IJIAR -International Journal of Innovations in Agriculture Research
Volume 2, Issue 1 | May 2025
e-ISSN: 2998-8357 Print ISSN: 2998-8349

CNN Applications Using Multimodal Data in Smart Agriculture

Dr. Thomas Müller[1], Dr. Lena Fischer[2]
[1]Institute of Crop Science, University of Hohenheim, Stuttgart, Germany
[2]Institute of Crop Science, University of Hohenheim, Stuttgart, Germany

Received: 17-03-2025; Revised: 10-04-2025; Accepted: 25-04-2025; Published: 22-05-2025

Abstract

Abstract
The utilization of Convolutional Neural Networks (CNNs) in smart farming has revolutionized traditional farming with its capability in making accurate, data-based decisions. The advent of multimodal sources of data RGB images, hyperspectral images, thermal sensors, LiDAR, as well as IoT-based agricultural sensors has made CNNs highly effective in automating as well as improving multiple agricultural activities. The following is a discussion on multiple usages of CNN in smart farming with emphasis on some important areas that encompass detection of crop diseases, detection of weeds, prediction of yield, monitoring soil as well as plant health, as well as precision irrigation. The discussion also explains improving CNN-based models' precision as well as stability with data fusion strategies that mitigate environmental variability as well as sensor-based limitations.
While remarkable progress has been achieved, some limitations are placed on the widespread utilization of CNNs in agriculture, some of which are data scarcity, model generalizability, computational complexity, as well as limitations in real-time deployments. In addition, multimodal data fusion requires efficient preprocessing, synchronization, as well as feature extraction strategies in order to allow sound decisions. Recent advancements suggest that the addition of CNN with transformers, GANs, edge computing, as well as XAI can also lead towards enhanced model functionality as well as usability in real-world farming.
This paper gives a full discussion on state-of-the-art CNN models that are currently implemented in farming, multimodal data utilization in improving model performances, as well as challenges that are currently experienced and research directions in the future. With CNN-based multimodal learning, smart farming can be optimized with little wastage of resources as well as increased harvest output, making AI-based farming a scalable as well as a feasible possibility.

Keywords:***Convolutional Neural Networks (CNNs), Intelligent Farming, Multimodal Data, Precision Farming, Deep Learning, Detection of Crop Diseases, Detection of Weeds, Yield Estimation, Data Fusion, Internet of Things in Farming, Remote Sensing, Farming with Computer Vision***

1.Introduction

The global agricultural sector is experiencing a revolutionary shift with the addition of Artificial Intelligence (AI) and deep learning approaches. The traditional farming methods that are labor-intensive as well as experience-based decisions are increasingly giving space to data-based, autonomous, as well as precise farming systems. Among all AI approaches, Convolutional Neural Networks (CNNs) have emerged as a powerful tool in smart farming, primarily because they are effective in analyzing as well as interpreting complex vision-based data. CNNs are very effective in handling multimodal data a combination of multiple sources of data in forms like RGB images, hyperspectral as well as multispectral images, thermal sensing, LiDAR, as well as sensor data from IoT devices which provides a more comprehensive picture about agricultural scenarios(1).

With more development in precision farming, CNNs are greatly utilized in critical farming functions that entail detecting plant growth, detecting weeds, detecting crop diseases, predicting yield, estimation of soil quality, as well as precision irrigation. All these functions entail handling varied input data that are gathered from autonomous farming machinery, ground sensors, satellite images, as well as drones. The feature extraction ability in CNNs from high-volume, high-dimensional data makes it highly suitable in both multimodal as well as image-based analysis in farming contexts. For instance, hyperspectral as well as multispectral sensing can reveal latent signs of plant stress that are not observable with a human eye, while thermal sensing can provide insights into moisture in soil as well as in plants. The combination of these varied sources can greatly boost precision farming decisions with minimal human input as well as optimized utilization of resources.

