International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 14, Issue 1, 2026

The Role of Artificial Intelligence and Machine Learning in Computer Vision: From Feature Engineering to Generative Synthesis

Dr. Jeevith S H
Associate Professor
Electronics & Telecommunication Engineering
Sri Siddhartha Institute of Technology, Tumkur, India
jeevith@ssit.edu.in
ORCID ID: 0000-0002-4950-9022

Article Received 7-12-2025, Revised 5-1-2026, Accepted 15-01-2026

Abstract: The domain that has completely changed after the introduction of Artificial Intelligence (AI) and Machine Learning (ML) is Computer Vision (CV), the sphere that allows machines to process and extract meaningful data out of the digital image and video. In this paper, the author conducts a thorough overview of the development of the AI/ML paradigms as they have progressed over augmenting the classical feature-based approaches to becoming the fundamental driving force of the current system of CV. We follow the history down to starting with the early classifiers, such as Support Vector Machines (SVM) using manually-constructed features to classifiers powered by deep learning ushered in by Convolutional Neural Networks (CNNs), capable of automatically extracting features and performing tasks such as object detection and image classification better than a human being. The review also examines the state of the frontiers, such as use of vision transformers, image synthesis and image augmentation by generative models, and multimodal learning which combines visual information with other modalities. By relying on a broad spectrum of applications, such as in the fields of medical image analysis as well as autonomous systems but also content-based retrieval and behavioral understanding, we combine results of the most relevant literature to point out breakthroughs, current challenges, as well as future trends. Such crucial problems as model interpretability, data-efficiency, algorithm bias, and real-time processing constraints are discussed in details. We conclude that AI/ML has not only improved computer vision but changed its very nature, making it possible to perform more than ever before in terms of perception, understanding, and creation of visual content with significant effects in the fields of science, industry, and society.

Keywords: Artificial Intelligence, Machine Learning, Computer Vision, Deep Learning, Convolutional Neural Networks, Object Detection, Image Segmentation, Generative Models, Explainable AI.

1. Introduction

Computer Vision seeks to mimic and go beyond the visual perceptions of humans in allowing machines to localise objects, comprehend scenes, track motions and recreate three dimensional surroundings out of two-dimensional data. In the past, this was realized, using elaborate algorithms that were constructed on the geometry model and manually biased features (e.g., SIFT, HOG). With the invention of Machine Learning, and later Deep Learning, the shift has been in the paradigm of designing features to the more directly graphical learning of representations directly out of data.

This development is in line with larger tendencies in the AI field, as rule-based systems have already been replaced by data-driven models in many areas, including behavioral economics (Ghule, 2025) and anomaly detection in finance (Ghori, 2018). Accepting raw pixels, which obtain intelligence, AI/ML in CV is the computational cerebral ganglion. The present paper defines the diverse role of AI/ML in CV, which is organized in a sequence of simple perception processes to complex generative and cognitive functions, and placed in the context of the other relevant literature.

2. Foundational Roles: Perception and Recognition

2.1. Automated Feature Learning with Deep Neural Networks

The historic invention was the design of Convolutional Neural Networks (CNNs), which makes the hierarchical process of feature extraction automatic, i.e., the extraction of edges and textures in lower layers, and the object components and classes in the deeper ones.

- ***Image Classification:*** CNNs, such as AlexNet, VGG, ResNet, and Inception, have dominated benchmarks (e.g.

ImageNet) reaching at the top of human accuracy rates. Such architectures as InceptionResNetV2 and ResNet50 can be applied to classification, but it is also a powerful feature extractor used in the Content-Based Image Retrieval (CBIR) task as shown in the architecture of reverse search engines (Marathe et al., 2022). In this case, ML gives the social framework with the ability of transforming pictures into succinct, semantically meaningful "feature vectors" that allow the wise similarity exploration.

- ***Object Detection and Segmentation:*** The next step above classification is object localization and object delineation, which can be achieved with such ML models as R-CNN, YOLO (You Only Look Once) and Mask R-CNN. This is essential to an autonomous driving application like medical image analysis, in which it is necessary to divide organs or tumors as the initial stage of radiomics pipelines (Amudala Puchakayala et al., 2023; Saha et al., 2025).

