International Journal of Smart Agriculture (IJSA)

ISSN: 2995-9829 (ONLINE)

VOLUME 4 ISSUE 1 (2026)

PUBLISHED BY
E-PALLI PUBLISHERS, DELAWARE, USA

International Journal of Smart Agriculture (IJSA)

Volume 4 Issue 1, Year 2026
ISSN: 2995-9829 (Online)
DOI: https://doi.org/10.54536/ijsa.v4i1.7582
https://journals.e-palli.com/home/index.php/ijsa

On-Device Edge AI for Precision Agriculture: A Systematic Scoping Review

K.Y.B.S. Fernando[1*], K.D. Thamarasee[2]

Article Information

Received: January 17, 2026

Accepted: March 23, 2026

Published: July 29, 2026

Keywords

Edge AI, Model Optimisation, On-Device Inference, Precision Agriculture, TinyML

ABSTRACT

Edge AI and Tiny Machine Learning (TinyML) have emerged as transformative paradigms for deploying machine learning inference directly on field hardware, addressing the connectivity, latency, and bandwidth constraints that render cloud-dependent systems impractical in real agricultural environments. However, no prior survey systematically maps fully on-device, cloud-independent inference across multiple precision agriculture domains or applies a multi-dimensional evaluation framework to support deployment decision-making. This paper presents a hybrid systematic-scoping review of 41 peer-reviewed and preprint studies published between January 2020 and March 2026, following PRISMA-ScR reporting guidelines. Studies are analysed across four domains (crop disease detection, irrigation and soil monitoring, livestock monitoring, and greenhouse monitoring) using a five-dimensional framework covering technical performance, resource efficiency, economic viability, deployment feasibility, and agricultural impact. Results show that classification accuracy ranges from 92.3% to 99.9% and regression performance reaches R^2 values of 0.85 to 0.99 across hardware spanning microcontrollers to AI accelerators, yet field validation rates vary considerably across domains and economic viability remains critically underreported, with only 7 of 41 studies disclosing hardware costs. This survey contributes a three-tier hardware taxonomy, the five-dimensional evaluation framework, and a structured analysis of systemic challenges and future research directions to advance edge AI from agricultural prototyping toward scalable real-world deployment.

INTRODUCTION

Sustainably feeding a global population projected to reach 9.7 billion by 2050, while managing accelerating climate disruption, water scarcity, and shrinking arable land, demands a fundamental transformation in agricultural monitoring, management, and decision-making (FAO *et al.*, 2024; UN DESA, 2022). In response, researchers and practitioners have developed a wide range of AI-powered precision farming systems that deploy machine learning across crop health monitoring, irrigation control, livestock management, and greenhouse automation, marking a decisive shift from manual observation toward data-driven agricultural intelligence. However, the dominant architecture underpinning these systems relies on cloud computing for data processing and model inference, creating fundamental operational vulnerabilities in rural field environments where unreliable connectivity, high communication latency, and bandwidth constraints render cloud-dependent pipelines impractical for real-time agricultural control (Kalyani & Collier, 2021; Yu *et al.*, 2025). A new paradigm has emerged to address these constraints: Edge AI and Tiny Machine Learning (TinyML) enable optimised machine learning models to execute inference directly on field-deployed hardware, from microcontrollers to single-board computers, without requiring any cloud or internet connectivity, making low-latency, autonomous agricultural intelligence viable at the point of observation.

Despite this rapid evolution, existing survey literature has not addressed on-device inference as a distinct paradigm within precision agriculture. Recent comprehensive reviews of AI and ML in agriculture, such as Kuan *et al.* (2025), evaluate models primarily by predictive accuracy and application breadth, with no systematic consideration of deployment hardware or on-device feasibility. Heydari and Mahmoud (2025) confirm that TinyML delivers measurable latency and power advantages over cloud-based inference, but span multiple domains including healthcare, industrial, and vehicular applications, with agriculture appearing only marginally and without a domain-specific evaluation framework. Madiwal *et al.* (2025) focus exclusively on edge AI in agriculture, yet they are restricted to a single domain and review a mix of cloud-connected and partially autonomous systems without a structured cross-study evaluation framework. Collectively, no prior survey systematically maps fully on-device, cloud-independent machine learning inference across multiple agricultural domains, or applies a multi-dimensional framework capturing technical performance, resource efficiency, economic viability, deployment feasibility, and agricultural impact.

To address these gaps, this survey is guided by the following research questions:

RQ1: What edge hardware platforms and lightweight ML architectures have been deployed for on-device inference in precision agriculture between 2020 and 2026?

RQ2: How do these systems perform when evaluated across five dimensions: technical performance, resource efficiency, economic viability, deployment feasibility, and agricultural impact?

RQ3: What are the primary challenges and open research directions for edge AI in resource-constrained and

[1] K.Y.B.S. Fernando, Robert Gordon University, Garthdee Road, Aberdeen, United Kingdom
[2] Informatics Institute of Technology Sri Lanka, Sri Lanka
[*] Corresponding author's e-mail: yeran.20221647@iit.ac.lk

connectivity-limited farming environments?

This paper presents a hybrid systematic-scoping review of on-device machine learning inference in precision agriculture, drawing on 41 peer-reviewed and preprint studies published between January 2020 and March 2026 and following the PRISMA Extension for Scoping Reviews (PRISMA-ScR) reporting guidelines (Tricco *et al.*, 2018). This survey makes three primary contributions. First, it provides a structured taxonomy of edge hardware platforms and lightweight ML architectures deployed across agricultural settings, organised across three device tiers from microcontrollers to AI accelerators. Second, it introduces and applies a five-dimensional evaluation framework covering technical performance, resource efficiency, economic viability, deployment feasibility, and agricultural impact, enabling systematic cross-study comparison of hardware-model trade-offs across four agricultural domains. Third, it identifies systemic open challenges including model drift, dataset bias, hardware robustness under field conditions, and the absence of standardised benchmarking protocols, and outlines concrete future research directions to advance edge AI from agricultural prototyping toward scalable real-world deployment. The remainder of this paper is structured as follows. Section 2 reviews related surveys and identifies the gap addressed by this work; Section 3 details the hybrid systematic-scoping methodology; Section 4 introduces edge AI foundations including hardware taxonomy and lightweight model architectures; Section 5 analyses included studies by application domain; Section 6 provides cross-domain critical analysis, documents key challenges, and outlines future research directions; and Section 7 concludes the survey.

LITERATURE REVIEW

Existing literature relevant to this survey spans three broad areas: IoT-based smart farming systems, artificial intelligence and machine learning applications in precision agriculture, and edge computing architectures for agricultural deployments. This section reviews the most directly related surveys within each area and identifies the collective gaps that motivate the present work.

IoT and Cloud-Based Smart Farming Surveys

Navarro *et al.* (2020) conducted a systematic review of IoT-based smart farming using PRISMA methodology, examining hardware platforms, communication protocols, and applications across crop monitoring, disease prevention, and irrigation control. The survey provides a structured taxonomy of sensor-based agricultural systems and documents a growing trend toward cloud and big data computing for data processing across the reviewed literature. However, while the survey acknowledges AI-driven data analysis within cloud-hosted pipelines, it does not examine where model inference is executed, whether systems can operate without network connectivity, or how models might be deployed on resource-constrained field hardware. The embedded hardware reviewed in that survey is treated as sensor acquisition and communication infrastructure rather than as inference platforms, with no discussion of lightweight ML models, model optimisation techniques, or on-device inference frameworks.

Building on this foundation, Kalyani and Collier (2021) surveyed the role of cloud, fog, and edge computing combinations in smart agriculture, covering literature from 2015 to 2021, and proposed a new architecture model integrating all three tiers for smart agricultural applications. The survey offers a structured classification of how fog and edge computing introduced non-cloud processing tiers into agricultural systems, reducing dependence on centralised infrastructure, and identifies latency, bandwidth, and real-time analytics limitations of cloud-only approaches as key motivations for bringing computation closer to the field. Nonetheless, the edge and fog devices discussed in that survey serve primarily as data processing nodes within hybrid cloud-fog-edge pipelines rather than as fully autonomous on-device inference platforms. Lightweight ML model optimisation, quantisation, and TinyML frameworks are outside the scope of that work, and no evaluation framework is applied to compare hardware-model trade-offs across agricultural domains.

Artificial Intelligence and Machine Learning in Precision Agriculture

Ugwu *et al.* (2025) conducted a systematic review of AI and ML applications in agriculture, synthesising 95 studies published between 2013 and 2023 across crop monitoring, yield prediction, and resource optimisation. The review provides a broad cross-domain synthesis of algorithmic performance, reporting accuracy rates of up to 93 percent across neural networks, decision trees, and deep learning models. However, the review evaluates models exclusively by predictive accuracy and reported agricultural outcomes, with no consideration of deployment platform, inference hardware, or where model execution takes place. Questions of model size, power consumption, and offline operability are absent from the methodology entirely, limiting the practical relevance of the findings for resource-constrained or connectivity-limited farming environments.