While having these strengths, implementing CNN-based solutions in field-based agricultural scenarios poses a variety of challenges. Unavailability of sufficient training data with appropriate labels, environmental fluctuations, computational limitations, as well as lack of generalizability across varying crop species and climatic regimes are some key challenges that have to be met. Furthermore, incorporating edge computing, real-time computing, as well as explainable AI (XAI) is becoming crucial in order to achieve scalability as well as reliability in CNN-based smart farming solutions. Recent developments in transfer learning, generative adversarial networks (GANs), as well as hybrid AI models are some directions that are likely to address these limitations, leading towards more accurate, dynamic, as well as more interpretable AI-based solutions in agriculture(2).

This paper gives a full exploration of CNN-based smart farming with multimodal data, elaborating on varying CNN models, key fields of application, methods of data fusion, challenges, as well as directions in research in the future. Leveraging strengths in both deep learning as well as multimodal data fusion, smart farming can boost food security, improve utilization of resources, as well as facilitate farming that is sustainable in the future.

2.Literature Survey

Survey of Literature

In recent years, the application of Convolutional Neural Networks (CNNs) in smart farming has gained a lot of attention, primarily due to increased sources of multimodal data in terms of RGB images, hyperspectral and multispectral images, thermal sensors, LiDAR, and IoT-based farming sensors. Numerous studies have proved that CNNs have the ability to convert traditional farming into a more autonomous, reliable, and data-based form. The following section presents a broad overview of research work in this domain with a focus on important developments, methodologies, and challenges in the application of CNN in monitoring crop health, detecting plant diseases, identifying weeds, estimating yield, and precision farming.

CNN-Based Crop Disease Detection

The most thoroughly researched application of CNN in farming is crop disease detection. Scientists have made effective use of CNN models in detecting plant diseases. Mohanty et al. (2016) began with the application of AlexNet as well as GoogLeNet in differentiating 26 plant disease categories from the PlantVillage database with over 99% accuracy. But in their work, they primarily made extensive use of images in RGB format, which is inappropriate in field-based scenarios. To address that, Ferentinos (2018) suggested a CNN-based approach that made use of field-acquired images from different environmental scenarios, leading to increased robustness. Furthermore, Barbedo (2019) experimented with hyperspectral images in combination with CNNs, which established that spectral features have a much greater role in improving disease classification accuracy over traditional RGB images.

Zhang et al. (2021), in a multimodal strategy, combined thermal and hyperspectral data with CNN models in order to enhance disease detection accuracy(3). The research identified that incorporating leaf thermal variations with spectral reflectance data made early detection of disease possible before symptoms could be seen. In more recent work, models have also utilized transformer-based models, with Doshi et al. (2023) coupling CNN with Vision Transformers (ViTs) to enhance feature extraction and disease classification, outperforming traditional CNN models.

Weed Detection and Crop Classification

CNNs have also proved highly effective in detecting weeds as well as categorizing crops, which not only allows targeted herbicide application but also minimizes excessive chemical use. Sa et al. (2017) suggested a CNN-based model for detecting weeds from drone images with much-improved weed segmentation compared to traditional image-processing methods. Similarly, Dyrmann et al. (2016) proposed a deep CNN model trained on multimodal data (RGB + hyperspectral + near-infrared) to classify crop and weed species. Their experiments established that multimodal methods outperform single-modal classification models, particularly in distinguishing visually similar weed species.

Doshi et al. (2022) recently used fusion-based CNN models combining RGB and thermal imaging to improve detection accuracy under varying lighting conditions. Yu et al. (2023) took it a step further by integrating LiDAR and hyperspectral inputs into CNN models, ensuring precise weed localization even in overlapping vegetation scenarios. These studies highlight the shift towards multimodal CNN models, which enable more scalable and reliable solutions for automated weed management.

