advances for computer vision fundamentals, such as classification [15], object detection [17], [19], [30] and segmentation [19], [30].

Sparse Spatiotemporal Transformers (SSTVOS) [15] use sparse attention operators to attain an overall score of 81.8 on YouTube-VOS 2019 and mean J&F score of 78.4 on DAVIS 2017 with just 10% of the feature matching FLOPs of previous techniques.

A systematic comparison of attention mechanisms in deep neural networks (DNNs) for image classification and object detection tasks has shown that moving the attention to spatial regions (visual attention) and to feature channels (internal attention) can help improve the accuracy [30].

Art style classification shows the versatility of AI beyond just a technical area. Classifier performance was improved by using self-attention GANs to generate synthetic training images, as in EfficientNet B0 where it was seen to be helpful with GAN data augmentation [25].

5. Comparative Analysis of Methods and Performance

5.1. Medical Image Analysis Performance

Medical imaging applications represent the most advanced and well-tested applications of AI. The performance measures from representative studies are summarized in Table 1.

Table 1: *Typical studies performance measures*

Application	Architecture / Model	Dataset	Accuracy (%)	Additional Performance Metrics	Reference
Alzheimer's Classification	Shallow CNN	MRI Patches	98.0	High Precision and Recall across disease stages	[20]
Brain Tumor Detection	InceptionV3 (Transfer Learning)	MRI Scans	100.0	High validation accuracy	[11]
Brain Tumor Detection	VGG-19 (Transfer Learning)	MRI Scans	95.0	High validation accuracy	[11]
Liver Tumor Detection	Hybrid CLAHE + CNN + GAN + Transfer Learning	MRI / CT Images	93.3	Sensitivity: 92.2%, Specificity: 94.5%, F1-score: 92.8%	[8]

Application	Architecture / Model	Dataset	Accuracy (%)	Additional Performance Metrics	Reference
Malaria Detection	Advanced CNN	Blood Smear Images	99.68	High sensitivity and specificity	[22]
Medical Image Classification	CTNGAN (GAN + Contrastive Learning + Transformer)	Benchmark Datasets	98.5	F1-score: 0.968	[5]
Biomedical Classification	HDLB-BTA (Swin-UNet + GRU + HFSA)	ISIC 2017	94.51	Not reported	[3]
Biomedical Classification	HDLB-BTA (Swin-UNet + GRU + HFSA)	ISIC 2020	95.38	Not reported	[3]

The facts show a few points of interest. First, transfer learning from pre-trained models (InceptionV3, VGG-19) has routinely been found to be very successful, and some research has reported achieving 100% validation accuracy [11]. Second, hybrid methods often outperform single paradigm methods when using a combination of AI methods or when using traditional image processing and deep learning methods [3, 8, 11]. Third, specialized architectures that are developed for specific medical imaging applications

(CTNGAN, HDLB-BTA) show balanced performance across various metrics, indicating that they might perform better in other medical imaging applications, i.e., better generalization [3, 5].

5.2. Image Enhancement and Super-Resolution

For the task of image enhancement and super-resolution, perceptual quality metrics are often used to assess tasks instead of classification accuracy:

Table 2: *Typical application with key evaluation metrics*

Application	Architecture / Model	Key Evaluation Metrics	Performance Summary	Reference
MRI Super-Resolution	CNNs, GANs, Transformers, Diffusion Models	PSNR, SSIM	Performance varies by architecture; transformer-based models show superior reconstruction capability	[7]
Infrared Image Enhancement	Cbc-SwinIR + MAXIM + AILI Fusion	FD, Entropy, MSE, PSNR	FD: +0.08, Entropy: +0.094, MSE: +0.00512, PSNR: -12.206	[27]
Mobile ISP Channel Optimization	Spatial Attention Network	PSNR, SSIM, Runtime	PSNR: 23.73, SSIM: 0.8487, Runtime: 60-100 ms for Full HD images	[16]
Image Denoising	Dual-Path Adversarial UNet	Qualitative and Quantitative	Outperforms state-of-the-art GAN-based methods on SIDD, DND, and PolyU datasets	[13]
Night Image Enhancement	LE-Net (CNN)	Qualitative and Quantitative	Superior enhancement under low-light and nighttime conditions	[23]

The deterioration of PSNR values, but improvement of fractal dimension and information entropy in the infrared image enhancement, suggests that the enhanced images may have more useful texture details and complexities for analysis, even if the traditional quality metrics are lower.