Kuan *et al.* (2025) published a systematic review of machine learning and computer vision in precision agriculture, guided by PRISMA methodology and analysing 213 articles from the past five years across crop monitoring, harvesting, soil management, and water management. The review documents the predominant use of CNNs across agricultural vision tasks and explores the potential of emerging architectures including Vision Transformers and Generative Adversarial Networks, identifying limited access to high-quality datasets and the need for adaptable solutions across diverse agricultural contexts as key open challenges. While the review notes that lightweight models such as MobileNet and YOLO variants are suitable for resource-constrained settings, and acknowledges connectivity limitations in remote agricultural areas, hardware platforms are not used as a systematic evaluation criterion across the reviewed studies. The review therefore does not provide a structured cross-study comparison of hardware-model pairings, resource efficiency, or deployment feasibility, which are precisely the dimensions needed to guide on-device deployment decisions in connectivity-limited farming environments.

Edge Computing and Edge AI for Agriculture

Yu *et al.* (2025) presented a detailed analysis of cloud-edge-device collaborative computing architectures in smart agriculture, covering applications including intelligent irrigation, environmental monitoring, livestock health tracking, and pest detection. The survey situates edge AI within broader hybrid system designs and clearly articulates that unstable rural connectivity, high communication latency, and bandwidth constraints make purely centralised cloud models insufficient for real-time agricultural control. However, the collaborative framework it examines inherently assumes ongoing interaction between cloud, edge, and device tiers, with device-level components assigned to front-line data acquisition and basic preprocessing rather than end-to-end model execution. Fully autonomous on-device inference, where a field-deployed microcontroller or single-board computer operates as a standalone system without any cloud or edge server involvement, lies outside the scope of that survey.

Heydari and Mahmoud (2025) conducted a PRISMA-based systematic review of TinyML and on-device inference across 26 experimental studies, confirming that TinyML delivers measurable latency and power advantages over cloud-based inference in bandwidth-constrained environments. However, smart agriculture appears only at the margins of the review, represented by five isolated examples: greenhouse forecasting, grape leaf disease detection, plant growth monitoring, gas recognition, and UAV-assisted soil moisture prediction. These are grouped under the broader category of 'Ecology'. The review does not provide a systematic mapping or a structured evaluation framework for hardware-model pairings specific to the agricultural sector. Furthermore, the authors acknowledge that because performance metrics were not standardized across the reviewed experiments, making definitive comparisons about deployment feasibility across diverse agricultural domains remains a challenge. Consequently, the multi-domain landscape of precision agriculture is not addressed as a unified field of study within their analysis.

Madiwal *et al.* (2025) conducted a survey specifically focused on edge AI and IoT for real-time crop disease detection, organising their literature review across four primary categories: IoT-edge system architectures, lightweight deep learning models for resource-constrained devices, data-centric methods such as augmentation and incremental learning, and communication protocols with sensor integration. The survey identifies five primary open challenges: scalability, energy efficiency, data sparsity, model generalisation, and secure communication. However, the survey is restricted to a single agricultural domain, namely crop disease detection, and the edge deployments it reviews span a mix of cloud-connected IoT setups and partially autonomous edge systems rather than focusing exclusively on fully on-device inference. No structured multi-dimensional evaluation framework is applied across the reviewed papers to compare hardware-model pairings, resource efficiency, or economic viability.

Synthesis and Identified Gap

Collectively, these seven surveys advance knowledge across IoT-based farming, AI and ML in agriculture, TinyML, and edge computing architectures, yet a clear and consistent gap emerges across all of them. None systematically maps fully on-device, cloud-independent machine learning inference across multiple agricultural domains. Navarro *et al.* (2020) and Kalyani and Collier (2021) treat edge devices as pipeline nodes rather than inference platforms and do not address lightweight model optimisation or TinyML frameworks. Ugwu *et al.* (2025) and Kuan *et al.* (2025) cover model architecture and agricultural application in depth but entirely overlook deployment hardware and on-device feasibility. Yu *et al.* (2025) situate edge AI within collaborative cloud-edge-device systems rather than standalone on-device deployments. Heydari and Mahmoud (2025) confirm TinyML's advantages but treat agriculture only marginally, without addressing precision agriculture as a unified field or applying a domain-specific evaluation framework. Madiwal *et al.* (2025) is restricted to a single agricultural domain and reviews a mix of cloud-connected and partially autonomous systems without a structured cross-study evaluation framework.

This survey directly addresses these gaps by providing a hybrid systematic-scoping review focused exclusively on on-device machine learning inference in precision agriculture, covering multiple domains, applying a five-dimensional evaluation framework, and explicitly mapping hardware-model pairings to support deployment decision-making in resource-constrained farming environments.

MATERIALS AND METHODS

Review Design

This survey adopts a hybrid systematic-scoping review methodology, combining the structured rigour of a systematic review with the broad mapping capability of a scoping review. The systematic component ensures methodological credibility through explicit search protocols, pre-defined inclusion and exclusion criteria, and consistent structured evaluation across all included studies using the five-dimensional framework described in Section 3.5. The scoping component accommodates the breadth of the field by capturing diverse terminology, varied hardware platforms ranging from microcontrollers to AI accelerators, and exploratory implementations across multiple agricultural domains. This hybrid design is particularly appropriate for edge AI in precision agriculture because the field is both technically specialised, requiring structured evaluation, and rapidly evolving, requiring a sufficiently broad lens to map its full landscape without prematurely restricting scope. The review methodology follows the PRISMA Extension for Scoping Reviews (PRISMA-ScR) reporting guidelines to ensure transparency and reproducibility (Tricco *et al.*, 2018).

Search Strategy

A systematic literature search was conducted across five major academic databases: IEEE Xplore, ACM Digital Library, ScienceDirect, Scopus, and arXiv. These databases were selected because they collectively provide the most comprehensive coverage of computer science, electrical engineering, and applied AI research relevant to the topic. In addition to these primary databases, supplementary searches were conducted via Google Scholar to capture relevant studies published in emerging and domain-specific journals. Further papers were identified through backward citation chasing of included studies. The search was restricted to publications between January 2020 and March 2026 to capture the period following the emergence of TinyML as a practical deployment paradigm while ensuring currency of findings.

The following keyword combinations were used across all databases:

1. "TinyML agriculture" OR "TinyML precision farming"
2. "edge AI smart farming" OR "edge AI crop"
3. "on-device machine learning agriculture"
4. "embedded deep learning agriculture"
5. "MobileNet ESP32 agriculture" OR "MobileNet Raspberry Pi crop"
6. "lightweight neural network IoT farming"
7. "quantisation pruning agriculture edge"

Boolean operators (AND, OR) were applied to combine terms across technology and domain categories. Where database-specific filters were available, results were restricted to peer-reviewed journal articles and conference papers published in English.

Inclusion and Exclusion Criteria

A clear set of inclusion and exclusion criteria was defined prior to screening to ensure consistent application by the author and to align the review scope with the research questions.

Inclusion criteria

1. Studies describing machine learning model inference executed directly on edge hardware (microcontrollers, single-board computers, or AI accelerators) without real-time cloud dependency
2. Studies involving at least one real agricultural application domain (crop disease detection, irrigation control, livestock monitoring, greenhouse management, or related domains)
3. Studies reporting at least one measurable performance metric (accuracy, latency, power consumption, or model size)
4. Published between January 2020 and March 2026
5. Written in English and published in peer-reviewed journals, conference proceedings, or preprint repositories

Exclusion criteria:

6. Studies where model inference occurs exclusively on cloud or remote servers
7. Purely theoretical or simulation-based studies without hardware validation
8. Studies using only traditional rule-based or non-machine-learning algorithms
9. Studies that do not report sufficient technical detail to apply the evaluation framework
10. Duplicate publications reporting the same system or dataset

Study Selection Process

The study selection followed a four-stage screening process consistent with PRISMA-ScR guidelines (Tricco *et al.*, 2018). In the first stage, database searches returned an initial pool of 169 records. After removal of duplicates across databases, 132 unique records remained. In the second stage, titles and abstracts were screened against the inclusion and exclusion criteria, reducing the pool to 62 candidate studies. In the third stage, full-text articles were retrieved and assessed in detail, resulting in 38 included studies. In the fourth stage, reference lists of included studies were examined for any additional relevant papers not captured by the database searches, a process known as backward citation chasing, which identified 3 additional studies meeting the inclusion criteria. The final included set therefore comprises 41 studies.

Figure 1 presents the PRISMA-ScR flow diagram illustrating the number of records identified, screened, assessed for eligibility, and included at each stage.

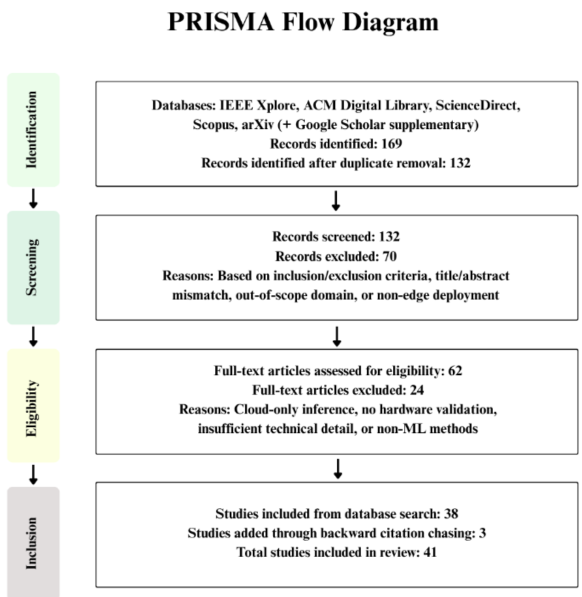

Figure 1: PRISMA-ScR flow diagram depicting the study selection process for the systematic scoping review.