Yield Prediction and Precision Farming

Yield prediction is another field in which CNNs have made a significant difference. Traditional prediction models relied on statistical and empirical approaches, which often lacked accuracy due to the complexity of environmental interactions. Kamilaris and Prenafeta-Boldú (2018) demonstrated that CNN models trained on satellite images and

Volume 2, Issue 1 | May 2025
e-ISSN: 2998-8357 Print ISSN: 2998-8349

climatic data can accurately predict crop yield at a regional level. Their research was one of the earliest to incorporate spatial and climatic inputs into CNN-based yield prediction models.

Further advancements came from Jin et al. (2020), who developed a CNN-based prediction model with multimodal data fusion from hyperspectral imaging, LiDAR, weather sensors, and soil nutrient analysis. This model significantly improved yield prediction accuracy by capturing both visual indicators of crop health and environmental factors. In another study, Chen et al. (2022) introduced a CNN-LSTM hybrid model that combined CNN feature extraction with LSTM's ability to process temporal dependencies, enabling more accurate seasonal yield forecasting.

Soil and Plant Health Monitoring

CNNs have also been explored for soil and plant health monitoring. Sharma et al. (2021) proposed a CNN model for soil texture classification using multispectral and hyperspectral data, achieving higher classification accuracy compared to traditional machine learning methods. Additionally, Liu et al. (2022) developed a multimodal CNN model integrating soil moisture sensors, thermal imaging, and hyperspectral data, effectively assessing plant nutrient deficiencies and optimizing precision fertilization strategies.

3.Existing and Proposed System

Conventional farming practices heavily depend on human labor, expert knowledge, and traditional machine learning techniques for plant monitoring, disease diagnosis, and yield forecasting. While models such as Support Vector Machines (SVMs), Random Forests, and Decision Trees have been applied in precision farming, they require extensive feature engineering and often struggle to adapt across different environmental conditions.

Most existing deep learning-based methods rely on single-modal data processing using CNNs. Early models were predominantly trained on RGB images for plant disease classification, weed detection, and crop yield estimation. While CNNs demonstrated high accuracy in controlled environments, their performance significantly declined in real-world agricultural settings due to challenges such as varying lighting conditions, occlusions, and diverse crop types.

Additionally, many existing systems lack multimodal data integration, limiting their ability to leverage hyperspectral imaging, thermal sensors, LiDAR, and IoT-based soil and weather data(4). The absence of data fusion techniques in conventional CNN-based models leads to reduced accuracy, poor adaptability, and unreliable decision-making in diverse farming environments. Furthermore, traditional CNN architectures are computationally expensive, making real-time deployment on edge devices and drones difficult.

Proposed System

To overcome the limitations of existing CNN-based agricultural solutions, a multimodal CNN framework is proposed, integrating multiple data sources:

- RGB images for visual crop analysis.
- Hyperspectral and multispectral imaging for plant health and disease detection.
- Thermal imaging to assess plant hydration and stress levels.
- LiDAR imaging for plant canopy structure analysis.
- IoT sensor data (temperature, humidity, soil moisture) for real-time precision farming insights.

The proposed system combines these multimodal data sources using advanced CNN architectures such as ResNet, EfficientNet, and Vision Transformers (ViTs) to extract meaningful patterns from heterogeneous agricultural datasets. By incorporating attention mechanisms and feature fusion networks, the model can integrate complementary information from multiple modalities, significantly enhancing accuracy in crop disease detection, yield prediction, and soil analysis.

Additionally, the system addresses computational constraints by implementing lightweight CNN models optimized for edge computing, enabling real-time processing on drones, IoT devices, and agricultural robots. Furthermore, the integration of Explainable AI (XAI) techniques ensures that decisions are interpretable and transparent, fostering trust among farmers and agronomists.

By leveraging multimodal data fusion, cutting-edge CNN architectures, and real-time edge computing, the proposed system aims to enhance precision farming, reduce resource wastage, improve crop productivity, and make AI-driven farming more accessible and sustainable.