2.2. Enhanced Traditional ML with Learned Features

In the pre-deep learning hegemony, not only were supervised ML classifiers critical but also let go of their importance. Systems would pull out features of handcraft which models would include:

- ***Support Vector Machines (SVM):*** Applicable to high dimensional feature space. They continue to be useful with extra functionality added to the expected results such as Bayesian Optimization (BO-SVM) used to tune hyperparameters, such as in drowsiness detection with highly accurate EEG signal classification results (Sardesai and Gedam, 2025). The first and most popular choice of classifier in early CV was the SVMs, following the feature extraction.
- ***Random Forests and Gradient Boosting:*** These ensemble algorithms are sound and they give feature significance measures. They are also very useful, especially in multimodal or hybrid systems where interpretability is appreciated, as is the case in predicting COPD progression based on a combination of clinical and imaging characteristics (Saha et al., 2025) or in prediction of soil quality in smart agriculture (Kumar et al., 2023).

3. Advanced Roles: Understanding, Generation, and Cognition

3.1. Scene Understanding and Multimodal Integration

Modern CV is interested in achieving a comprehensive perception of visual images, relationships between objects, context, as well as intent.

- **Multimodal Machine Learning (MML):** The model of multimodal machines is capable of fusing visual information with other modalities: text (image captioning), audio (video analysis), sensor measurements. MML research can deal with fundamental issues of representation, translation, alignment, and fusion (Sardesai et al., 2025). As an example, an AI-assisted autism detection system may be a combination of facial expression detection (CV) and gait sensor and behavioral logs (S, Shalini et al., 2025).
- **Video Analysis and Action Recognition:** Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks, which are skilled at the modeling of sequences, are utilised to the video frames to identify actions, track behaviors and determine future states. This is comparable to how they are applied in multi variable time series prediction of electricity demand (Ghori, 2019).

3.2. Generative and Synthetic Vision

AI/ML has enabled CV to see not only, but also to build and enhance.

- **Generative Adversarial Networks (GANs) and Diffusion Models:** This can be used to create photorealistic images, image-to-image translation (e.g. day to night, CT to MRI), and generate data to train. This is vital in healthcare in which statistics are meager and confidential. GAN-based frameworks, such as Spatio-Convolutional GAIN (S-CGAIN), are also modified to do more important tasks such as missing data imputation in complicated data (Bansal et al., 2025).
- **Data Augmentation and Domain Adaptation:** ML can enhance model robustness and generalizability through the automatic augmentation of training datasets (e.g. by random rotations, contrast manipulations or style transfer) to help deep learning solve the problem of data hunger.

3.3. Vision for Autonomous Systems and Robotics

The biggest sense of autonomous agents is the CV. ML combines the perceptions and decision-making.

- **Simultaneous Localization and Mapping (SLAM):** ML-accelerated SLAM involves visual data to assist robots and drones in creating the map of an unfamiliar environment and locating themselves in the middle of it.

- **Human-Robot Interaction:** CV models can help robots to recognize hand gestures (Sheela et al., 2022), have the ability to identify faces, and apprehend the human activity to collaborate naturally.

4. Domain-Specific Applications Powered by AI/ML in CV

1. **Healthcare and Medical Imaging (Radiomics):** It is an outstanding model. Disease detection (COPD, cardiac risks), monitoring of its progression, and planning of its treatment are analyzed using AI/ML models, in particular, CNNs and radiomics pipelines, which operate with X-rays, CTs, and MRIs (Sheela and Shalini, 2024; Amudala Puchakayala et al., 2024). These changes can be characterized as the shift of the role to the quantitative previewing analytics.
2. **Environmental Monitoring:** CV uses satellite and Synthetic Aperture Radar (SAR) images (Kavitha et al., 2017) to monitor deforestation, urban planning, and water consequential pollution. ML models categorize land use, identify the changes and the quantification of environmental measures.
3. **Intelligent Transportation Systems (ITS):** The CV drives the traffic analyzes and license plate technology alongside in person radar. CV information combined with vehicle-to-vehicle communication favors resource distribution in Vehicle Ad-Hoc Networks and improves the safety (Sheela et al., 2023).
4. **Security and Surveillance:** Real-time video analytics with the help of the efficient ML models enhance the automated threat detection, analysis of how crowds move, and identification of anomalies in the social environment.