This is a reminder of the need for task-specific evaluation criteria.

5.3. Object Detection and Segmentation

For object detection and segmentation, various metrics are used including Intersection over Union (IoU), precision, recall and F-measure.:

Table 3: *Object detection and segmentation with key evaluation metrics*

Application	Architecture Model	/ Dataset	Key Evaluation Metrics	Performance Summary	Reference
Medical Image Segmentation	SAM-AutoMed	Multiple Medical Datasets	Not reported	Outperforms SAM and SAM-Med2D across benchmark evaluations	[10]
Video Segmentation	SSTVOS (Sparse Transformer-based VOS)	YouTube-VOS 2019	Overall Score	81.8	[15]
Video Segmentation	SSTVOS (Sparse Transformer-based VOS)	DAVIS 2017	Mean J&F	78.4	[15]
Remote Sensing Salient Object Detection	CIFNet	Benchmark Datasets	MAE, F-measure	MAE: +3.45% improvement, F-measure: +1.38% improvement	[14]
Crater Detection	OWLv2 (Vision Transformer)	IMPACT Dataset	Recall, Precision	Recall: 94.0%, Precision: 73.1%	[24]

The SSTVOS architecture highlights the fact that they can obtain competitive performance and significantly decrease the computational complexity, including a 10-fold decrease in FLOPs in comparison with dense attention architectures [15]. This is an essential efficiency/performance compromise for real-time video processing applications.

5.4. Computational Efficiency and Deployment

With the deployment of AI models on edge devices and mobile platforms, computational efficiency is becoming increasingly critical:

Table 4: *Computational efficiency and deployment application with models*

Application	Architecture Model	/ Model Parameters	Inference Speed	Deployment Platform	Reference
Aerial Scene Categorization	Accelerated Feature Extraction Network	Not reported	43-53% faster than comparable methods	RTX 2080Ti, Titan V, Jetson Edge Devices	[9]
Remote Sensing Object Detection	CIFNet	40.40 Million	58 frames per second (FPS)	Not reported	[14]
Mobile ISP Optimization	SmallNet	21 KB	16 FPS	Mobile NPU	[16]
Mobile ISP Optimization	Channel Spatial Attention Network	Not reported	60-100 ms per Full HD frame	MediaTek Dimensity 1000+	[16]

The Smallnet architecture achieves remarkable efficiency with only 21 KB of parameters while maintaining acceptable image quality, demonstrating that careful architectural design can enable deployment on severely resource-constrained devices [16]. The CIFNet model balances accuracy and speed, achieving real-time performance (58 fps) suitable for operational remote sensing applications [14].

5.5. Comparative Insights

Several key insights emerge from this comparative analysis:

1. Architecture selection is important: Different architectures are good for different jobs. CNNs work very well for classification and detection, U-Nets work very well for segmentation, transformers work very well for picking up the

global context, and GANs work very well for generation and augmentation.

2. Hybrid Approaches demonstrate Potential: Synergistic combination of several AI paradigms (CNN+RNN, CNN+Transformer, GAN+Contrastive Learning) has consistently shown to outperform single architecture solutions [3], [5], [6], [11].

3. Transfer Learning speeds up development: Fine-tuning pre-trained models on domain-specific data yields very good performance with small amounts of labeled data, especially in medical imaging where labeled data is rare [11], [20].

4. Efficiency-Performance trade-off: Mobile and embedded architecture-specific lightweight designs can achieve performance comparable to that of the base design with orders of magnitude fewer parameters and inference times [9], [14], [16].