4. CNN applications in smart agriculture using multimodal data

Role of CNNs in Smart Agriculture

1.Introduction to CNNs in Smart Agriculture

Convolutional Neural Networks (CNNs) are highly effective for computer vision, pattern recognition, and image classification, making them a powerful tool for smart agriculture. The agricultural sector is undergoing a digital transformation, with increasing reliance on automated monitoring, precision farming, and AI-driven decision-making. CNNs play a crucial role in processing vast amounts of multimodal agricultural data, such as drone imagery, satellite images, IoT sensor data, and hyperspectral imaging, providing valuable insights into crop health, soil quality, weed detection, and yield prediction. Unlike traditional machine learning techniques that require extensive manual feature extraction, CNNs automatically learn hierarchical features, making them well-suited for complex agricultural environments(5).

2.CNN-Based Crop Disease Detection

One of the most impactful applications of CNNs in smart agriculture is crop disease detection. If plant diseases are not identified early, they can cause significant economic losses. Traditional disease detection methods rely on manual scouting, which is labor-intensive and prone to human error. CNNs have revolutionized plant disease identification by analyzing leaf patterns, color variations, and texture anomalies to classify infected crops with high accuracy.

Research has shown that CNN architectures like AlexNet, VGGNet, ResNet, and EfficientNet can classify plant diseases with accuracy rates exceeding 95% when trained on large datasets such as PlantVillage. More advanced CNN models now incorporate hyperspectral and thermal imaging, allowing early disease detection even before visible symptoms appear, significantly improving intervention strategies and disease management in smart farming.

3.Precision Weed Detection and Management

Weeds can severely impact crop yields by competing for essential resources such as nutrients, water, and sunlight. Conventional weed control methods, including manual weeding and excessive herbicide application, are often inefficient and environmentally harmful(6).

CNNs enable precision weed detection by distinguishing between crops and weeds based on shape, texture, and spectral properties. Drone-based CNN models analyze high-resolution images to identify and localize weeds, enabling targeted herbicide application, which reduces chemical use and promotes sustainable farming.

Recent advancements in CNN-based weed classification incorporate multispectral and hyperspectral imaging, improving detection accuracy even in complex field conditions with overlapping vegetation.

4.Yield Estimation and Crop Growth Monitoring

Accurate yield prediction is crucial for agricultural planning, resource allocation, and supply chain management. Traditional yield estimation models relied on historical yield records, climate data, and statistical regression models, which often lacked precision.

CNNs, when trained on multimodal data sources such as satellite imagery, soil quality measurements, and weather data, can predict crop yield with greater accuracy and reliability. Advanced CNN models integrate Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks to analyze both spatial and temporal trends in agricultural data, enabling real-time yield forecasting.

The ability of CNNs to process high-resolution remote sensing data makes them particularly useful for large-scale precision farming applications.

5. Soil and Plant Health Analysis

Maintaining soil fertility and plant health is essential for sustainable agriculture. CNNs are widely used for soil texture classification, nutrient deficiency detection, and moisture level estimation using hyperspectral imaging and IoT sensor data(7).

Deep learning models trained on multimodal inputs, including thermal imaging and LiDAR, help assess plant stress, water deficiencies, and nutrient imbalances with high precision. CNN-based soil analysis tools provide real-time recommendations for optimized irrigation and fertilization, helping farmers enhance productivity while reducing resource wastage.

6. Livestock and Farm Monitoring

Beyond crop monitoring, CNNs are also transforming livestock management by analyzing animal behavior, health, and movement patterns. CNN-powered vision systems installed in dairy farms, poultry houses, and cattle farms help detect diseases, injuries, and abnormal behavior, allowing for early intervention and improved livestock care.

By integrating thermal cameras and infrared sensors, CNN models can identify fever symptoms, respiratory issues, and hydration levels in livestock. This not only reduces mortality rates but also improves overall farm efficiency and livestock well-being.

