IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 1, August 2025
Impact Factor: 7.67

Towards Interpretable and Efficient Agricultural Image Classification: A Review of Autoencoder-Enhanced YOLOv8 Architectures with Spatial Attention and Feature Compression

[1]Ms. Deepak B. Patil and [2]Mr. Prathmesh S. Powar
[1]Research Scholar,[2]Asst. Professor, [1]Computer Science and Engineering,
[2]Computer Science and Engineering,
[1]Ashokrao Mane Group of Institutions, Vathar, Kolhapur, India
[2]Ashokrao Mane Group of Institutions Kolhapur, Vathar, India
[1]deepakbpatil87@gmail.com , [2]psp@amgoi.edu,in

Abstract: *This review presents a comprehensive analysis of recent advancements in agricultural image classification using deep learning, emphasizing three key innovations: feature compression via autoencoders, spatial attention mech- anisms, and model interpretability. Autoencoders efficiently reduce high-dimensional agricultural imagery, while attention modules like CBAM and PAM enhance spatial focus and feature refinement. YOLOv8, known for its lightweight design, is explored for crop classification tasks, with added interpretability through Grad-CAM and SHAP. Through an extensive literature survey, we compare model architectures, datasets, and performance out- comes across tasks like disease detection, crop type classification, and weed identification. The review identifies current research gaps, including the limited integration of compression and interpretability in unified frameworks. We conclude by proposing future directions toward efficient, interpretable, and real-time deployable deep learning systems for precision agriculture*.

Keywords: Grad-CAM, Image classification, Agricultural classification, autoencoder

I. INTRODUCTION

The advent of deep learning has revolutionized agricultural image analysis, offering unprecedented accuracy and automation in tasks like crop classification, disease detection, and weed identification. However, the high dimensionality of remote sensing and UAV-acquired images, combined with the need for interpretability and real-time processing, presents unique challenges. Traditional convolutional neural networks (CNNs), though effective, often struggle with overfitting, large memory footprints, and lack of transparency in decision-making. To address these limitations, researchers have turned to a synergistic integration of three key techniques: feature compression using autoencoders, spatial attention mechanisms, and interpretable machine learning models.

Autoencoders provide robust dimensionality reduction, enabling efficient training on high-resolution data. Spa- tial attention modules such as CBAM and PAM enhance feature focus by guiding the model to attend to discrimi- native regions, especially in complex agricultural scenes. Additionally, interpretability techniques like Grad-CAM and SHAP are gaining traction for unveiling the inner workings of these models—ensuring trust and transparency in decision-critical environments. This review aims to comprehensively explore the latest innovations and compar- ative performance of these components, highlighting their individual and combined potential in building accurate, lightweight, and interpretable systems for modern precision agriculture.

II. AUTOENCODERS FOR FEATURE COMPRESSION IN AGRICULTURE

Autoencoders (AEs) have gained significant traction in agricultural image analysis as a means to reduce data dimensionality, denoise images, and generate compressed yet informative feature representations. In high-resolution agricultural imaging tasks such as hyperspectral analysis, plant disease detection, and remote sensing the abundance of redundant or noisy pixel information creates challenges for efficient and accurate classification. Autoencoders, particularly their advanced variants like VariationalAutoencoders (VAEs) and Graph-Regularized AEs, offer a solution by learning compact, latent representations that capture essential patterns while discarding noise and irrelevant details.A prominent application of VariationalAutoencoders in agriculture was demonstrated by [1], who tackled the lack of explainability in plant disease classification. Using a disentangled VAE on the PlantVillage dataset, they achieved not only high classification accuracy on crops like potato virus Y but also enabled visual interpretability without relying on external heatmaps. Their approach allowed the generation of image variants controlled by specific latent features, thereby bridging classification performance and interpretability a rare achievement in agricultural deep learning models. In the hyperspectral domain, [2] employed stacked autoencoders to compress high-dimensional hyperspectral images (HSIs) from datasets like Salinas and Botswana. These images, while rich in information, are computationally expensive to process. The authors used autoencoder-based compression followed by a 3D-2D CNN classifier, resulting in peak signal-to-noise ratios (PSNR) exceeding 60 and classification accuracies above 99%. Their dual emphasis on reconstruction fidelity and classification efficiency demonstrated how AEs can enhance both robustness and practicality for real-world deployment.The foundational work by [3] was among the first to integrate autoencoders in hyperspectral remote sensing. They applied PCA on spectral dimensions and deep autoencoders on spatial features, improving classification accuracy over SVM and PCA-SVM baselines. Similarly, [4] introduced graph-regularized autoencoders to preserve the spatial continuity inherent in agricultural landscapes. By encoding spatial–spectral relationships, they achieved superior classification performance compared to classical AEs, especially in tasks like land cover identification.