Five-Dimensional Evaluation Framework

To enable structured and consistent cross-study comparison, this survey introduces a five-dimensional (5D) evaluation framework applied uniformly to all included studies. Each dimension captures a distinct aspect of system suitability for real-world agricultural deployment. All metrics are reported in the format used by the original study, and where data for a specific dimension is unavailable, this is noted explicitly rather than assumed.

Dimension 1 - Technical Performance: Evaluates the accuracy of the ML model and inference latency. Higher accuracy and lower latency indicate stronger technical performance.

Dimension 2 - Resource Efficiency: Evaluates the computational and memory demands of the deployed model, including RAM usage, flash memory footprint, model size after optimisation, and power consumption.

Dimension 3 - Economic Viability: Evaluates the approximate hardware cost of the deployment platform in US dollars where reported. Where cost data is unavailable, this is noted explicitly as N/R.

Dimension 4 - Deployment Feasibility: Evaluates the practical requirements for deploying the system in a real field environment, including internet connectivity dependency, ease of installation, and reported robustness to environmental conditions.

Dimension 5 - Agricultural Impact: Evaluates the real-world applicability and reported outcomes of the system in an agricultural context, including whether the system was tested in a real farm environment, whether it addresses a validated agricultural problem, and whether performance metrics are reported under field conditions rather than only in controlled laboratory settings.

Each included study is assessed against all five dimensions where data is available. The framework is applied in Section 5 to analyse individual studies by domain and in Section 6 to support cross-study comparative benchmarking.

Edge AI Foundations

Edge Hardware Taxonomy

Edge devices deployed in agricultural AI systems span a wide range of computational capability, power consumption, and cost. For this survey, hardware platforms are organised into three tiers based on processing architecture and resource profile, as summarised in Table 1.

Table 1: Edge Hardware Taxonomy for On-Device Agricultural AI Deployment

Tier	Category	Representative Devices	Typical Specs	Cost (USD)
Tier 1	Microcontrollers (including MCUs with integrated NPUs)	ESP32, STM32, Arduino Nano 33 BLE	80–400 MHz, 256 KB–8 MB RAM, ~0.15–1 W	<$20
Tier 2	Single-Board Computers	Raspberry Pi 3, 4 and 5	1.2–2.4 GHz ARM, 1–8 GB RAM, 2–7 W	$20–$90
Tier 3	AI Accelerators	NVIDIA Jetson Nano, Google Coral Dev Board	GPU/TPU-accelerated inference, 5–45 W	$80–$2,000

Some deployments augment a Tier 2 host with a peripheral accelerator such as the Google Coral USB Accelerator or Intel Neural Compute Stick 2, which attach via USB rather than operating as standalone platforms. These configurations occupy an intermediate position between Tier 2 and Tier 3.

The choice of hardware tier directly determines which model architectures and optimisation techniques are feasible for a given deployment. Tier 1 devices require the most aggressive compression strategies to fit within kilobyte-level memory budgets, while Tier 3 devices can execute complex multi-modal inference pipelines with minimal optimisation overhead.

Model Architectures for Edge Deployment

Standard deep learning architectures were designed for server-class hardware and are not directly deployable on edge devices due to their large parameter counts and high memory requirements. A class of architectures has been developed specifically to deliver competitive accuracy within the memory and computational budgets of edge hardware.

MobileNet replaces standard convolutions with depthwise separable convolutions, which factorize a full convolution into two cheaper operations, substantially reducing both parameter count and inference cost while retaining classification accuracy suitable for edge deployment.

EfficientNet applies a compound scaling strategy that jointly optimises model depth, width, and input resolution, producing a family of variants that span from editions optimised for MCU-class hardware to standard editions deployable on more capable edge platforms.

YOLO-family models extend the CNN architecture to real-time object detection through single-pass inference, where the model simultaneously predicts bounding boxes and class probabilities across the entire image in one forward pass. Nano and small variants are designed for Tier 2 and Tier 3 edge hardware.

Vision Transformers (ViT) apply self-attention mechanisms across image patches rather than convolutions. While computationally heavier than MobileNet or YOLO-nano, quantised ViT models are deployable in ensemble configurations on Tier 2 hardware.

Model Optimisation Techniques

Deploying models on Tier 1 devices often requires compression beyond architectural design alone. Four principal techniques are used to compress models for edge deployment.

Quantisation reduces the numerical precision of model weights and activations from 32-bit floating point to 8-bit integers or lower, reducing model size and accelerating inference on hardware without floating-point units. Post-training quantisation is applied after training, while Quantisation-Aware Training (QAT) simulates reduced precision during training itself, typically yielding better accuracy retention.

Pruning removes redundant weights, channels, or layers based on their contribution to output accuracy. Structured pruning, including channel pruning and backbone reconstruction, produces architecturally compact models that reduce inference time without requiring specialised hardware support.

Knowledge Distillation trains a smaller student model to replicate the output behaviour of a larger teacher model, allowing the student to operate within edge memory budgets while retaining accuracy learned from the teacher.

Genetic Algorithm (GA) optimisation uses an evolutionary search process to identify optimal model hyperparameters or architecture configurations for a given hardware constraint.

Deployment Frameworks

Several software frameworks bridge the gap between model training environments and resource-constrained edge hardware.

TensorFlow Lite for Microcontrollers (TFLM) is a low-level runtime for executing quantised models on Tier 1 devices with kilobyte-level memory. It requires no operating system or dynamic memory allocation.

Edge Impulse is an end-to-end platform covering data collection, model training, optimisation, and deployment on embedded targets. Its EON Compiler applies additional compression at the deployment stage to reduce RAM and flash footprint beyond standard TFLM output.

ONNX Runtime supports cross-platform deployment from multiple training frameworks including PyTorch and scikit-learn, and is suited to Tier 2 devices where a lightweight operating system is available.

TensorRT is NVIDIA's inference optimisation engine for Jetson-class hardware. It applies layer fusion, precision calibration, and kernel auto-tuning to maximise throughput on GPU-accelerated Tier 3 devices.

Together, these hardware tiers, architectures, optimisation strategies, and deployment frameworks form an interconnected pipeline that governs how an agricultural problem is translated into on-device inference. Figure 2 consolidates these layers into a unified Edge AI deployment stack, illustrating the conditional relationships between hardware capability and the feasibility of each subsequent design decision.

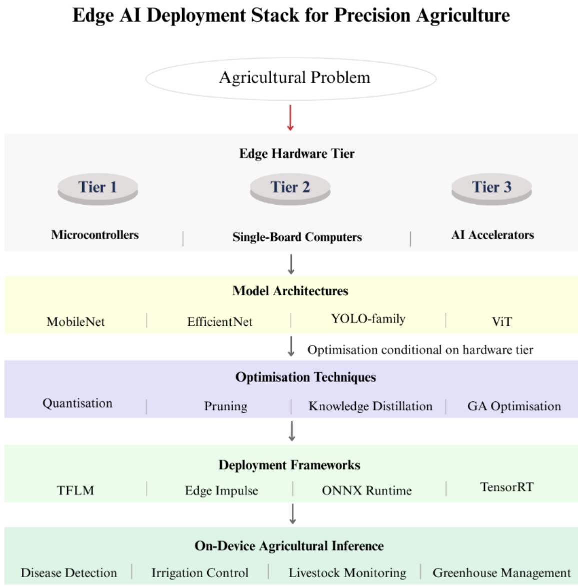

Figure 2: Conceptual overview of the edge AI deployment stack for precision agriculture.

Applications

The following section presents the findings of the review organised across four core application domains identified through the systematic screening process: crop disease detection, irrigation and soil monitoring, livestock monitoring, and greenhouse and environmental monitoring. Each domain is analysed through the five-dimensional evaluation framework introduced in Section 3.5, assessing included studies against Technical Performance, Resource Efficiency, Economic Viability, Deployment Feasibility, and Agricultural Impact. Where data for a specific dimension is unavailable within a study, this is noted explicitly. A summary evidence table is provided within each subsection to support structured cross-study comparison. All performance metrics are reported in the format used by the original study.

Crop Disease Detection

Crop disease detection is a critical edge AI application in precision agriculture, as traditional approaches rely on expert inspection or cloud-based processing, both of which are impractical in remote environments without reliable connectivity. Deploying lightweight models directly on edge devices enables real-time, offline diagnostics, ensuring timely interventions, reducing data transmission bottlenecks, and enhancing the scalability of smart farming solutions.

Hardware selection dictates the fundamental trade-off between computational capability and resource constraints. At the constrained end, an Arduino Nano 33 BLE Sense achieves a 726.6 KB RAM footprint at 7,648 ms inference via a hybrid TensorFlow and Edge Impulse pipeline (Gookyi *et al.*, 2024). At the mid-range, the Raspberry Pi 5 running Tiny-LiteNet via Knowledge

Distillation achieves 98.6% accuracy at 16 ms without dedicated accelerators (Nyakuri *et al.*, 2025). A Raspberry Pi 4 paired with a Google Coral USB Accelerator yields the best cost-to-performance ratio across all benchmarked platforms (Majeed *et al.*, 2024). While Jetson platforms offer superior throughput (Majeed *et al.*, 2024), with DenseNet169 achieving 99.1% accuracy on Jetson Nano (Yousif *et al.*, 2025), the Jetson AGX Orin costs $1,999 and draws 45 W during inference versus 2 W for the Coral USB accelerator (Majeed *et al.*, 2024), making cost and power equally critical constraints.