5.Multimodal Data Integration in Smart Agriculture

1. Introduction to Multimodal Data Integration

The current frontier in smart agriculture utilizes multimodal data integration to unite multiple agricultural datasets which delivers better all-around information about farming status. The former agricultural analytics system extracted information through a dedicated data source such as RGB images or weather measurements but it proved inefficient because of incomplete contextual information. Modern Convolutional Neural Networks (CNNs) along with deep learning methods have transformed precision agriculture through their ability to combine RGB, hyperspectral, thermal data and LiDAR and IoT sensor data. Multimodal learning through the combination of spatial along with spectral and thermal and environmental data results in better crop health diagnosis and disease recognition and weed identification and yield estimation accuracy(8).

Smart agricultural landscapes produce extensive information from both geospatial information systems and unmanned aerial vehicles (UAVs) and ground sensors and meteorological stations. The information that each data source provides remains distinct from the others.

The visible crop features including leaf shapes and colors become visible in RGB image acquisition.

The detection capabilities of hyperspectral imaging work through the analysis of small spectral shifts which show signs of plant health problems and sickness.

The technology measures water stress conditions and transpiration processes of plants through thermal imaging sensors.

The 3D crop canopy details and biomass evaluation process depends on LiDAR (Light Detection and Ranging) sensor technology.

The real-time environmental insights are provided through IoT sensor measurements of soil moisture together with temperature ecology and humidity monitoring plus nutrient analysis ability.

A combination of this wide array of data sources analyzed through CNN-based systems enables all-inclusive choices that lead to better crop yields as well as disease management and more efficient resource distribution in farming.

2. CNNs for Multimodal Data Fusion

Premade architectures based on CNNs excel at multimodal data integration because they effectively extract complex patterns in spatial, spectral and temporal data from large datasets. Two main strategies exist for multimodal data fusion while different methods are used to develop these frameworks.

CNN receives raw data from different modalities as a joint feature set through early fusion before its initial processing stage. Using tensors that combine RGB, hyperspectral and thermal bands permits the CNN model to discover intermodal feature associations since its inception.

This method of fusion obtains separate features from each modality through independent CNN branches before uniting them within middle-layer connections. Such an approach enables the system to maintain specific modal-based information along with advanced feature schemata(9).

The final predictions from independent CNN pipelines belonging to each modality combine through ensemble techniques or voting methods in late fusion approaches. Different data sources with varying resolution and acquisition periods can be processed successfully through this method.

Such integration of different data forms provides stronger capabilities to find diseases in their initial stages. ABC AI models trained solely on RGB picture data can fail to discover developing diseases yet hyperspectral and thermal picture systems show advanced spectra and temperature patterns which alert about sicknesses before they become visible. Multimodal CNN models elevate the classification accuracy by 15-20 percent above single-modal approaches which makes them essential tools for disease control and early intervention.

The system detects weeds by using this technology alongside exact herbicide application techniques.

Multimodal CNN models have transformed weed detection capabilities particularly for environments containing crops and weeds that have challenging visual similarities. Crop and weed structures become easier to distinguish through the joint application of LiDAR depth data and RGB and multispectral image observation. Since multimodal sensors

are attached to drones these devices become able to examine extensive field spaces, find weed-infested regions and activate precise herbicide employment thus minimizing chemical consumption while conserving environmental quality.
Crop Growth Analysis together with Yield Prediction are possible through integration of certain sensor data.
CNN models that work with multiple data sources enhance the prediction process for agricultural yields because this field previously relied solely on historical records and climate models. CNN architectures successfully predict accurate yield outputs by processing satellite and UAV and moisture sensor and meteorological information simultaneously. Global yields become more precise by ninety percent due to the implementation of hyperspectral imaging system with Lidar aerial scanning and IoT sensors which work together to analyze plant health status and canopy structure and soil quality conditions.