5. Task-Specific Metrics are necessary: There may be other aspects of the performance that are important to standard metrics such as accuracy. Medical imaging demands for sensitivity and specificity, perceptual quality metrics are needed for enhancement tasks, and real-time applications call for considerations of computational efficiency [8], [13], [27].

6. Emerging Trends and Innovations

6.1. Foundation Models and Large-Scale Pre-training

Foundation models have become a success in natural language processing, and similar techniques are applied to computer vision. The Segment Anything Model (SAM) is a groundbreaking innovation a large visual model pre-trained on a wide range of visual data to be used for various segmentation tasks [10]. This work shows how foundation models can be adapted to perform domain-specific tasks without sacrificing their general ability, e.g., in medical imaging with SAM-Med2D [10] and SAM-AutoMed.

6.2. Multi-Modal and Cross-Modal Learning

The integration of multiple data modalities is one of the next challenges in medical imaging and beyond. Multimodal digital twins incorporating imaging data, clinical data, genomic data, and real-time physiological monitoring data can provide personalized healthcare by facilitating accurate risk stratification and personalized treatment

recommendations [1]. Graph neural networks enable the integration of multiple data modalities by encoding different modal relationships as graph structure [1].

With the introduction of the new approach, vision-language models are becoming increasingly popular for tasks that demand semantic understanding. The use of vision-language models for crater detection illustrates the use of textual descriptions to direct visual analysis [24].

6.3. Self-Supervised and Contrastive Learning

Representations learned from a large amount of unlabeled data are becoming popular using a technique called self-supervised learning, especially in areas where annotations are costly. A comparative model which learns from both positive and negative examples has proven to be very effective in learning powerful feature representations [5]. The CTNGAN framework successfully combined the principles of contrastive learning and generative adversarial networks (GANs) and transformers, yielding up to 98.5% accuracy in medical imaging benchmarks, showcasing the potential of this combination [5].

6.4. Attention Mechanisms and Interpretability

The notion of attention has grown from spatial to multiple dimensions and geometric spaces. Multi-manifold attention is employed in order to learn attention maps for models of different input spaces using Euclidean, Symmetric Positive Definite and Grassmann manifolds, which results in more descriptive attention maps [17]. Such powerful attention mechanisms have been shown to boost the performance of the model and also to increase the interpretability of the model by highlighting the regions of the images and features that contributed to the prediction [17], [30].

6.5. Neural Architecture Search and AutoML

With the advent of automated machine learning (AutoML) techniques such as neural architecture search (NAS), one is discovering optimal architectures for specific tasks and hardware platform. This allows the efficient designs to be generated automatically by considering NPUs/coprocessors of target deployment devices in platform-aware NAS [16]. This shift empowers more people to use AI, mitigating the need for a high level of architecture engineering skills.

6.6. Lightweight and Efficient Architectures

With edge computing and the increasing deployment of mobile devices, an emphasis on lightweight architectures has spurred innovation. Network pruning, quantization (16-bit, 8-bit and below), knowledge distillation and depth wise separable convolutions are some of the techniques used to deploy complex models on limited resource devices [16], [21]. The emergence of models such as Small-net (21 KB) and efficient super-resolution networks has shown that with careful architectural design it is possible to obtain competitive performance with a small amount of computational effort [16], [21].

6.7. Synthetic Data and Data Augmentation

Dealing with data scarcity problems, in the medical imaging domain in particular, synthetic training data generation has become a common task using models like GANs and diffusion models. Self-attention GANs have been applied to successfully generate synthetic art images to improve the style classification [25]. In medical imaging, GANs can be used to produce realistic synthetic data samples, reducing data sparseness and class imbalance [5]. The rise of synthetic data generation might speed up AI innovation in data-constrained fields, and can also provoke debates on how well AI models transfer to real-world data.