[5] investigated the stability of various feature extraction methods and confirmed that VAEs produced more consistent and meaningful representations than PCA and traditional AEs. Their findings emphasized the importance of probabilistic encoding, particularly when dealing with normalization-sensitive agricultural datasets. Beyond classification, [6] proposed an autoencoder-augmented OFDM (Orthogonal Frequency Division Multiplexing) framework to accelerate image transmission in IoT-enabled smart agriculture. Their approach reduced latency and congestion while preserving image integrity, showing that autoencoders can serve not only as learning modules but also as communication optimizers in agricultural cyber-physical systems. In unsu- pervised classification, [7] leveraged convolutional autoencoders with K-means clustering, achieving strong results on general-purpose datasets like MNIST and CIFAR-10. While not agriculture-specific, the architecture offers transferable insights for crop clustering or anomaly detection. Meanwhile, [8] improved spatial coherence in hyperspectral AEs by modifying the loss function to enforce neighborhood similarity, thereby producing smoother and semantically rich latent features for land cover classification.

Despite these advances, a critical research gap remains: most autoencoder-based studies stop at feature extraction or dimensionality reduction, rarely integrating AEs into full classifier pipelines especially with modern lightweight models like YOLOv8. This disconnection limits the full potential of AEs as pre-classification modules, especially in spatially rich agricultural contexts.

III. YOLOV8 AND LIGHTWEIGHT CLASSIFICATION NETWORKS IN AGRICULTURAL APPLICATIONS

YOLOv8, the latest in the You Only Look Once family of object detection models, has expanded beyond detection into the realm of classification with the introduction of its classifier variant (e.g., yolov8n-cls.pt). Originally optimized for speed and accuracy in real-time object detection, recent adaptations of YOLOv8 now address the growing demand for lightweight, interpretable, and mobile-friendly deep learning models especially in agriculture, where devices deployed in fields and farms require computational efficiency without compromising accuracy. This section reviews recent work focused on tailoring YOLOv8 and its lightweight variants to agricultural and environmental tasks. In the domain of

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 1, August 2025 Impact Factor: 7.67

plant disease recognition, [9] proposed a lightweight YOLOv8 architecture enhanced with ODConv and Wise-IoU (WIoU) loss for accurate rice blast disease detection. Their model achieved a 66.6% reduction in parameters and a 61.9% decrease in model size, while also improving mAP by 5.2% over the baseline YOLOv8n. This design was explicitly optimized for mobile deployment, highlighting YOLOv8's adaptability to resource-constrained agricultural settings.

In industrial agriculture, [10] applied YOLOv8n in combination with MobileNetV3 and SimAM attention to detect defects in power transmission insulators. Although not a crop-focused application, the insights are transferable to agricultural sensor networks. With a 90.6% mAP@50 and a 40.54% increase in training speed, this hybrid architecture showed how integrating YOLOv8 with lightweight backbones and efficient attention mod- ules can drastically improve detection speed and accuracy under real-time constraints. Real-time classification of agricultural products has also benefitted from YOLOv8 innovations. [11] developed a cascaded YOLOv8 system enhanced by SRGAN (for super-resolution) and Siamese data augmentation to classify star anise varieties. Their system achieved a remarkable 96.37% mAP at 146 FPS while using only 7.4% of the YOLOv3 model size, validating YOLOv8's efficiency for classification tasks in resource-limited environments.