To meet the memory and compute constraints of edge hardware, model compression is widely applied across the corpus, though the rigor of these optimisations varies considerably. Post-training INT8 quantisation via TensorFlow Lite is most prevalent, maintaining accuracies above 92% (Gookyi *et al.*, 2024; Salabi & Manthila, 2025; Tudevdagva & Shimada, 2023), while Reddy *et al.* (2025) used TensorRT 8.2 for Jetson Nano deployment. Silva and Almeida (2024) advanced this with Pruning and Quantisation-Aware Training for thermal image classification on Edge TPUs, achieving 4.7 ms inference. Appe *et al.* (2025) demonstrated that Vision Transformer architectures can be edge-deployed within a quantised CNN and ViT ensemble on a Raspberry Pi 4, maintaining sub-500 ms latency. By contrast, Nyakuri *et al.* (2025) employed training-time compression via Knowledge Distillation and squeeze-and-excitation depthwise blocks rather than post-training quantisation.

A critical limitation is the scarcity of true field testing, with most studies relying on controlled laboratory conditions. Reddy *et al.* (2025) and Nyakuri *et al.* (2025) report high accuracies of 98.6% without real-farm validation. Bharathi *et al.* (2025) addressed this by deploying EfficientNetB3 across three real greenhouse plots for 60 days, proving a single edge device can handle simultaneous disease detection and precision irrigation control. Salabi and Manthila (2025) validated their IoT camera system under real field conditions with varied lighting and occlusion, achieving 93.4% accuracy, and Tudevdagva and Shimada (2023) conducted a pilot with 60 smallholder farmers, validating against expert agronomists and demonstrating an 18% reduction in crop yield loss. Several non-field-validated studies also rely on the PlantVillage benchmark dataset, raising questions about generalisation to diverse crop varieties, regional conditions, and real-world imaging variability, confirming that field deployment must prioritise robustness over marginal accuracy gains. Table 2 summarises the included studies on crop disease detection across key evaluation dimensions.

Table 2: Summary of Edge AI Deployments for Crop Disease Detection

Author/Year	Hardware	ML Model	Performance Metrics	Size/RAM	Cost	Field Tested
Nyakuri *et al.*, 2025	RPi 5	Tiny-LiteNet	98.6%, 16 ms	1.2 MB / 570 MB	$80	No
Reddy *et al.*, 2025	Jetson Nano + RPi 4	MobileNetV2	98.6%, 80-120 ms	14.3 MB / N/R	N/R	No
Gookyi *et al.*, 2024	Arduino Nano BLE	Custom CNN	94.8%, 7,648 ms	344 KB / 726 KB	N/R	No
Appe *et al.*, 2025	RPi 4B	Ensemble (CNN+ViT)	96.4%, <500 ms	3.2-45 MB / N/R	N/R	No
Bharathi *et al.*, 2025	RPi 4B	EfficientNetB3	98.1%, ~430 ms	N/R / N/R	N/R	Yes
Salabi & Manthila, 2025	RPi 4B	MobileNetV2	93.4%, 0.56 s	8.1 MB / N/R	N/R	Yes
Tudevdagva & Shimada, 2023	RPi 4	Custom CNN	92.3%, 0.9 s	N/R / N/R	N/R	Yes
Yousif *et al.*, 2025	Jetson Nano	Multiple CNNs	Up to 99.1%, 46-252 ms	N/R / 818-955 MB	N/R	No
Majeed *et al.*, 2024	Jetsons + RPi 4 + Coral	Six CNNs	Up to 98.3%, 0.35-2.9 ms	N/R / N/R	$75 - $1,999	No
Silva & Almeida, 2024	RPi 4B + TPU/NCS2	Four CNNs	Up to 98.1%, 4.7-2,800 ms	5-22 MB / N/R	N/R	No

The literature confirms that edge AI can run complex crop disease detection models at low latencies. However, a significant gap remains in translating laboratory prototypes into ruggedised, field-ready systems. Future research should shift from marginal accuracy gains on narrow benchmarks toward longitudinal field trials, economic viability assessments, and greater model generalisation across diverse crop varieties and real-world imaging conditions.

Irrigation and Soil Monitoring

Precision agriculture faces critical challenges from water scarcity and the inefficiency of traditional timer-based irrigation systems. Deploying machine learning models directly on edge devices enables autonomous soil monitoring and irrigation control without continuous cloud connectivity, minimising data transmission latency, reducing bandwidth costs, and ensuring robust operation in remote agricultural environments.

A fundamental architectural divergence exists between hybrid multi-device systems and purely microcontroller-based deployments. Two studies utilise a split architecture where a single-board computer handles AI inference while a secondary microcontroller manages sensor actuation, enabling Reinforcement Learning (Kasera & Acharjee, 2024) and Decision Tree (Al-Jame & Manthila, 2026) deployments. Another study adopts a tiered approach where on-device TinyML inference runs directly on a microcontroller, supported by a local edge server for storage and dashboard visualisation (Taueatsoala *et al.*, 2026). A notably unconventional paradigm is presented by Hayajneh *et al.* (2024), where a hexacopter UAV serves as a flying edge server to deliver over-the-air model updates to distributed ESP32 nodes, an architecture particularly suited to connectivity-deprived rural environments. Solar power is a common design feature, sustaining autonomous off-grid operation across multiple deployments (Baseca *et al.*, 2025; Evangeline *et al.*, 2026; Hayajneh *et al.*, 2024; Kasera & Acharjee, 2024; Kumeresan *et al.*, 2025), while LoRa and LoRaWAN are widely adopted for long-range communication (Al-Jame & Manthila, 2026; Baseca *et al.*, 2025; Kasera & Acharjee, 2024), enabling reliable field coverage without internet infrastructure.

The selection of machine learning algorithms for soil moisture monitoring and irrigation control reveals a clear tension between predictive accuracy and computational footprint. LSTM networks have been successfully deployed on microcontrollers, achieving near-perfect predictive performance (Baseca *et al.*, 2025; Hayajneh *et al.*, 2024). Other studies implement static models such as CNNs (Kumeresan *et al.*, 2025) or lightweight ANNs (Verma *et al.*, 2025), with the latter reaching 3 ms inference via aggressive integer quantisation. Gradient Boosting has demonstrated exceptional regression accuracy on constrained microcontrollers (Taueatsoala *et al.*, 2026), while TinyML deployments relying on undisclosed model architectures raise reproducibility concerns (Evangeline *et al.*, 2026). Reinforcement Learning enables actively learned optimal irrigation schedules but requires an SBC rather than a microcontroller for inference (Kasera & Acharjee, 2024).

The robustness of water conservation claims varies significantly with field testing rigor. Credible evidence comes from longitudinal field deployments: an 87-day trial demonstrating 55% water reduction (Kasera & Acharjee, 2024), a 197-day orchard deployment achieving 35% savings (Baseca *et al.*, 2025), a 30-day multi-soil trial reporting 38.7% savings (Verma *et al.*, 2025), a 6-week semi-arid experiment yielding 38% savings (Al-Jame & Manthila, 2026), and a 20-day open-field test achieving 35 to 42% reduction (Kumeresan *et al.*, 2025). By contrast, studies conducted in controlled or unvalidated environments report estimated figures of 15 to 25% (Taueatsoala *et al.*, 2026) and 40 to 60% (Evangeline *et al.*, 2026), neither of which carries external validity. Notably, Hayajneh *et al.* (2024) report no water savings data, as the system is scoped exclusively to soil moisture forecasting and UAV-assisted model delivery without irrigation actuation. Table 3 summarises the included studies on irrigation and soil monitoring across key evaluation dimensions.

Table 3: Summary of Edge AI Deployments for Irrigation and Soil Monitoring

Author/ Year	Hardware	ML Model	Performance Metric	Water Savings	Power / RAM / Model Size	Hardware Cost	Field Tested
Al-Jame & Manthila, 2026	RPi 4 + STM32	Decision Tree	>94% accuracy	38%	<2.5 W / N/R	N/R	Yes
Baseca *et al.*, 2025	Arduino Edge Control	LSTM	R2=0.9854, RMSE=0.3016	35%	20 mAh/day / N/R	N/R	Yes
Kumeresan *et al.*, 2025	ESP32	Custom CNN	96.23% accuracy	35 to 42%	280 mW active / N/R	N/R	Yes
Verma *et al.*, 2025	ESP32-WROOM-32	Lightweight ANN	R2=0.945, MAE=3.15%	38.7%	N/R / 2.9 KB size	Under $25	Yes
Evangeline *et al.*, 2026	ESP32 DevKit	TinyML (TFLite)	N/R	40 to 60%	N/R / N/R	~ $300 USD	No
Taueatsoala *et al.*, 2026	ESP32 + RPi 5	Gradient Boosting	R2=0.9973 , MAPE=0.99%	15 to 25%	N/R / 0.8 MB size	N/R	No
Kasera & Acharjee, 2024	Arduino Mega + RPi 4	RL-PPO	97.88% accuracy	55%	N/R / N/R	N/R	Yes

Hayajneh *et al.*, 2024	ESP32 + UAV	DNN & LSTM	LSTM $R2=0.9996$	N/R	N/R / 32,768 bytes RAM	N/R	Yes

A collective analysis indicates a decisive shift toward highly quantised models on resource-constrained microcontrollers. Basic parameters such as volumetric soil moisture and temperature are universally monitored, while advanced agronomic indicators like macronutrients are monitored in only one study (Kasera & Acharjee, 2024), remaining underrepresented across the domain. While robust field trials provide credible evidence of substantial water savings (Al-Jame & Manthila, 2026; Baseca *et al.*, 2025; Kasera & Acharjee, 2024; Kumeresan *et al.*, 2025; Verma *et al.*, 2025), a considerable portion of the literature relies on controlled or unvalidated environments that fail to capture the complexities of open farm conditions (Evangeline *et al.*, 2026; Taueatsoala *et al.*, 2026). Hardware cost data is absent from six of the eight reviewed studies, leaving economic viability largely unassessed and preventing meaningful comparison of smallholder affordability. Addressing these gaps by integrating chemical soil sensors, mandating multi-season field validations, and ensuring full model architecture transparency is essential for the advancement of future predictive irrigation systems.