7. Challenges and Limitations

7.1. Data Quality and Availability

Although synthetic data generation has made progress, there is still a fundamental limitation of having a limited number of annotated cases, especially in medical imaging where expert annotations are time-consuming and costly [1], [2], [5]. Class imbalance, with the availability of a few examples of one class and a substantially larger number of examples of another class, can lead to the model being optimized for the majority class, and thus to sub-optimal performance on the minority class, which may be clinically relevant [5] and [8].

Harmonization of data between imaging devices, imaging protocols and institutions is a significant challenge. Systematic biases may arise due to different image acquisition parameters, different scanner manufacturers, and different imaging protocols, affecting the performance of the model

when used in a different clinical environment [1], [7]. The model needs to be generalizable through the use of standardized datasets and imaging protocols [1].

7.2. Model Interpretability and Trust

Deep neural networks are "black boxes" that are difficult to use in the clinic and approve for regulation. Clinicians and patients need to trust a model, which means understanding why a model makes certain predictions, and model interpretability is crucial [1], [4]. Although attention mechanisms are useful to understand what information from which region of the image is used when making predictions, they fail to explain the full decision-making process [17], [30]. Evaluation metrics and standardized benchmarking strategies need to be transparent to allow a fair benchmarking between methods, and to allow clinical validation [1]. The absence of standard evaluation procedures makes it hard to determine if the reported performance gains are true gains and not due to the choice of data sets and evaluation method.

7.3. Computational Requirements and Efficiency

Lightweight designs have come a long way, but there are still many state-of-the-art models that are not lightweight. Although the Vision Transformers have achieved better results on many tasks, they are still more expensive in terms of computational resources than CNNs [5], [12], [17]. This computing power requirement restricts mobile deployment and real-time applications.

One of the key issues is the tradeoff between accuracy and computational complexity. The computational demands can be reduced by quantization and pruning, with a corresponding loss in performance [16]. Careful architecture design and hardware-aware optimization are needed to find optimal points on the accuracy-efficiency Pareto frontier.

7.4. Generalization and Domain Shift

Models learned from one domain tend to perform poorly on another domain, which is referred to as domain shift or distribution shift. Medical imaging might be affected by the lack of generalizability from one hospital to another with different equipment and protocols [1], [7]. Some of these difficulties are addressed by transfer learning

and/or domain adaptation techniques, but not fully overcome [11] and [20].

The gap between simulated and real data is still an issue [23], [25] when a model trained on synthetic data or enhanced data is used to model real images. Synthetic data generation can provide additional data in small training sets, but care needs to be taken to validate the model to ensure that its predictions can be transferred from synthetic to real data.

7.5. Clinical Integration and Regulatory Deployment.

Language barriers are only part of the challenge to get research prototypes into clinically deployed systems. Significant engineering efforts are required for integration with existing clinical workflows, electronic health record (EHR) systems and Picture Archiving and Communication Systems (PACS) [1], [2]. The regulatory landscape for medical AI systems is still developing and is complex and lengthy.

Coordinating the work of the AI researchers, clinicians, and healthcare administrators is crucial but can be difficult to achieve [1]. Various priorities, terms and criteria for evaluation can prevent progress between these communities.

7.6. Ethics and Privacy Issues

Medical imaging data is typically sensitive, containing patient information, which presents privacy concerns that need to be addressed using approaches like federated learning, differential privacy, and secure multi-party computation. If not properly monitored and mitigated, algorithmic bias (i.e., models that perform differently by demographic group) can contribute to or amplify healthcare disparities [4].

Artificial data and data augmentation also bring up concerns about the ability of models trained on synthetic data to accurately predict for real patients, especially for rare diseases or "edge cases" that are not well represented in the training data [5], [25].

8. Future Directions and Research Opportunities

8.1. Enhanced Interpretability and Explainable AI

Developing Explainable AI (XAI) techniques that deliver clinically relevant explanations for AI model predictions is a pivotal research pathway. Future work should include methods beyond

attention visualization that offer causal explanations, counterfactual reasoning and uncertainty quantification [1], [4]. The incorporation of domain knowledge via physics-informed neural networks and knowledge graphs could prove to be useful in improving both interpretability and performance.