For aerial agricultural monitoring, [12] introduced YOLOv8-LD, incorporating pruning techniques, an ASBiFPN neck, and MPDIoU loss for drone-based object detection. With an 81% reduction in parameters and a 67% smaller model size, they retained high accuracy (mAP@0.5 ↑21%), demonstrating that YOLOv8 can be pruned and still perform competitively for high-resolution aerial tasks. Pest detection, a critical problem in precision agriculture, was addressed by [13] using LP-YOLO, a YOLOv8 derivative enhanced with Efficient Channel and Spatial Attention (ECSA) and pruning strategies. Their system reduced parameters by 70.2% and improved FPS by 40.7%, with only a minor 0.8% drop in mAP, offering an optimal balance between accuracy and efficiency for on-field pest monitoring.

In crop-level weed detection, [14] improved the YOLOv8s backbone with D-PP-HGNet and novel dual downsampling techniques, raising accuracy from 91.2% to 95.8% while cutting model size by nearly 60%. These optimizations make embedded deployment in agricultural robots more feasible without compromising model precision. Addressing disease detection in tea leaves, [15] introduced YOLOv8-RCAA, which uses a RepVGG backbone combined with CBAM attention. Achieving 98.14% mAP, this anchor-free model out- performed YOLOv5, SSD, and Faster-RCNN, underscoring YOLOv8's dominance in lightweight, real-time crop disease classification. Finally, [12] developed YOLOv8-PG for detecting pigeon eggs in poultry farms. The integration of EMA attention and Fasternet Block yielded a 4.45% mAP75 improvement, while reducing parameters by 24.69% and compute costs by 22.89%, reinforcing YOLOv8's role in robotic automation and animal farming.

Collectively, these studies underscore the flexibility and efficiency of YOLOv8 when customized for agricultural use cases. However, despite growing applications in detection, its use in classification pipelines especially in conjunction with autoencoders or attention modules remains sparse. This gap suggests significant opportunities for future exploration into fully integrated, lightweight classification systems for agricultural intelligence.

IV. SPATIAL ATTENTION MECHANISMS IN CROP CLASSIFICATION

The growing complexity and diversity of agricultural imagery driven by high-resolution remote sensing, drone surveillance, and ground-level imaging have introduced new challenges in crop classification. Traditional convolutional neural networks (CNNs), while effective at general feature extraction, often lack the ability to selectively emphasize spatially relevant features, especially when dealing with subtle visual cues such as small lesions, overlapping plant structures, or heterogeneity in plant morphology. Spatial attention mechanisms have emerged as a powerful tool to address these challenges by allowing models to focus on informative regions of input images, enhancing both classification accuracy and interpretability. This section reviews key developments in spatial attention modules including CBAM, SE, Dual Attention, and custom attention within the context of crop classification. [16] introduced a Geo-CBAM module that fused spatial and channel attention with geographic contextual information for crop classification using Sentinel-2 satellite imagery. Their model achieved 97.82% classification accuracy across six counties in the U.S., significantly outperforming both CNN and Random Forest baselines. Notably, the attention module adapted well to spatial heterogeneity between different regions, demonstrating how geo-aware spatial attention enhances robustness in large-scale agricultural mapping.

Similarly, [17] focused on crop-type identification from multi-spectral satellite imagery. Their model employed a Dual Attention Module (DAM) that jointly leveraged spectral and spatial attention streams to refine feature representations. Using over 200,000 Sentinel-2 image samples across six crop categories, they reported an impressive 98.54% accuracy surpassing strong baselines such as XGBoost and standard CBAM. The dual-stream design enabled better modeling of textural patterns and reflectance variations unique to each crop type, especially in high-dimensional data.

At the leaf level, [18] applied CBAM to ResNet50 for cassava disease classification. The spatial attention component helped isolate lesion-affected regions from healthy leaf areas, resulting in a 97% classification accuracy and clear improvements over standard CNNs. This work highlighted the role of spatial attention in disease localization, particularly for crops where visual symptoms are subtle and spatially dispersed.In segmentation based tasks, [19] incorporated CBAM with CARAFE (Content-Aware Re- Assembly of FEatures) into a DeepLabv3+ architecture to tackle overlapping disease regions in rice leaves. Their model significantly outperformed baseline DeepLabv3+ in pixel-wise accuracy, with CBAM improving spatial focus and CARAFE enhancing boundary resolution. This combination proved especially effective in distinguishing between closely related diseases such as bacterial blight and brown spot.