5. Key Issues and Research Areas

Although dramatic advances have been made, there are still issues, most of which are the points of interest in the wider research of AI ethics and systems:

1. **Explainability and Interpretability (XAI):** This is a limitation of the "black box" character of deep neural networks and is one of the reasons why these networks have not yet found high-stakes application in such domains as medicine or law. It is of urgent necessity to render the decisions of CV models transparent and reliable enough, and this problem is mirrored in debates about Responsible AI (Puchakayala, 2022) and how algorithms can be fair regarding consumers (Ghule, 2025).
2. **Data Efficiency and Learning with Less Labels:** Deep learning usually needs tremendous number of labeled data. The field of semi-supervised, self-supervised, and few-shot learning studies are set to minimize such a dependency, as do problems in disaster management model development due to little past events (Ghori, 2021).
3. **Algorithmic Bias and Fairness:** CV models are capable of taking in and exerting more bias on the training data, causing biased results at various demographics. Bias works against datasets and models: it is an ethical requirement to actively debias them.
4. **Computational Efficiency and Real-Time Processing:** When complex models (e.g. large vision transformers) need to be deployed on edge devices with restricted power (IoT sensors, mobile phones), the models need to be compressed, quantized, and deployed using architecture designs that are efficient.
5. **Robustness and Adversarial Attacks:** CV models can be easily fooled by adversarial perturbations which are well-designed noise that is they add to an image to be misclassified. The strength is essential to the safety-critical applications.

6. Conclusion and Future Directions

It can be argued that AI and Machine Learning are no longer a tool but the material of contemporary computer vision. They have facilitated the jump of the primitive pattern recognition to the advanced scene perception, predictive analytics and content generation. With the development of the field, such convergence of CV with other branches of AI as multimodal learning, responsible AI, and streamlined ML practices will become essential in order to construct a not only intelligent, but also accurate, and efficient system, not to mention fair and consistent with human values. AI/ML in CV role is also, therefore, both core and radical in nature, as it progresses innovation in practically all fields of the global economy.

The future of AI/ML in CV is becoming increasingly integrated, efficient and more cognizant:

- **Foundation Models for Vision:** Large scale, pre-trained models (as in CLIP, vision-language understanding) that can be trained with a consequent application of little additional fine-tuning to large set of downstream tasks.

- **Neuromorphic and Energy-Efficient Vision:** Systems Hardware and algorithms inspired by the human visual system to enable ultra-low-power and real-time vision processing.
- **Causal Computer Vision:** Required to infer more than correlation to get the causal mechanisms behind visual scenes, which make it possible to reason and generalize.
- **Ethical and Federated Vision:** Learn how to build the frameworks of privacy preserving CV with methods of federated learning, where models are trained on devices which do not share raw visual data.