8.2. Federated Learning and Privacy-Preserving AI.

To address privacy concerns while allowing collaboration among institutions, federated learning methods allow training models from distributed datasets without centralized sensitive patient data [1]. For future work, more efficient federated learning methods, such as handling non-IID data across sites, and secure aggregation protocols to prevent information leakage should be explored.

8.3. Few-Shot and Zero-Shot Learning

Robust models that can learn from a small amount of data or generalize to unseen categories without further training would make data requirements a huge jump. In this direction, SAM is an example of a foundation model, but the development of more powerful "prompt-based models," "meta-learning models," and "transfer learning models" is required [10].

8.4. Real-Time and Edge AI

In the near future, the deployment of advanced AI models on edge devices, for autonomous vehicles, mobile health monitoring, and industrial quality control, will be made possible by continued innovation in efficient architectures and hardware acceleration [9], [14], [16], [21], [23]. Neural architecture search for specific hardware platforms, mixed-precision quantization, and hardware-software co-design are among the directions for research.

8.5. Multi-Modal Foundation Models

Multi-modality foundation models, combining multiple imaging modalities, clinical text, genomic information and physiological signals, are a frontier in medical AI [1]. Such models might allow for more thorough patient evaluation and more individualized treatment plans. Difficulties are related to designing efficient fusion architectures, dealing with missing modalities and maintaining the robustness of the model.

8.6. Continual and Lifelong Learning.

The field of medicine is constantly changing, and imaging technology is advancing as well. The new systems must be able to learn new information, but without catastrophic forgetting past information. As clinical practice and imaging procedure change, research into continual learning, lifelong learning, and online adaptation will be necessary to ensure the model's performance [1], [7].

8.7. Standardization and Benchmarking

The field would benefit from having the benchmarks, evaluation protocols, and reporting guidelines standardized such that they would allow for fair comparison with other methods and clinical validation of the field [1]. Large, diverse, well-annotated data sets with standardized train-test splits and evaluation metrics will further propel the progress and help ensure reproducibility.

8.8. Human-AI Collaboration

Increasingly, studies will need to move beyond the idea of AI taking the place of experts and look to the paradigms of human-AI collaboration [1], [29]. It is noteworthy that interactive segmentation tools, uncertainty quantification for diagnosis using artificial intelligence and decision support systems that provide explanations for their decision making are promising approaches.

9. Conclusion

Over the last six years (2021-2026), there have been significant strides in AI-powered digital image processing, with deep learning architectures reaching new heights of performance in various application areas. The most mature application area is medical imaging, where AI systems have been shown to perform with a diagnostic accuracy of over 98% for various tasks, including Alzheimer's classification and malaria detection. However, modern AI methods have shown impressive abilities in autonomous driving, remote sensing, industrial quality control, and even in basic computer vision work, apart from healthcare. These advances are due to a number of important innovations in the architectural field. CNNs are still the backbone of deep learning, with advancements like deeper models, attention mechanisms, and hybrid architectures. U-Net variants are the current state-of-the-art for segmentation tasks especially combined with

transformer components and adversarial training. While CNNs are widely used for image classification, they are less effective at tasks that need to understand the global context due to slow computation, Vision Transformers can be powerful alternatives at higher computational cost. GANs have transformed the landscape of image synthesis, enhancement, and data augmentation, especially in the face of limited data in medical imaging.

Multi-paradigm architectures, in general, have been shown to achieve better performance than single paradigm approaches. Examples of integration are CNNs with RNNs, transformers with convolutional layers, and GANs with contrastive learning, which are multi-paradigm approaches that take advantage of complementary capacities of various types of neural networks. Transfer learning from very large pre-trained models has been found to be extremely useful, which allows good performance with a small amount of data for a specific domain.