For real-time agricultural applications, [20] developed ECENet a lightweight model combining CBAM with EfficientNetB0. Their design introduced parallel spatial and channel attention pathways optimized for low-resource deployment. Tested on a corn weed classification dataset, ECENet achieved 99.92% accuracy with minimal computational overhead, making it ideal for mobile and embedded systems in precision farming.Expanding beyond generic attention modules, [21] proposed a novel bit-plane and correlation- based spatial attention mechanism within ResNet101. Designed specifically for plant disease detection, the module captured localized lesion patterns by modeling feature correlations across spatial neighborhoods. Evaluated on the AI Challenger and PlantVillage datasets, their model reached up to 99.82% accuracy, outperforming SE and CBAM in small lesion detection.

[22] tackled tomato disease severity classification using SEV-Net a network embedding spatial and channel attention within ResNet blocks. With 97.59% accuracy, the model supported real-time inference on Android platforms and provided visual explanations through Grad-CAM and saliency maps. This made it both performant and interpretable a crucial combination for practical agricultural diagnostics.Finally, [23] designed the Global Spatial Coordinate Attention Module (GSCAM), which com- bines the strengths of SE and CBAM while preserving computational efficiency. Though tested on fine-grained benchmarks like birds and flowers, its emphasis on global location-aware cues has direct applicability to detailed crop classification tasks, especially when precise spatial information is critical.

In summary, spatial attention mechanisms have significantly advanced crop classification by enhancing feature localization, improving generalization across geographies, and enabling lightweight real-time models. However, most of these approaches remain centered around segmentation or disease detection. The integration of spatial attention within end-to-end classification pipelines particularly when combined with dimensionality-reduction modules like autoencoders and lightweight classifiers such as YOLOv8 remains underexplored. This opens a promising frontier for future research into compact, interpretable, and spatially aware agricultural AI systems.

V. INTERPRETABLE DEEP LEARNING IN AGRICULTURE

As deep learning becomes increasingly integral to agriculture, particularly in tasks like crop classification, disease detection, and yield prediction, a critical challenge has emerged: interpretability. For agricultural practitioners and stakeholders to trust and adopt AI-driven insights, models must offer transparent reasoning. The movement toward interpretable deep learning, although widely studied in medical domains, offers valuable lessons that are transferrable to agriculture.

Medical imaging research provides a strong foundation for interpretability techniques. [24] examined the transparency of deep learning models in chest radiography. Their comparative study of SHAP, LIME, GradCAM, and LRP revealed that Grad-CAM excelled quantitatively, while LIME delivered semantically rich explanations. The study highlighted the potential of multimodal interpretability a concept that could be extended to agricultural settings where multispectral

ISSN: 2581-9429
Impact Factor: 7.67

or multimodal sensor data is often used. Similarly, [25] explored CNNs applied to breast mammograms and emphasized how Grad-CAM could reliably identify malignant regions, offering visual correlation with pathology types. These insights mirror the challenges in plant disease classification, where distinct regions of leaves or crops may exhibit disease symptoms that demand explain- able localization.

Grad-CAM, in particular, emerges as a recurring tool for making convolutional models interpretable. [26] successfully integrated Grad-CAM with CNNs like ResNet and EfficientNet for COVID-19 X-ray detection, enabling medical professionals to validate predictions with saliency maps. In agriculture, such integration could aid agronomists in visually verifying regions affected by pest infestations or nutrient deficiencies. Similarly, [27] used Grad-CAM++ and SHAP with DenseNet201 to explain multiclass lung disease classifications, deploying their model on Android for field diagnostics an approach highly relevant to mobile agricultural advisory systems in rural regions.