Livestock Monitoring

Traditional manual observation is limited by labour costs, human error, and impracticality at scale. Deploying machine learning models directly on wearable and remote edge devices enables early disease detection and continuous behavioural tracking at the individual animal level. Livestock environments introduce unique challenges around battery longevity, processing capacity, and network reliability that drive specialised architectural choices.

A primary division in the literature lies between camera-based computer vision and wearable sensor deployments. Camera-based systems on NVIDIA Jetson devices enable non-invasive monitoring by processing high-resolution visual feeds, demonstrated across digital dermatitis detection (Aravamuthan *et al.*, 2024), physiological state classification across ten commercial dairy farms (Mahato & Neethirajan, 2025), and individual cattle identification (Smink *et al.*, 2024). Wearable systems use IMUs to track individual animal movements continuously; STM32-based collar deployments with accelerometers and GPS achieve fine-grained behaviour recognition directly on-device, reaching 93.20% accuracy at 11 ms per sample (Liu *et al.*, 2026). Wearable sensors provide uninterrupted animal-specific data regardless of occlusion or lighting, but impose stricter constraints on memory footprint and energy consumption than stationary camera nodes. Audio-based monitoring constitutes a distinct third approach, enabling respiratory disease detection (Qiao *et al.*, 2026) and vocalisation classification in poultry without visual input (Srinivasagan *et al.*, 2025).

Research distribution across species reveals imbalances affecting broader applicability. Cattle are heavily represented across individual identification (Smink *et al.*, 2024), physiological classification (Mahato & Neethirajan, 2025), hoof disease detection (Aravamuthan *et al.*, 2024), and behaviour recognition (Jao *et al.*, 2025; Liu *et al.*, 2026). Swine receive considerable attention through models for piglet behaviour recognition (Luo *et al.*, 2025), farrowing event detection and counting (Kong *et al.*, 2025), and posture tracking in commercial pig housing (Zha *et al.*, 2023). Poultry monitoring has grown in research activity; audio-based systems yield strong results, with a RegNet classifier on Raspberry Pi 4B reaching 96.03% accuracy for respiratory disease detection across three commercial farms (Qiao *et al.*, 2026) and a 1D CNN on a Syntiant TinyML board achieving 96.6% accuracy for six-class vocalisation classification (Srinivasagan *et al.*, 2025). A vision-based YOLO-CCA model on a Jetson AGX Orin additionally demonstrated strong caged chicken counting performance in dense farm environments (Wu *et al.*, 2025). Sheep remain the least represented species, with a single ESP32-based IMU deployment for behaviour classification (Suhaimi *et al.*, 2025). Edge AI also extends to aquatic species; YOLOv7 on a Raspberry Pi 4 has been demonstrated for Atlantic salmon mortality detection (Ranjan *et al.*, 2023).

Communication protocols dictate the scalability and energy efficiency of remote monitoring systems. Bluetooth Low Energy (BLE) suits wearable deployments, enabling multi-device coordination across up to 200 metres between fence-mounted and neck-worn units (Zhang & Kanjo, 2025). Long-range 4G LTE enables offloading across wide pastoral areas, while on-device inference eliminates continuous raw data transmission (Liu *et al.*, 2026). These choices shape architecture directly: systems in low-connectivity areas must perform complete on-device inference and transmit only lightweight classifications rather than raw sensor streams.

Behaviour classification is the most common task category, with many systems accurately categorising routine posture and activity behaviours using compressed models on microcontrollers (Jao *et al.*, 2025; Liu *et al.*, 2026; Suhaimi *et al.*, 2025; Zhang & Kanjo, 2025). While behaviour tracking serves as a welfare proxy, it lacks the diagnostic specificity needed for early disease intervention. Direct health monitoring, including dermatological lesion detection (Aravamuthan *et al.*, 2024) and respiratory disease identification (Qiao *et al.*, 2026), alongside welfare-oriented vocalisation classification (Srinivasagan *et al.*, 2025), relies on dedicated camera-based or audio sensors. This highlights a persistent limitation of wearable systems, which struggle to detect complex physiological anomalies without additional sensing modalities beyond inertial measurements. Zhang and Kanjo (2025) represent a step toward addressing this through a dual-device system combining neck-mounted accelerometers with fence-mounted cameras.

Battery life and computational constraints shape the architectural complexity of wearable deployments. Models are aggressively optimised through post-training quantisation and neural architecture pruning to fit restricted microcontroller memory (Jao *et al.*, 2025; Liu *et al.*, 2026). The S-ResNet 0.1 model occupies just 55.73 KiB of flash while achieving 93.20% accuracy, sustained by a 9,000 mAh battery with solar recharging (Liu *et al.*, 2026). Camera-based systems on fixed power supplies support larger architectures, with Jetson-class devices enabling real-time high-frame-rate inference (Aravamuthan *et al.*, 2024; Mahato & Neethirajan, 2025; Zha *et al.*, 2023); their stationary nature, however, prevents continuous monitoring of individual animals across expansive landscapes. Table 4 summarises the included studies on livestock monitoring across key evaluation dimensions.

Edge AI for cattle and swine monitoring shows strong

Table 4: Summary of Edge AI Deployments for Livestock Monitoring

Author/Year	Hardware	Sensor Type	ML Model	Animal/Task	Performance Metrics	Hardware Cost	Field Tested
Aravamuthan *et al.*, 2024	Jetson Xavier NX + OAK-1	RGB camera	Tiny YOLOv4	Dairy cattle / hoof disease detection	mAP=0.895, Acc=0.873, 40 FPS	N/R	No
Luo *et al.*, 2025	Jetson Orin NX	RealSense D435i	YOLOv8-Piglet	Piglets / multi-behavior recognition	Prec=92.6%, Rec=91.2%, mAP=91.8%, 163.2 ms	N/R	No
Jao *et al.*, 2025	Sipeed Maixduino (K210)	OV2640 camera	YOLOv2-MobileNet0.75	Cattle / behavior classification	Acc=87.5%, 35 ms	N/R	No
Qiao *et al.*, 2026	RPi 4B	USB microphone	RegNet	Chickens / respiratory disease	Acc=96.03%, F1=93.26%, 1933 ms/5s clip	$72.71	Yes
Liu *et al.*, 2026	STM32U585 (wearable collar)	IMU + GPS	S-ResNet 0.1	Cattle / behavior recognition	Acc=93.20%, 11 ms	N/R	Yes
Suhaimi *et al.*, 2025	ESP32 Mini C3	MPU6050 IMU	Random Forest	Sheep / behavior classification	Acc=92.1%, F1=0.927, <50 ms	N/R	No
Zhang & Kanjo, 2025	Coral Dev Micro + RPi Pico	ADXL345 + cameras	CNN-LSTM + MCUNet	Cattle / multi-behavior recognition	Acc=98.94% & fused=96.87%, 75 ms	$100 + $25-30/unit	No
Mahato & Neethirajan, 2025	Jetson Xavier NX	RGB camera	YOLOv11-nano	Dairy cattle / physiological state	Acc=94.2%, mAP50=0.947, 24 FPS	$3,399 (100 cows)	Yes
Kong *et al.*, 2025	RPi 4B	RGB camera	MGDT-YOLO	Piglets / detection and counting	Prec=88.5%, Rec=91.7%, mAP=93.8%, 87 ms	N/R	Partial
Smink *et al.*, 2024	Jetson AGX Orin	RTSP RGB camera	YOLOv5s + TRBA	Dairy cattle / ear tag identification	Acc=84.2%, ~24.69 FPS	N/R	Yes
Zha *et al.*, 2023	Jetson AGX Xavier	RGB camera	YOLOv5s DCCA + DeepSort	Pigs / posture recognition and tracking	Prec=96.6%, Rec=95.4%, mAP=97.0%, 25–46 ms	N/R	Partial
Srinivasagan *et al.*, 2025	Syntiant TinyML board	Microphone	1D CNN	Chickens / vocalization classification	Acc=96.6%, F1=0.95, 47–91 ms	N/R	Yes

Ranjan *et al.*, 2023	RPi 4 + PoE HAT	RGB camera	YOLOv7	Atlantic salmon / mortality detection	mAP=93.4%, F1=0.89, N/R	N/R	Yes
Wu *et al.*, 2025	Jetson AGX Orin	RGB camera	YOLO-CCA	Chickens / flock counting	Acc=93.2%, F1=96.7%, 90.9 FPS	N/R	Yes

maturation, driven by advances in lightweight computer vision and efficient wearable sensors. Sheep and aquatic species remain underrepresented, and robust on-device health diagnostics are constrained by sensor limitations of current microcontroller platforms. Hardware cost data is absent from eleven of the fourteen reviewed studies, representing a critical gap in the Economic Viability dimension that prevents meaningful assessment of deployment affordability. Future research should develop multi-modal wearable sensors combining inertial tracking with physiological data acquisition, and extend validated deployments to a wider range of species and real-world environments.