Predictive trends indicate more advanced AI systems in the future. The state-of-the-art in current research includes foundation models that are pre-trained on a wide range of data and then fine-tuned for specific tasks, multi-modal architectures that combine in a coherent way several different types of data, self-supervised learning algorithms that are used to decrease the amount of annotation required, and highly sophisticated attention mechanisms that boost performance and explainability. With an increasing trend towards edge processing, innovative lightweight architectures with competitive performance and very little computational burden are being developed allowing the deployment of edge applications on mobile and embedded devices.

Although there has been considerable improvement, there are still great challenges. Even with the introduction of labeled data, its availability is still restricted, especially in the medical imaging domain. Model interpretability and trust are still important challenges for clinical adoption. Real-time and mobile deployment are constrained by computational demands of cutting-edge architectures. Generalization to other domains and institutions still has challenges, with models frequently failing to yield accurate results

with data from other sources. In addition to technical challenges, the integration with current clinical workflows, regulatory clearance and ethical concerns regarding privacy and algorithmic bias are potential obstacles to clinical deployment.

Future research directions are evident. There are several aspects that warrant further research and development, including improved interpretability techniques using explainable AI methods, privacy-preserving approaches like federated learning, few-shot and zero-shot learning to mitigate data needs, ongoing efficient architectural innovations for edge deployment, multi-modal foundation models consolidating multiple data modalities, continuous learning systems that evolve with new data,

consistent benchmarks and evaluation protocols, and human-AI collaboration paradigms that complement human expertise.

The effect of 'modern' AI on digital image processing is undeniable. The more complex the architecture, the bigger and more varied the datasets, the more complex the deployment issues, the more AI-powered image processing systems will be embedded in the healthcare, autonomous systems, industrial process, and scientific research fields. The future of intelligent image processing systems, which were laid out during the 2021-2026 period, is expected to bring even more amazing features and applicability into the wider context of society.