Beyond visual explanations, hybrid interpretability methods improve trust in time-series and signal-based data, which is also applicable in precision agriculture. [28] combined SHAP, Grad-CAM, and Partial Dependence Plots to analyze CNN and LSTM models on ECG signals, highlighting the importance of both global and local interpretation. [29] further demonstrated that interpretable models could achieve high accuracy without sacrificing explainability by combining Grad-CAM and SHAP for arrhythmia classification, leading to improved clinician trust. This dual strategy can be extended to interpret spectral signatures in remote sensing based crop analysis or time-series data from soil sensors.

The robustness and generalizability of interpretability tools are equally critical. [30] emphasized this by proposing the Forward Backward Interpretability framework and identifying the stability of Grad-CAM under data perturbations. [31] broadened the scope by applying Grad-CAM across architectures CNNs, ViTs, and Swin Transformers demonstrating its utility beyond traditional networks. With Transformer-based models gaining popularity in agricultural image analysis, such visual tools become increasingly vital.

Interpretability tools like Grad-CAM, SHAP, and LIME have proven effective across domains for rendering black-box models more transparent. Drawing parallels from medical and signal processing domains, these methods can significantly enhance trust and usability of deep learning in agriculture. Ensuring explainability will not only improve user confidence but also foster adoption in mission-critical applications like disease diagnostics, soil health assessment, and yield forecasting, where human oversight remains essential.

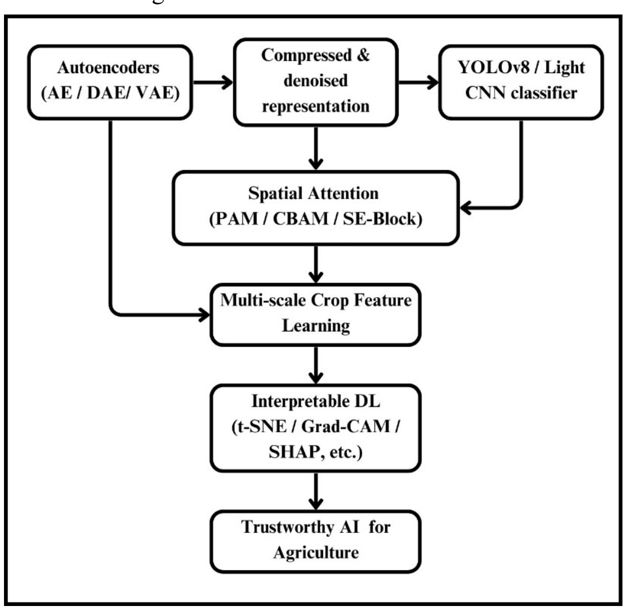

Fig 1 :Overview of Key Deep Learning Enhancements in Agricultural Image Classification

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 1, August 2025 Impact Factor: 7.67

VI. COMPARATIVE EVALUATION OF INTERPRETABILITY TECHNIQUES FOR DEEP LEARNING IN AGRICULTURE-INSPIRED HEALTHCARE CONTEXTS

Table 1: Comparative Summary of Interpretability Techniques

Study	Method	Unique Insight
Alam et al., 2023	SHAP, LIME, Grad-CAM, LRP	Multimodal interpretability enhanced trust
Balve& Hendrix, 2024	SHAP, LIME, Grad-CAM	Strong visual correlation with pathology
Ali, 2025	Grad-CAM with CNNs	Clinically validated explanations
Verma, 2022	SHAP, PDP, Grad-CAM	QRS highlighting increased interpretability
Zeng, 2024	Grad-CAM, SHAP	Full accuracy + visual saliency
Mahamud, 2024	SHAP, Grad-CAM++, LIME	Android-based interpretable model deployment
Kokate, 2018	Grad-CAM vs F-B Interpretability	Identified robustness gaps
Shen & Huang, 2024	Grad-CAM on CNNs/Transformers	Transformer explainability validated

The integration of autoencoders, attention mechanisms, lightweight classification backbones, and interpretability tools like t-SNE or Grad-CAM forms a robust pipeline for agricultural classification. These components work synergistically to address challenges such as noise, scale variance, and model transparency. An overview of this integrated approach is illustrated in Figure 1 , highlighting the complementary nature of compression, attention, classification, and interpretability in precision agriculture. Interpretability is a pivotal factor for deploying deep learning models in critical domains such as agriculture and healthcare, where decisions must be explainable to domain experts. This section presents a comparative evaluation of widely used interpretability methods Grad- CAM, SHAP, LIME, and others across recent studies (2023–2025) that deal with medical and pathology-oriented image datasets. These works serve as analogs for developing interpretable models in agriculture, particularly in scenarios involving plant disease diagnosis, yield prediction, or remote sensing.