Greenhouse and Environmental Monitoring

Greenhouse and controlled environment agriculture represent a fundamentally distinct edge AI domain compared to open-field systems. Unlike field-based irrigation, greenhouses demand continuous multi-parameter regulation across temperature, humidity, CO2 levels, and light intensity simultaneously. The enclosed and sensor-rich nature of these environments makes them ideal candidates for autonomous edge intelligence, enabling real-time on-device inference without cloud dependency. The breadth of monitoring tasks within this domain, spanning climate forecasting, disease detection, and adaptive control, makes it one of the most technically varied application areas reviewed.

The hardware landscape reveals a decisive shift toward more computationally capable platforms. Tasks requiring deep learning inference on continuous time-series data or real-time image processing demand the GPU-accelerated capacity of platforms such as the NVIDIA Jetson Nano or Raspberry Pi, which incur higher power envelopes reaching up to 10W (Proietti *et al.*, 2021). Multi-device architectures pairing a capable inference platform with lightweight sensor nodes remain prevalent, as seen in deployments combining Raspberry Pi with Arduino for hydroponic control (Phukan, 2022) and ESP32 with ESP8266 for hybrid climate management (Venkataramanan *et al.*, 2025). However, when inference tasks are sufficiently constrained, ultra-low-power microcontrollers remain highly competitive, with an Arduino Uno achieving 99.9% accuracy at 13 microseconds inference latency via TinyML toolkits (Sanchez-Iborra *et al.*, 2023), confirming that task complexity rather than platform prestige should dictate hardware selection.

The defining characteristic of this domain is the diversity of monitoring tasks, each demanding a fundamentally different modelling approach. Time-series climate forecasting favours recurrent and hybrid architectures, with a CNN-LSTM model achieving a MAE of 1.2°C for 12-hour greenhouse temperature prediction (Morales-García *et al.*, 2023a) and an LSTM Encoder-Decoder successfully detecting growth anomalies through multi-sensor sequence reconstruction (Proietti *et al.*, 2021). For continuous environmental optimisation, a Genetic Algorithm-tuned 1D-CNN achieved R^2=0.99 for plant growth condition monitoring (Ugli & Lee, 2025), while a DNN deployed on Raspberry Pi 3 regulated hydroponic parameters at 88.5% accuracy (Phukan, 2022). Visual monitoring tasks demand a different paradigm entirely, with a channel-pruned YOLOv5s model on Jetson Nano achieving 96.7% mAP at 13.8 ms latency for real-time melon disease detection, compressed to just 1.1 MB through backbone reconstruction and channel pruning (Xu *et al.*, 2022). For adaptive control, reinforcement learning has also shown strong potential, with a Q-learning agent achieving 83.3% energy reduction for greenhouse lighting control (Salem *et al.*, 2025), while ensemble methods offer near-perfect accuracy with minimal computational overhead for structured decision support (Sanchez-Iborra *et al.*, 2023). Table 5 summarises the included studies on greenhouse and environmental monitoring across key evaluation dimensions.

Six of the nine reviewed systems were validated in real greenhouse environments, reflecting a stronger field-testing culture in this domain. Real deployments demonstrated measurable agricultural impact, including a 15% increase in crop yield and 30% reduction in water consumption (Venkataramanan *et al.*, 2025) and accelerated tomato plant growth in an edge-controlled hydroponic system (Phukan, 2022), while controlled prototype studies also reported improved plant growth metrics under symmetry-aware climate optimisation (Ugli & Lee, 2025). Energy efficiency emerged as a prominent reporting dimension, with Salem *et al.* (2025) demonstrating 83.3% power reduction and Morales-García *et al.* (2023b) confirming that Arduino-based MLP inference consumes as little as 0.0001 Joules per execution. Notably, hardware cost data was entirely absent across all nine studies, representing a critical economic viability gap for practical adoption assessments. While multi-parameter monitoring is common across the domain, unified modelling of the full greenhouse climate spectrum within a single system remains rare, with several studies focusing narrowly on a single variable such as temperature (Morales-García *et al.*, 2023a; 2023b) or light intensity (Salem *et al.*, 2025).

A collective analysis highlights strong technical performance and deployment feasibility across the

Table 5: Summary of Edge AI Deployments for Greenhouse and Environmental Monitoring

Author/Year	Hardware	Application Task	ML Model	Performance Metric	Power / Model Size	Field Tested
Proietti *et al.*, 2021	RPi 4 + Jetson Nano	Anomaly Detection	LSTM Encoder-Decoder	F1=0.481, 8.32-30.84 ms latency	7.3W (RPi 4) / 10W (Jetson) / N/R size	Yes
Venkataramanan *et al.*, 2025	ESP32 + ESP8266	Climate Control & Yield Prediction	Random Forest + LSTM	88% yield accuracy	240 mA / N/R	Yes
Salem *et al.*, 2025	ESP32	Lighting Control	Q-learning (RL)	< 100ms convergence	1W (83.3% reduction vs open-loop) / N/R	No
Phukan, 2022	RPi 3 + Arduino	Hydroponic Environment Control	DNN	88.5% accuracy	N/R / N/R	Yes
Sanchez-Iborra *et al.*, 2023	Arduino Uno	Irrigation & Fertigation Decision Support	DT + RF	99.9% accuracy, 13 µs latency	0.11 µA inference, 3.2 mA LoRa / 3,392 B Flash, 346 B SRAM	No
Ugli & Lee, 2025	Jetson Nano	Growth Optimization	1D-CNN (GA-optimized)	R^2=0.99, SEP=0.35	N/R / N/R	No
Xu *et al.*, 2022	Jetson Nano	Greenhouse Disease Detection	Pruned YOLOv5s (PYSS)	96.7% mAP, 13.8 ms latency	N/R / 1.1 MB (85% reduction)	Yes
Morales-García *et al.*, 2023a	Jetson Nano	Temperature Forecasting	CNN-LSTM	MAE=1.2°C, R^2=0.947	N/R / N/R	Yes
Morales-García *et al.*, 2023b	Arduino + RPi + Jetson	Temperature Prediction	MLP + CNN	R^2=0.85, RMSE=2.05	~0.0001 J/exec (Arduino MLP) / N/R	Yes

domain, yet several gaps persist. Model degradation over time represents a significant operational challenge, with temperature forecasting models recommended for retraining by day 9 of deployment and requiring mandatory retraining by day 19 to prevent accuracy deterioration (Morales-García *et al.*, 2023a). The complete absence of economic viability reporting across all nine studies further limits practical scalability assessments. Future work must prioritise cost-transparent deployments, broader sensor integration, and autonomous retraining pipelines to advance greenhouse edge AI toward wider agricultural adoption.

Critical Analysis and Discussion

Cross-Domain Hardware and Architecture Patterns

Across all four domains, hardware selection is the single most consequential architectural decision, directly shaping model complexity, energy budget, and deployment feasibility. Ultra-constrained microcontrollers dominate structured sensor inference tasks, exemplified by an irrigation and fertigation decision support system achieving 99.9% accuracy within 346 bytes of SRAM (Sanchez-Iborra *et al.*, 2023) and a wearable cattle behaviour classifier fitting within 55.73 KiB of flash at 11 ms inference (Liu *et al.*, 2026), while mid-tier SBCs serve tasks exceeding microcontroller memory limits, including audio-based disease detection (Qiao *et al.*, 2026) and fish mortality monitoring (Ranjan *et al.*, 2023). Jetson-class hardware predominantly serves computationally intensive visual tasks such as multi-behaviour recognition and livestock counting (Luo *et al.*, 2025; Wu *et al.*, 2025), though it also supports non-visual tasks including time-series forecasting (Morales-García *et al.*, 2023a) and growth condition optimisation (Ugli & Lee, 2025). The strongest empirical validation of this hierarchy comes from Morales-García *et al.* (2023b), who evaluated identical temperature prediction tasks across Arduino, Raspberry Pi, and Jetson simultaneously, finding that Arduino MLP inference consumed as little as 0.0001 joules per execution, confirming that task complexity rather than hardware prestige should govern platform selection.