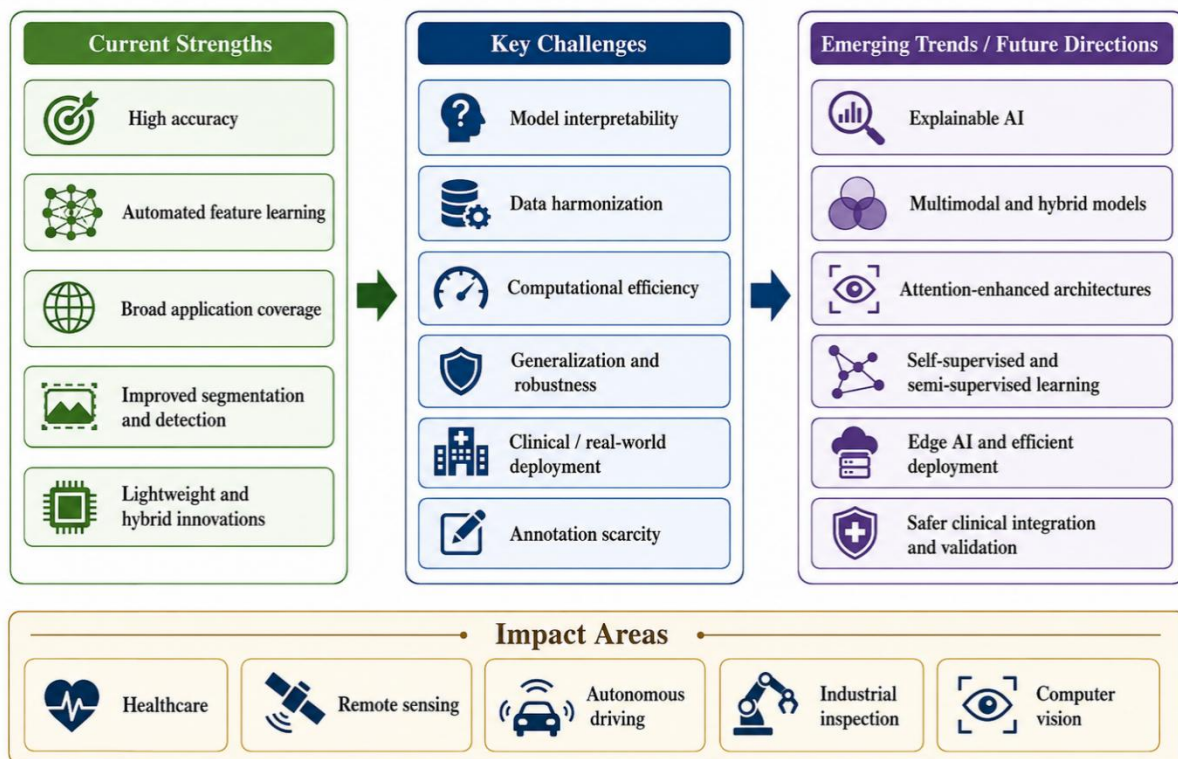


Figure 3 Challenges and Future Directions of AI-Based Image Processing

References

- [1] Maheswari et al., "AI Driven Medical Imaging and Computer Vision Techniques for Enhanced Disease Diagnosis and Treatment Planning," 2025. <https://doi.org/10.71443/9789349552548-08>
- [2] Hammad et al., "Machine Learning for Image Processing in Healthcare," *Advances in Computational Intelligence and Robotics Book Series*, 2025. <https://doi.org/10.4018/979-8-3373-0548-6.ch005>
- [3] Vijayaraghavan et al., "Deep Learning-Enhanced Hybrid Metaheuristic Fusion Model for Biomedical Image Analysis," 2025. <https://doi.org/10.4018/979-8-3373-0523-3.ch009>

- [4] Jain et al., "Deep Learning Techniques for Medical Image Segmentation," *Advances in Computational Intelligence and Robotics Book Series**, 2025. <https://doi.org/10.4018/979-8-3693-9045-0.ch004>
- [5] Heng et al., "Integrating GANs, Contrastive Learning, and Transformers for Robust Medical Image Analysis," *IEEE Journal of Biomedical and Health Informatics**, 2025. <https://doi.org/10.1109/jbhi.2025.3604845>
- [6] Chand, "Mathematical Modeling and Statistical Analysis of Novel Hybrid Deep Learning Models for Efficient Demarcation and Classification of Brain Tumors," *Panamerican Mathematical Journal**, vol. 35, no. 2s, 2024. <https://doi.org/10.52783/pmj.v35.i2s.3009>
- [7] Li et al., "Enhance the Image: Super Resolution using Artificial Intelligence in MRI," 2024. <https://doi.org/10.48550/arxiv.2406.13625>
- [8] Priyah et al., "Hybrid contrast-limited adaptive histogram equalization and Deep Learning techniques for improving liver tumor detection," *Future Technology**, vol. 4, no. 3, 2025. <https://doi.org/10.55670/fpll.futech.4.3.7>
- [9] Tu et al., "Accelerated Feature Extraction and Refinement for Improved Aerial Scene Categorization," *IEEE Transactions on Geoscience and Remote Sensing**, 2024. <https://doi.org/10.1109/tgrs.2024.3403877>
- [10] Sun et al., "Medical image analysis using improved SAM-Med2D: segmentation and classification perspectives," *BMC Medical Imaging**, 2024. <https://doi.org/10.1186/s12880-024-01401-6>
- [11] Billah et al., "A Hybrid Approach to Brain Tumor Detection: Combining Deep Convolutional Networks with Traditional Image Processing Methods for Enhanced MRI Classification," 2024. <https://doi.org/10.15680/ijmrset.2024.0710001>
- [12] Mandala et al., "Hybrid Approach for MRI Segmentation using Deep Learning and Machine Learning Algorithms," 2023. <https://doi.org/10.21203/rs.3.rs-3377680/v1>
- [13] Yu et al., "Dual-Path Adversarial Denoising Network Based on UNet," *Sensors**, vol. 25, no. 15, 2025. <https://doi.org/10.3390/s25154751>
- [14] "ORSI Salient Object Detection via Cross-Scale Interaction and Enlarged Receptive Field," *IEEE Geoscience and Remote Sensing Letters**, 2023. <https://doi.org/10.1109/lgrs.2023.3249764>
- [15] Duke et al., "SSTVOS: Sparse Spatiotemporal Transformers for Video Object Segmentation," in *Proc. Computer Vision and Pattern Recognition (CVPR)**, 2021. <https://doi.org/10.1109/CVPR46437.2021.00585>
- [16] Ignatov et al., "Learned Smartphone ISP on Mobile NPUs with Deep Learning, Mobile AI 2021 Challenge: Report," in *Proc. Computer Vision and Pattern Recognition Workshops (CVPRW)**, 2021. <https://doi.org/10.1109/CVPRW53098.2021.00284>
- [17] Konstantinidis et al., "Multi-Manifold Attention for Vision Transformers," *IEEE Access**, 2023. <https://doi.org/10.1109/access.2023.3329952>
- [18] Esmailzahi et al., "DPAN: A Deep Lightweight Attention-based Image Super Resolution Network using Multi-dimensional Filter Design Technique," *IEEE Signal Processing Letters**, 2023. <https://doi.org/10.1109/lsp.2023.3326387>
- [19] Chimaobi et al., "Using Convolutional Neural Network for Image Classification and Segmentation," *Computer Engineering and Intelligent Systems**, vol. 13, no. 1, 2022. <https://doi.org/10.7176/ceis/13-1-03>