From the review as suggested in Table 1 , Grad-CAM emerged as the most widely adopted method across various CNN-based architectures including DenseNet, ResNet, and EfficientNet. It demonstrated robust performance in visualizing relevant image regions (e.g., lesions in lungs, QRS complexes in ECGs) while maintaining consistency under model and data perturbations [30]. Studies such as [26] and [19] found Grad-CAM particularly effective in high-stakes scenarios like COVID-19 and cardiac diagnostics, boosting expert confidence.LIME, though less quantitatively dominant, proved useful in generating class-discriminative and medically intuitive explanations [24], [25]. SHAP offered a global-local interpretability balance, as seen in [27], [28], and performed well when paired with saliency-based methods like Grad-CAM++.Interestingly, [31] extended Grad-CAM utility beyond CNNs to transformer based models, indicating its adaptability for newer architectures. Collectively, these findings advocate for a hybrid interpretabilitypipeline, combining spatial heatmaps (e.g., Grad-CAM++) with feature attribution techniques (e.g., SHAP/LIME), especially for field-deployed applications in agriculture where model trustworthiness is essential.

VII. COMPARATIVE EVALUATION OF INTERPRETABILITY TECHNIQUES FOR DEEP LEARNING IN AGRICULTURE-INSPIRED HEALTHCARE CONTEXTS

Despite recent advancements, significant gaps remain in the integration of interpretability, attention mechanisms, and compression techniques within agricultural deep learning pipelines. Most studies focus on either classification accuracy or model efficiency, often neglecting explainability and generalizability in real-world agricultural scenarios. While spatial attention modules like CBAM have improved disease localization and multiscale feature extraction, they are rarely embedded within lightweight, interpretable classifiers such as YOLOv8. Similarly, autoencoders are widely used for feature compression but are seldom integrated in end-to-end classification models. Additionally, although Grad-CAM and SHAP have shown promise in healthcare, their adoption in agricultural image analysis particularly in UAV or hyperspectral data is still limited. A lack of benchmarking datasets and standardized evaluation protocols further impedes reproducibility and progress. These gaps underscore the need for unified, hybrid frameworks that combine interpretability, efficiency, and accuracy in a scalable and agriculturally relevant manner.

ISSN: 2581-9429
Volume 5, Issue 1, August 2025
Impact Factor: 7.67

VIII. CONCLUSION AND FUTURE SCOPE

This review synthesizes recent developments in feature compression, spatial attention, and interpretability techniques highlighting their individual strengths and their potential synergy in advancing agricultural deep learning. Autoencoders have demonstrated powerful denoising and dimensionality reduction capabilities, especially in hyper- spectral and UAV datasets. YOLOv8, with its lightweight classification variant, has emerged as a strong candidate for mobile-friendly crop monitoring applications. When integrated with spatial attention modules like CBAM or PAM, these models show enhanced multi-scale learning and localized focus critical for precision agriculture. However, the black-box nature of such deep learning systems still hinders trust and adoption in real world farming scenarios.

Future research should focus on developing unified pipelines that combine autoencoder-based compression, YOLO-based classification, and attention-driven spatial modeling while embedding interpretability modules like Grad-CAM and SHAP for transparent decision-making. Benchmarking these hybrid models on large-scale, real-world agricultural datasets, including temporal and multispectral data, will be crucial. The incorporation of explainable AI on edge devices also presents exciting opportunities for real-time, farmer-friendly deployment. Such integrated, interpretable, and efficient systems will significantly propel the next generation of AI-powered precision agriculture.

REFERENCES

[1] H. Habaragamuwa, Y. Oishi, and K. Tanaka, “Achieving Explainability for Plant Disease Classification with Disentangled VariationalAutoencoders,” *ArXiv*, vol. abs/2102.03082, 2021, doi: 10.1016/j.engappai.2024.107982.