Split and hybrid architectures appear across multiple domains, ranging from tightly coupled SBC-microcontroller pairs (Al-Jame & Manthila, 2026; Phukan, 2022) to dual-node BLE-fused deployments combining a Coral Dev Micro with a Raspberry Pi Pico (Zhang & Kanjo, 2025) and a hexacopter UAV serving as a flying edge server (Hayajneh *et al.*, 2024). Hardware accelerators offer a targeted solution to the mid-tier to Jetson-class performance gap, with Coral Edge TPU augmentation on Raspberry Pi 4 delivering the best cost-to-performance ratio across benchmarked platforms (Majeed *et al.*, 2024) and achieving 4.7 ms inference (Silva & Almeida, 2024), approaching Jetson-class speed at substantially lower cost and power draw. However, several Jetson-class studies across the crop disease and livestock domains were validated exclusively in laboratory settings (Aravamuthan *et al.*, 2024; Luo *et al.*, 2025; Yousif *et al.*, 2025), raising the concern that such systems may not transfer to the constrained, cost-sensitive platforms required in real agricultural deployments.

Model Optimisation and Compression Trends

Model compression is not an optional refinement but a fundamental prerequisite, as the memory and compute constraints of field-appropriate hardware make unoptimised models non-viable. Post-training INT8 quantisation via TensorFlow Lite is the dominant strategy in crop disease detection and the most frequently documented approach across the reviewed corpus, with its accessibility and workflow compatibility explaining its breadth of application: maintaining accuracy above 92% in disease classification (Salabi & Manthila, 2025; Tudevdagva & Shimada, 2023), reducing an irrigation ANN to 2.9 KB at 3 ms inference (Verma *et al.*, 2025), and enabling Arduino Nano BLE deployment via a combined TFLite and Edge Impulse pipeline (Gookyi *et al.*, 2024).

Training-time compression offers notable advantages in the examples documented: Knowledge Distillation produced a 1.2 MB Tiny-LiteNet at 98.6% accuracy (Nyakuri *et al.*, 2025), combined pruning and quantisation-aware training enabled 4.7 ms Edge TPU inference on thermal images (Silva & Almeida, 2024), channel pruning with backbone reconstruction reduced YOLOv5s by 85% to 1.1 MB for greenhouse disease detection (Xu *et al.*, 2022), and GNN-based channel pruning guided by reinforcement learning compressed S-ResNet to a 55.73 KiB flash footprint (Liu *et al.*, 2026). MCU-targeted architecture selection represents a complementary strategy, with Zhang and Kanjo (2025) deploying MCUNet, an architecture designed via neural architecture search specifically for microcontroller constraints, for livestock behaviour recognition. Across irrigation, greenhouse, and livestock domains, deliberate selection of classical algorithms such as Decision Trees, Gradient Boosting, and Random Forests achieves competitive accuracy on severely constrained hardware (Al-Jame & Manthila, 2026; Sanchez-Iborra *et al.*, 2023; Suhaimi *et al.*, 2025; Taueatsoala *et al.*, 2026), with the most extreme example fitting within 3,392 bytes of Flash on an Arduino Uno (Sanchez-Iborra *et al.*, 2023), confirming that algorithm selection is itself a resource efficiency decision.

A persistent limitation is the inconsistency of compression reporting across the corpus, with multiple studies omitting model size, quantisation method, or pre- and post-compression accuracy deltas, the most severe instance being a TinyML irrigation deployment where model architecture, size, and performance metrics are entirely undisclosed (Evangeline *et al.*, 2026). Beyond reporting gaps, Yousif *et al.* (2025) apply no quantisation or compression, with pruning deferred to future work, limiting the transferability of that model to the cost- and power-constrained platforms required for real agricultural deployment.

Field Deployment and Validation Gaps

Field validation represents the critical bridge between technical performance and genuine agricultural utility, yet the reviewed literature reveals a persistent and domain-variable gap between laboratory prototypes and real-world deployments. Field testing rates vary considerably across the four domains: crop disease detection is the weakest at 3 out of 10, despite being the most technically mature by reported accuracy; irrigation and soil monitoring is the strongest at 6 out of 8, and is correspondingly the only domain where agricultural impact is consistently quantified through longitudinal water savings data; livestock monitoring achieves 7 out of 14 fully field-tested with 2 partial deployments; and greenhouse monitoring sustains 6 out of 9. The correlation between higher field testing rates and more concrete agricultural outcome reporting suggests that laboratory-confined studies systematically overestimate deployment readiness. High-accuracy claims from lab-only studies warrant particular scrutiny. Reddy *et al.* (2025) and Nyakuri *et al.* (2025) both report 98.6% accuracy in crop disease detection, and Yousif *et al.* (2025) achieve 99.1%, yet none are validated under variable lighting, occlusion, or environmental noise conditions of actual farms. Similarly, strong livestock detection metrics from Aravamuthan *et al.* (2024) and Luo *et al.* (2025) were obtained entirely in laboratory settings on Jetson-class hardware, reinforcing the concern raised in Section 6.1 that Jetson-class systems validated only in laboratory settings may not transfer to the cost-constrained platforms required in real agricultural deployments.

Longitudinal deployment depth further distinguishes meaningful validation from superficial testing. The most credible evidence comes from extended trials including a 197-day orchard deployment (Baseca *et al.*, 2025), an 87-day irrigation study (Kasera & Acharjee, 2024), and a 60-day greenhouse trial (Bharathi *et al.*, 2025). Critically, Morales-García *et al.* (2023a) is the only study across all 41 papers to explicitly document model degradation over time, reporting that temperature forecasting accuracy deteriorated sufficiently to require preventive retraining by day 9 and mandatory retraining by day 19, a finding that raises unresolved questions about long-term model stability that the broader corpus largely ignores.

Economic Viability Gap

Hardware cost is the most underreported dimension across the entire reviewed corpus, with only 7 of 41 studies reporting any cost data and the greenhouse domain maintaining complete cost silence across all nine studies. This absence reflects a broader pattern in which studies prioritise technical demonstration over practical adoptability, treating economic accessibility as outside their scope.

The sparse available data reveals a meaningful spread. At the lowest confirmed hardware cost, Verma *et al.* (2025) at under $25 per node represents the most accessible single-unit price point in the corpus. Nyakuri *et al.* (2025) at $80 and Qiao *et al.* (2026) at $72.71 suggest cooperative or subsidised viability for smallholder contexts. Zhang and Kanjo (2025) deploy a two-device system at $25-30 per wearable node and $100 per static fence unit, with a total estimated deployment cost of $3,000-6,000 for 100-200 animals. Majeed *et al.* (2024) benchmark across a $75-$1,999 hardware range, explicitly identifying the $135 RPi4+Coral combination as the best cost-to-performance option. Evangeline *et al.* (2026) at approximately $300 is notable as the only cost-reporting deployment with no ML performance metric disclosed. Mahato and Neethirajan (2025) report a 5-year operational cost of $3,399 for a 100-cow Jetson-class system, a total cost of ownership figure not directly comparable to the hardware costs above. Without broader cost transparency, accessibility for low- and middle-income farming contexts cannot be meaningfully evaluated.

Key Challenges

Despite the technical progress demonstrated across the reviewed corpus, several systemic barriers persistently impede the transition from laboratory prototype to production-ready agricultural deployment. Most recur across multiple application domains rather than being isolated to individual studies.

Dataset bias and domain generalisation represent a foundational vulnerability. Most models are trained and evaluated on curated datasets collected under controlled conditions, and only Hayajneh *et al.* (2024) explicitly address cross-environment generalisation, achieving $R^2 = 0.929$ on a new geographic domain via cross-site Transfer Learning. No study tests for domain shift under the variable lighting, occlusion, and seasonal variation of real farm conditions, meaning reported accuracy figures cannot be assumed to transfer to open-field deployments.

Model drift and long-term stability remain almost entirely unaddressed across the corpus. As documented in Section 6.3, only one study explicitly records accuracy deterioration under deployment, yet no study implements on-device continual learning or autonomous retraining pipelines to address distributional shift arising from seasonal change or evolving crop growth stages.

Physical hardware robustness in agricultural environments is consistently underreported. Dust, moisture, vibration, and animal interference impose demands that few studies acknowledge, with enclosure specifications or ingress protection ratings almost never disclosed. Software-level real-time constraints are similarly underexplored, with no study systematically evaluating latency tolerances under the variable conditions of live farm operation.

Reproducibility is further undermined by inconsistent reporting standards and the absence of shared benchmarking frameworks. Model architecture, compression method, training dataset composition, and evaluation conditions are irregularly disclosed across the corpus, while no common evaluation protocol exists to enable meaningful cross-study comparison.

Within the livestock domain specifically, the sensor limitations of current wearable deployments impose a persistent diagnostic ceiling. MCU-based wearables are constrained predominantly to inertial data, restricting output to behavioural proxies rather than direct physiological indicators. Bridging this gap requires multi-modal integration of inertial, biochemical, and environmental sensing within a single low-power wearable architecture.

Power management and connectivity assumptions remain inadequately addressed across all domains. Several systems assume persistent Wi-Fi or 4G availability without accounting for rural connectivity gaps, and battery-powered deployments rarely disclose duty cycling strategies beyond total consumption figures. Sustained off-grid operation, as demonstrated in the longitudinal deployments documented in Section 6.3, remains an exception rather than a standard design requirement across the reviewed corpus.

Future Research Directions

The gaps and challenges identified across this review point toward six concrete directions that future research must prioritise to advance edge AI and TinyML from agricultural prototyping toward scalable, real-world deployment.