3. Challenges in Multimodal Data Integration

Multiple obstacles limit the adoption of CNN-based smart agriculture where multiple data sources are integrated.
The utilization of CNN models becomes difficult due to inconsistent spatial resolutions and differing time intervals as well as formats among different modal data.
Processing high-resolution hyperspectral alongside LiDAR and multispectral data becomes difficult through real-time applications because these requires extensive computing power and big storage capacity within low-resource environments.
The acquisition of labeled multimodal datasets incorporating data collection requirements both heavy financial resources and significant personnel involvement. Research groups utilize self-supervised learning together with data augmentation methods as solutions to minimize this issue.
CNN models that use multimodal data from one particular geographical area struggle to adapt to new regions because of climate variations together with variations in soil structure and crop types. Scientists focus on developing transfer learning along with domain adaptation algorithms to assure better inter-region adaptability.

4. Future Prospects and Research Directions

The application of CNN-based multimodal data integration in smart agriculture shows great promise because researchers present the following approaches:
Similar to real-time processing the system implements lightweight CNN models on both edge devices and drones which enable them to process multimodal agricultural data in real time.
XAI enhances CNN interpretability to allow farmers to understand the predictions made by models in addition to their decision-making processes.
Fusion with Transformer-Based Models: Integrating Vision Transformers (ViTs) and CNNs for better contextual feature extraction in multimodal agricultural applications.
GANs enable the generation of synthetic multimodal datasets through this approach to handle data scarcity problems.

6.CNN Architectures for Agricultural Tasks

1. Introduction to CNN Architectures in Agriculture

The basic architecture of Convolutional Neural Networks (CNNs) operates as the main technical component for computer vision applications which execute automated detection of crop diseases and weed identification along with yield forecasting and analysis of soil health and an oversight of livestock. CNNs prove advantageous in agricultural tasks because they analyze and extract complex patterns found in extensive collections of unstructured images. Sequential training of spatial features by convolutional layers and pooling layers and full connected layers enables CNNs to analyze multiple varieties of data from drones along with satellites, IoT sensors and hyperspectral cameras and LiDAR systems.
Several CNN architectures have been optimized for agricultural work because they provide distinct benefits through accuracy and computational performance along with easy scalability and environmental resistance. This part discusses popular CNN architectural frameworks utilized in agricultural work which improves both precision farming systems and sustainable agricultural practices.

2. Classical CNN Architectures in Agriculture

LeNet-5: The Foundation of CNNs

Yann LeCun introduced LeNet-5 as an early CNN architectural design for image classification purposes in 1998. LeNet-5 originated as a digit recognition system but present-day applications utilize it for identifying different agricultural specimens including fruits along with leaves and grains. The architectural limitations prevent LeNet-5

e-ISSN: 2998-8357 Print ISSN: 2998-8349

from processing complex agricultural datasets because it cannot effectively deal with large high-resolution images and multiple input modalities.

AlexNet: The Breakthrough in Deep Learning

The breakthrough in deep learning occurred with AlexNet when Krizhevsky et al. presented an innovation which delivered exceptional results in the ImageNet challenge in 2012. AlexNet brought advanced image classification through its five convolutional layers and three fully connected layers which are actively used for both leaf classification and crop disease identification. The success of AlexNet comes at the expense of extensive computing demands which hinder its effectiveness when used on real-time farm equipment operating at the edge.

VGGNet: Deep but Computationally Expensive

VGGNet presented a deep architecture in VGG16 and VGG19 while implementing smaller convolutional filters of dimension 3×3 through its development by Simonyan and Zisserman (2014). The model displays excellent performance in disease detection and weed recognition because it produces detailed hierarchical feature structures. Huge computational expenses caused by its numerous parameters restrict the use of VGGNet for resource-limited agricultural applications on field-based edge systems.