[2] A. Afrin, Md. R. Haque, and Md. Al Mamun, “Enhancing Hyperspectral Image Compression Through Stacked Autoencoder Approach,” *2024 6th International Conference on Electrical Engineering and Information & Communication Technology (ICEEICT)*, pp. 1372–1377, 2024, doi: 10.1109/ICEEICT62016.2024.10534540.

[3] Z. Lin, Y. Chen, X. Zhao, and G. Wang, “Spectral-spatial classification of hyperspectral image using autoencoders,” *2013 9th International Conference on Information, Communications & Signal Processing*, pp. 1–5, 2013, doi: 10.1109/ICICS.2013.6782778.

[4] X.-T. Fan, J. Chen, and Y. Qian, “Graph Regularized Autoencoder Based Feature Extraction for Hyperspectral Image Classification,” *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, pp. 2166–2169, 2021, doi: 10.1109/IGARSS47720.2021.9554068.

[5] C. Dong, T. Xue, and C. Wang, “The Feature Representation Ability of VariationalAutoEncoder,” *2018 IEEE Third International Conference on Data Science in Cyberspace (DSC)*, pp. 680–684, 2018, doi: 10.1109/DSC.2018.00108.

[6] D. Li, X. Liu, Y. Shao, Y. Sun, S. Cheng, and J. Liu, “Autoencoder-based OFDM for Agricultural Image Transmission,” *2022 Tenth International Conference on Advanced Cloud and Big Data (CBD)*, pp. 157–162, 2022, doi: 10.1109/CBD58033.2022.00036.

[7] A. Chefrour and S. Drissi, “K-CAE: Image Classification Using Convolutional AutoEncoder Pre-Training and K-means Clustering,” *Informatica (Slovenia)*, vol. 47, 2023, doi: 10.31449/inf.v47i7.4499.

[8] S. Koda, F. Melgani, and R. Nishii, “Unsupervised Spectral–Spatial Feature Extraction With Generalized Autoencoder for Hyperspectral Imagery,” *IEEE Geoscience and Remote Sensing Letters*, vol. 17, pp. 469–473, 2020, doi: 10.1109/LGRS.2019.2921225.

[9] S. Jin*et al.*, “Study on lightweight rice blast detection method based on improved YOLOv8.,” *Pest ManagSci*, 2025, doi: 10.1002/ps.8790.

[10] G. Tan *et al.*, “Real-time detection method of intelligent classification and defect of transmission line insulator based on LightWeight-YOLOv8n network,” *J. Real Time Image Process.*, vol. 22, p. 53, 2025, doi: 10.1007/s11554-025-01627-9.

[11] H. Chen, F. Zhang, C. Guo, J. Yi, and X. Ma, “SA-SRYOLOv8: A Research on Star Anise Variety Recognition Based on a Lightweight Cascaded Neural Network and Diversified Fusion Dataset,” *Agronomy*, 2024, doi: 10.3390/agronomy14102211.

[12] M. Jiang, Z. Si, K. Yang, and Y. Zhang, “Improved YOLOv8-Based Lightweight Object Detection on Drone Images,” pp. 426–434, 2024, doi: 10.1007/978-981-97-5600-1_37.

[13] Y. Yu *et al.*, "LP-YOLO: A Lightweight Object Detection Network Regarding Insect Pests for Mobile Terminal Devices Based on Improved YOLOv8," *Agriculture*, 2024, doi: 10.3390/agriculture14081420.

[14] J. Huang *et al.*, "A Lightweight Model for Weed Detection Based on the Improved YOLOv8s Network in Maize Fields," *Agronomy*, 2024, doi: 10.3390/agronomy14123062.

[15] J. Wang, M. Li, C. Han, and X. Guo, "YOLOv8-RCAA: A Lightweight and High-Performance Network for Tea Leaf Disease Detection," *Agriculture*, 2024, doi: 10.3390/agriculture14081240.