References

1. Amudala Puchakayala, P. R., Sthanam, V. L., Nakhmani, A., Chaudhary, M. F., Kizhakke Puliyakote, A., Reinhardt, J. M., & Bodduluri, S. (2023). Radiomics for improved detection of chronic obstructive pulmonary disease in low-dose and standard-dose chest CT scans. Radiology, 307(5), e222998.
2. Bansal, A., Puchakayala, P. R. A., Suddala, S., Bansal, R., & Singhal, A. (2025, May). Missing Value Imputation using Spatio-Convolutional Generative Adversarial Imputation Network. In 2025 3rd International Conference on Data Science and Information System (ICDSIS) (pp. 1-6). IEEE.
3. Ghori, P. (2018). Anomaly detection in financial data using deep learning models. International Journal Of Engineering Sciences & Research Technology, 7(11), 192-203.
4. Ghori, P. (2019). Advancements in Machine Learning Techniques for Multivariate Time Series Forecasting in Electricity Demand. International Journal of New Practices in Management and Engineering, 8(01), 25-37. Retrieved from https://ijnpme.org/index.php/IJNPME/article/view/220
5. Ghori, P. (2021). Enhancing disaster management in India through artificial intelligence: A strategic approach. International Journal of Engineering Sciences & Research Technology, 10(10), 40–54.
6. Ghule, P. A. (2025). AI in Behavioral Economics and Decision-Making Analysis. Journal For Research In Applied Sciences And Biotechnology, Учредители: Stallion Publication, 4(1), 124-31.
7. Kavitha, T. M., Sheela, S., Thenappan, S., & Pradeep Kumar, M. S. (2017). Synthetic aperture radar (SAR) imaging using global back projection (GBP) algorithm for airborne radar systems. In Proceedings of the International Conference on Current Trends in Engineering, Science and Technology (ICCTEST), 1071–1076. Grenze Scientific Society.
8. Kumar, P. R., Meenakshi, S., Shalini, S., Devi, S. R., & Boopathi, S. (2023). Soil Quality Prediction in Context Learning Approaches Using Deep Learning and Blockchain for Smart Agriculture. In Effective AI, Blockchain, and E-Governance Applications for Knowledge Discovery and Management (pp. 1-26). IGI Global.
9. Marathe, S., Arosh, S., & Mondal, T. (2022, December). A CNN-inspired reverse search engine for content-based image retrieval. In 2022 5th International Conference on Contemporary Computing and Informatics (IC3I) (pp. 996-1001). IEEE.
10. Puchakayala, P. R. A. (2022). Responsible AI Ensuring Ethical, Transparent, and Accountable Artificial Intelligence Systems. Journal of Computational Analysis and Applications, 30(1).
11. Saha, P., Bodduluri, S., Nakhmani, A., Chaudhary, M. F., Amudala Puchakayala, P. R., Sthanam, V., & Bhatt, S. P. (2025). Computed tomography radiomics features predict change in lung density and rate of emphysema progression. Annals of the American Thoracic Society, 22(1), 83-92.
12. Sardesai, S., & Gedam, R. (2025, February). Hybrid EEG Signal Processing Framework for Driver Drowsiness Detection Using QWT, EMD, and Bayesian Optimized SVM. In 2025 3rd International Conference on Integrated Circuits and Communication Systems (ICICACS) (pp. 1-6). IEEE.
13. Sardesai, S., Kirange, Y. K., Ghori, P., & Mahalaxmi, U. S. B. K. (2025). Secure and intelligent financial data analysis using machine learning, fuzzy logic, and cryptography. Journal of Discrete Mathematical Sciences and Cryptography, 28(5-B), 2163–2173.
14. Shalini, S., Nandini, C., Lakshmi, M. R., Biswas, K., Divyashree, L., Kumar, M. A., & Monika, V. (2025). AI-enhanced comprehensive autism detection and support platform for all ages. In Proceedings of the International Conference on Innovative Computing &

ISSN 2347–3657

Volume 14, Issue 1, 2026

Communication (ICICC 2024). https://doi.org/10.2139/ssrn.5241652

15. Shalini, S., Sree, K. M., Prajwal, M. H., Reddy, N. N., & Reddy, P. G. (2024, April). BIOWATCH: Detecting the Lake Pollution to Safeguard Biotic Life. In 2024 International Conference on Emerging Technologies in Computer Science for Interdisciplinary Applications (ICETCS) (pp. 1-6). IEEE.
16. Sheela, S., & Shalini, S. (2024). Prediction of cardiac disabilities in diabetic patients. In Futuristic trends in network & communication technologies (IIP Series, Vol. 3, Book 4, Part 2, Chapter 2, pp. 123–129). Integrated Intelligent Publication.
17. Sheela, S., Nataraj, K. R., & Mallikarjunaswamy, S. (2023). A comprehensive exploration of resource allocation strategies within vehicle Ad-Hoc Networks. Mechatron. Intell. Transp. Syst, 2(3), 169-190.
18. Sheela, S., Nataraj, K. R., & Rekha, K. R. (2022, November). Design and Implementation of Hand Gesture for Various Applications. In Proceedings of the International Conference on Cognitive and Intelligent Computing: ICCIC 2021, Volume 1 (pp. 721-728). Singapore: Springer Nature Singapore.