3. Modern CNN Architectures for Smart Agriculture

ResNet: Solving the Vanishing Gradient Problem

He et al. (2016) presented Residual Networks (ResNet) through skip connections that enabled deeper networks to train without the vanishing gradient issue. The precision agriculture sector implements ResNet-18, ResNet-50 and ResNet-101 alongside various other system variants throughout their operations.

An analysis of crop diseases employs hyperspectral and thermal data as multiple data types to achieve combined disease classification.

Crops need segmentation treatment within actual agricultural fields for weed detection.

The analysis of soil texture depends on remote sensing image data.

ResNet-based architectures demonstrate enhanced robustness features while extracting complex information from extensive agricultural datasets which solidifies them as prime solutions for such challenging datasets.

InceptionNet: Multi-Scale Feature Extraction

Szegedy et al. (2015) developed InceptionNet through the creation of Inception modules which enable the network to analyze features simultaneously at various scales. Remote sensing applications in agriculture find particular benefit from this approach since leaf texture along with plant shape and disease symptoms present at different spatial scales. The application of InceptionNet contributes to agricultural tasks through the following usages:

Pest detection in drone-captured images.

Crop classification using hyperspectral imaging.

Agricultural livestock observation occurs through a combination of thermal along with RGB sensing methods.

The agricultural sector prefers using InceptionNet instead of traditional CNNs because it maintains simultaneous extraction of global and local features.

EfficientNet: Balancing Performance and Efficiency

The research team of Tan and Le (2019) developed EfficientNet through their innovative approach to scale all three aspects of depth width and resolution in a single step. The combination of better accuracy and fewer parameters in EfficientNet makes it an excellent choice for real-time edge-based agricultural work which includes:

Real-time plant disease detection on mobile devices.

Automatic weed classification forms a critical part of autonomous robotic weeders.

Predictions about crop yield levels require analysis of satellite as well as drone collected imagery.

EfficientNet demonstrates betterprecision rates alongside reduced processing requirements thus becoming appropriate for IoT devices with low power consumption and robotic farming systems with edge-based artificial intelligence.

4. Advanced CNN Architectures for Multimodal Agricultural Data

Hybrid CNN-LSTM for Temporal Crop Analysis

çünkü tarım verisi genellikle mekanik sırasında sürekli veriler bulunan (tarım gelişim aşamları ve iklim eğilimleri) nedenle LSTMs ile ilişkilendirilen sınırlıCHEDULED alanı (CNN) yapılardır. The implementation of CNN-LSTM models takes place in these domains:

Agricultural yield prediction makes use of historical satellite photographs alongside climate information for evaluation.

Perfected irrigation control depends on analyzing soil moisture evolutions.

The evaluation of early stress identification depends on measuring vegetation indices through time periods.

The combination of CNN and LSTM methods works best to track settings where both image recognition and pattern sequence analysis is needed.

Vision Transformers (ViTs) for Agricultural Image Processing

Vision Transformers (ViTs) became popular in computer vision because they develop self-attention mechanisms that allow models to establish distant relational connections. The application of ViT models has benefited smart agriculture through the following purposes:

Fine-grained plant disease classification.

Crop segmentation in satellite imagery.

Machine automation performs quality evaluations of post-harvested agricultural produce.

The global contextual understanding capabilities of ViTs exceed those of CNNs thus making them suitable for processing extensive remote sensing data collections. The processing requirements of these models extend beyond typical farm operations which slows down their ability to operate in real time through lower-resource agricultural regions.

7.Conclusion and Future work

The detection of crop diseases and weed categorization along with yield predictions and soil health monitoring and livestock surveillance have been successfully achieved by CNNs which operate on multimodal data types that include RGB pictures, hyperspectral information, thermal data, LiDAR data and IoT sensor outputs. Distinct deep learning models have enhanced modern agricultural practices by enabling better precision alongside increased efficiency as well as sustainability which leads to optimal resource management for farmers.