[16] Y. Wang, Z. Zhang, L. Feng, Y. Ma, and Q. Du, "A new attention-based CNN approach for crop mapping using time series Sentinel-2 images," *Comput. Electron. Agric.*, vol. 184, p. 106090, 2021, doi: 10.1016/j.compag.2021.106090.

[17] S. T. Seydi, M. Amani, and A. Ghorbanian, "A Dual Attention Convolutional Neural Network for Crop Classification Using Time-Series Sentinel-2 Imagery," *Remote. Sens.*, vol. 14, p. 498, 2022, doi: 10.3390/rs14030498.

[18] S. P. Khabusi, P. Pheroijam, and S. Kshetrimayum, "Attention-Based Approach for Cassava Leaf Disease Classification in Agriculture," *2023 International Conference on Energy, Power, Environment, Control, and Computing (ICEPECC)*, pp. 1–6, 2023, doi: 10.1109/ICEPECC57281.2023.10209444.

[19] W. Zeng and M. He, "Rice disease segmentation method based on CBAM-CARAFE-DeepLabv3+," *Crop Protection*, 2024, doi: 10.1016/j.cropro.2024.106665.

[20] A. Dheeraj and S. Chand, "Deep learning based weed classification in corn using improved attention mechanism empowered by Explainable AI techniques," *Crop Protection*, 2025, doi: 10.1016/j.cropro.2024.107058.

[21] X. Wang and W. Cao, "Bit-Plane and Correlation Spatial Attention Modules for Plant Disease Classification," *IEEE Access*, vol. 11, pp. 93852–93863, 2023, doi: 10.1109/ACCESS.2023.3309925.

[22] Yun Zhao, Jiagui Chen, Xing Xu, Jingsheng Lei, and Wujie Zhou, "SEV-Net: Residual network embedded with attention mechanism for plant disease severity detection," *ConcurrComput*, vol. 33, 2021.

[23] H. Guo, Z. Guo, and Z. Pan, "GSCAM: Global Spatial Coordinate Attention Module for Fine-Grained Image Recognition," pp. 431–442, 2021, doi: 10.1007/978-3-030-87355-4_36.

[24] M. U. Alam, J. Hollm'en, J. R. Baldvinsson, and R. Rahmani, "SHAMSUL: Systematic Holistic Analysis to investigate Medical Significance Utilizing Local interpretability methods in deep learning for chest radiography pathology prediction," *Nordic Machine Intelligence*, 2023, doi: 10.5617/nmi.10471.

[25] A.-K. Balve and P. Hendrix, "Interpretable breast cancer classification using CNNs on mammographic images," *ArXiv*, vol. abs/2408.13154, 2024, doi: 10.48550/arXiv.2408.13154.

[26] A. A. Ali, "Interpretable Deep Learning Framework for COVID-19 Detection: Grad-CAM Integration with Pre-trained CNN Models on Chest X-Ray Images," *Int J Sci Res SciEngTechnol*, 2025, doi: 10.32628/ijsrset25121158.

[27] E. Mahamud, N. Fahad, M. Assaduzzaman, S. M. Zain, K. O. M. Goh, and Md. K. Morol, "An explainable artificial intelligence model for multiple lung diseases classification from chest X-ray images using fine-tuned transfer learning," *Decision Analytics Journal*, 2024, doi: 10.1016/j.dajour.2024.100499.

[28] S. Verma, "Development of Interpretable Machine Learning Models to Detect Arrhythmia based on ECG Data," *ArXiv*, vol. abs/2205.02803, 2022, doi: 10.48550/arXiv.2205.02803.

[29] W. Zeng, L. Shan, C. Yuan, and S. Du, "Advancing cardiac diagnostics: Exceptional accuracy in abnormal ECG signal classification with cascading deep learning and explainability analysis," *Appl. Soft Comput.*, vol. 165, p. 112056, 2024, doi: 10.1016/j.asoc.2024.112056.

[30] A. Kokate, "A study of interpretability mechanisms for deep networks," 2018.

[31] Y. Shen and X. Huang, "A Comparative Visualization Analysis of Neural Network Models Using Grad-CAM," *Science and Technology of Engineering, Chemistry and Environmental Protection*, 2024, doi: 10.61173/yzp9wt79.