- [20] Bamber et al., "Medical image classification for Alzheimer's using a deep learning approach," *Journal of Engineering and Applied Science**, 2023. <https://doi.org/10.1186/s44147-023-00211-x>
- [21] Qin et al., "Progressive Interaction-Learning Network for Lightweight Single-Image Super-Resolution in Industrial Applications," *IEEE Transactions on Industrial Informatics**, 2023. <https://doi.org/10.1109/TII.2022.3190350>
- [22] Sifka et al., "Malaria Detection Using Advanced Deep Learning Architecture," *Sensors**, vol. 23, no. 3, 2023. <https://doi.org/10.3390/s23031501>
- [23] Li et al., "A deep learning based image enhancement approach for autonomous driving at night," *Knowledge Based Systems**, 2021. <https://doi.org/10.1016/J.KNOSYS.2020.106617>
- [24] Bauer et al., "Vision-Language Model for Accurate Crater Detection," 2026.
- [25] Pérez, "Improving art style classification with synthetic images from self-attention generative adversarial network," 2022. <https://doi.org/10.11606/d.3.2022.tde-25112022-091647>
- [26] Eli et al., "Deep Learning Applications in Medical Image Analysis: Advancements, Challenges, and Future Directions," 2024. <https://doi.org/10.48550/arxiv.2410.14131>
- [27] Gan et al., "Enhancement of Infrared Imagery through Low-Light Image Guidance Leveraging Deep Learning Techniques," *International Journal of Intelligent Systems**, 2024. <https://doi.org/10.1155/2024/8574836>
- [28] "Recent advances and clinical applications of deep learning in medical image analysis," *Medical Image Analysis**, 2022. <https://doi.org/10.1016/j.media.2022.102444>
- [29] Lin et al., "Computational Approaches for Acute Traumatic Brain Injury Image Recognition," *Frontiers in Neurology**, 2022. <https://doi.org/10.3389/fneur.2022.791816>
- [30] Obeso et al., "Visual vs internal attention mechanisms in deep neural networks for image classification and object detection," *Pattern Recognition**, 2022. <https://doi.org/10.1016/J.PATCOG.2021.108411>
- The Role of Modern Artificial Intelligence in Digital Image Processing: A Comprehensive Literature Review (2021-2026)