Various obstacles still face deep learning solutions including restricted data availability and limited computing capacity and the need for universal systems across environmental areas and clear understanding of processed information. Numerous CNN models need substantial processing power which prevents their deployment for real-time processing on edge devices and drones and also prevents their utilization in mobile applications. CNN models operate as black-boxed systems which present challenges for farmers and agronomists when trying to understand confident decision-making processes. The insufficient availability of big labeled agricultural data sets makes it difficult to develop robust generalized models for training.

The complete adoption of CNNs for smart agriculture requires scientists to create deep learning models which perform efficiently with high scalability and allow interpretation from end-users. Smart farming will advance through solutions featuring low-caliber CNN architecture with self-supervised learning together with edge AI applications running in real-time.

Future Work

The advancement of CNN-based agricultural solutions requires attention to several research areas which will build better effectiveness and real-world usability:

Development of Lightweight CNN Models for Real-Time Processing

Modern CNN system architectures demand intensive computational hardware that makes it complicated to carry out real-time processing within resource-constrained environments. Scientists should dedicate their research to lightweight CNN model development along with pruning algorithms and quantification methods and TinyML applications to fit edge devices and drones and IoT sensors for deployment purposes.

Multimodal Data Fusion for Enhanced Accuracy

Future CNN-based models should connect multimodal agricultural data through aspects such as hyperspectral imaging combined with thermal data and LiDAR as well as IoT sensor readings without interference. Improved fusion approaches between features enable CNNs to recognize advanced data source interactions resulting in better monitor accuracy for crops and soil conditions and irrigation systematicity.

Self-Supervised and Transfer Learning for Overcoming Data Scarcity

The collection of labeled agricultural data remains scarce and expensive to acquire. The next phase of research should center on using self-supervised learning and semi-supervised learning together with domain adaptation approaches because these methods decrease dependence on expertly labeled data without sacrificing generalization abilities for assorted environmental and crop distributions and soil conditions.

Acknowledgement: Nil

Conflicts of interest

The authors have no conflicts of interest to declare

References

1. El Sakka M, Ivanovici M, Chaari L, Mothe J. A Review of CNN Applications in Smart Agriculture Using Multimodal Data. Sensors. 2025;25(2):472.
2. Kamilaris A, Prenafeta-Boldú FX. A Review of the Use of Convolutional Neural Networks in Agriculture. Journal of Agricultural Science. 2018;156(3):312-322.
3. Garg S, Pundir P, Jindal H, Saini H, Garg S. Towards a Multimodal System for Precision Agriculture Using IoT and Machine Learning. arXiv preprint arXiv:2107.04895. 2021.
4. Lottes P, Behley J, Milioto A, Stachniss C. Fully Convolutional Networks with Sequential Information for Robust Crop and Weed Detection in Precision Farming. IEEE Robotics and Automation Letters. 2018;3(4):2870-2877.
5. Smitt C, Halstead M, Ahmadi A, McCool C. Explicitly Incorporating Spatial Information to Recurrent Networks for Agriculture. arXiv preprint arXiv:2206.13406. 2022.
6. Ghanbari A, Shirdel G, Maleki F. Semi-Self-Supervised Domain Adaptation: Developing Deep Learning Models with Limited Annotated Data for Wheat Head Segmentation. arXiv preprint arXiv:2405.07157. 2024.
7. Zabawa L, Tzirakis P, Schuller BW. Multimodal Fusion with Recurrent Neural Networks for Continuous Affect Recognition. Information Fusion. 2020;65:255-264.
8. Zhang Y, Wu T, Li M, Li X. Multimodal Crop Cover Identification Using Deep Learning and Remote Sensing Data. Multimedia Tools and Applications. 2023;82:17140.
9. Zhou J, Chen J, Ye J, et al. Crop Yield Prediction Using Multimodal Meta-Transformer and Temporal Graph Neural Networks. IEEE Transactions on Geoscience and Remote Sensing. 2023;61:1-13.