The most immediate need is the establishment of standardised benchmarking and reporting protocols across the field. Community-agreed minimum reporting requirements covering model size, quantisation method, pre- and post-compression accuracy delta, hardware cost, and field trial duration would enable the kind of systematic cross-study comparison that the current corpus cannot support, and the five-dimensional evaluation framework applied in this review offers one candidate template for such standardisation.

Continual learning and on-device adaptation pipelines represent a critical architectural frontier. As documented in Sections 6.3 and 6.5, model drift under real deployment conditions remains almost entirely unaddressed across the corpus, underscoring the need for lightweight on-device fine-tuning mechanisms capable of adapting to distributional shifts introduced by seasonal change and evolving crop growth stages without full retraining cycles.

Multi-modal sensor fusion remains substantially underexplored across the livestock, irrigation, and greenhouse domains. Crop disease detection, as a predominantly visual monitoring task, is the natural exception. Wearable livestock systems should integrate physiological sensors alongside inertial measurement units to move beyond behavioural proxies toward direct health diagnostics, while irrigation and greenhouse deployments should expand beyond one or two monitored parameters toward full agronomic and climate spectrum sensing.

Smallholder-oriented hardware design should be elevated from an incidental outcome to an explicit research objective. The per-unit distributed cost model demonstrated by Zhang and Kanjo (2025) and the sub-$25 deployment by Verma *et al.* (2025) offer replicable design templates, and future studies should treat hardware cost and off-grid autonomy as primary reporting metrics rather than secondary considerations.

Longitudinal multi-season field trials should replace the short-duration and single-session validations that dominate the current corpus. Cross-environment generalisation testing is almost entirely absent, with Hayajneh *et al.* (2024) representing the only cross-site generalisation attempt across all 41 papers, and it remains an essential next step toward globally applicable systems. Future deployments should additionally document physical hardware robustness under real agricultural conditions, including ingress protection, thermal tolerance, and resistance to dust, moisture, and mechanical interference.

Finally, the independent convergence of reinforcement learning for closed-loop control optimisation across the irrigation (Kasera & Acharjee, 2024) and greenhouse (Salem *et al.*, 2025) domains suggests a promising but underexplored paradigm. Kasera and Acharjee's (2024) RL-PPO already demonstrates that non-tabular, function-approximation approaches are deployable on SBC-class edge hardware, and future work should extend this capability toward MCU-class constrained devices where, as Salem *et al.*'s (2025) tabular Q-learning illustrates, architectural limitations currently restrict the complexity of learnable control policies.

CONCLUSION

This paper presented a hybrid systematic-scoping review of on-device machine learning inference in precision agriculture, examining 41 peer-reviewed and preprint studies published between January 2020 and March 2026 and following PRISMA-ScR reporting guidelines (Tricco *et al.*, 2018). The review addressed three research questions covering the hardware platforms and model architectures deployed for edge inference, their performance evaluated across five dimensions, and the primary challenges and open directions facing the field.

Findings confirm consistent hardware-model alignment: MCU-class devices dominate structured sensor inference, mid-tier SBCs serve tasks exceeding microcontroller limits, and Jetson-class platforms support computationally intensive visual inference. Post-training INT8 quantisation is the dominant compression strategy, though training-time approaches deliver stronger accuracy retention. Technical performance is strong (classification accuracy 92.3–99.9%; regression R^2 = 0.85–0.99), but uneven across evaluation dimensions. Irrigation and soil monitoring achieves the highest field validation rate (6/8 studies) and is the only domain with consistently quantified agricultural impact. Economic viability is the most underreported dimension, with only 7 of 41 studies disclosing hardware costs.

Three primary contributions are made: a three-tier hardware taxonomy, a five-dimensional evaluation framework applied uniformly across all 41 studies, and a documented set of systemic challenges with six evidence-grounded future research directions. Limitations include English-language database restrictions, a corpus skewed toward livestock monitoring, and heterogeneous reporting standards that constrained the completeness of the five-dimensional evaluation. The most immediate priority for the field is the establishment of standardised benchmarking and reporting protocols, alongside advances in continual learning, multi-modal sensor fusion, smallholder-oriented hardware design, longitudinal field validation, and reinforcement learning for closed-loop agricultural control.

REFERENCES

Al-Jame, F., & Manthila, P. (2026). AI-Driven Smart Irrigation System Using Edge-Based Embedded Controllers. *Progress in Electronics and Communication Engineering, 3*(2), 23–30. https://doi.org/10.31838/ECE/03.02.04

Appe, S. N., N, B. G., & Agarwal, S. (2025). An Edge-Deployable Lightweight Ensemble Framework For Grape Leaf Disease Detection Using Vision Transformers And CNNs. *International Journal of Environmental Sciences, 11*(16). https://theaspd.com/index.php/ijes/article/download/3634/2700/8693

Aravamuthan, S., Walleser, E., & Döpfer, D. (2024). Benchmarking analysis of computer vision algorithms on edge devices for the real-time detection of digital dermatitis in dairy cows. *Preventive Veterinary Medicine, 231,* 106300. https://doi.org/10.1016/j.prevetmed.2024.106300

Baseca, C. C., Dionísio, R., Ribeiro, F., & Metrôlho, J. (2025). Edge-Computing Smart Irrigation Controller Using LoRaWAN and LSTM for Predictive Controlled Deficit Irrigation. *Sensors (Basel, Switzerland), 25*(22), 7079. https://doi.org/10.3390/s25227079

Bharathi, M. P., Sinchana, K. S., & Yashashwini, B. S. (2025). Smart Agro-IoT System with Edge-AI for CropLeaf Disease Detection and Precision Irrigation. *IJARCCE, 14*(9). https://doi.org/10.17148/IJARCCE.2025.14913

Evangeline, A. B., Alex, A. M., Saran, R. S., & J, S. (2026). Smart Irrigation System for Precision Farming Using IoT, TinyML, Hybrid GSM and WiFi and Chatbot. *International Research Journal on Advanced Engineering Hub (IRJAEH), 4*(01), 27–31. https://doi.org/10.47392/

IRJAEH.2026.0004

FAO, IFAD, UNICEF, WFP, & WHO. (2024). The State of Food Security and Nutrition in the World 2024. Financing to end hunger, food insecurity and malnutrition in all its forms. FAO; IFAD; UNICEF; WFP; WHO. https://doi.org/10.4060/cd1254en

Gookyi, D. A. N., Wulnye, F. A., Arthur, E. A. E., Ahiadormey, R. K., Agyemang, J. O., Agyekum, K. O.-B. O., & Gyaang, R. (2024). TinyML for smart agriculture: Comparative analysis of TinyML platforms and practical deployment for maize leaf disease identification. *Smart Agricultural Technology, 8,* 100490. https://doi.org/10.1016/j.atech.2024.100490

Hayajneh, A. M., Aldalahmeh, S. A., Alasali, F., Al-Obiedollah, H., Zaidi, S. A., & McLernon, D. (2024). Tiny machine learning on the edge: A framework for transfer learning empowered unmanned aerial vehicle assisted smart farming. *IET Smart Cities, 6*(1), 10–26. https://doi.org/10.1049/smc2.12072

Heydari, S., & Mahmoud, Q. H. (2025). Tiny Machine Learning and On-Device Inference: A Survey of Applications, Challenges, and Future Directions. *Sensors, 25*(10), 3191. https://doi.org/10.3390/s25103191

Jao, J. R., Vallar, E. A., & Hameed, I. (2025). Smart Cattle Behavior Sensing with Embedded Vision and TinyML at the Edge. *Engineering Proceedings, 118*(1), 81. https://doi.org/10.3390/ECSA-12-26519

Kalyani, Y., & Collier, R. (2021). A Systematic Survey on the Role of Cloud, Fog, and Edge Computing Combination in Smart Agriculture. *Sensors, 21*(17), 5922. https://doi.org/10.3390/s21175922

Kasera, R. K., & Acharjee, T. (2024). A Comprehensive IoT edge based smart irrigation system for tomato cultivation. *Internet of Things, 28,* 101356. https://doi.org/10.1016/j.iot.2024.101356

Kong, N., Liu, T., Li, G., Xi, L., Wang, S., & Shi, Y. (2025). Attention-Guided Edge-Optimized Network for Real-Time Detection and Counting of Pre-Weaning Piglets in Farrowing Crates. *Animals, 15*(17), 2553. https://doi.org/10.3390/ani15172553

Kuan, Y. N., Goh, K. M., & Lim, L. L. (2025). Systematic review on machine learning and computer vision in precision agriculture: Applications, trends, and emerging techniques. *Engineering Applications of Artificial Intelligence, 148,* 110401. https://doi.org/10.1016/j.engappai.2025.110401

Kumeresan, G., Parkavi, M., Kanibharathi, P., Praneshwaran, M., & Premalatha, M. (2025). AI-Integrated Smart Irrigation System Using ESP32 and LoRa. *International Journal of Research Publication and Reviews, 6.* https://ijrpr.com/uploads/V6ISSUE11/IJRPR56340.pdf

Liu, H., Song, P., Xin, X., Rong, Y., Gao, J., Wang, Z., & Zhang, Y. (2026). A Cattle Behavior Recognition Method Based on Graph Neural Network Compression on the Edge. *Animals, 16*(3), 430. https://doi.org/10.3390/ani16030430

Luo, Y., Lin, K., Xiao, Z., Chen, Y., Yang, C., & Xiao, D. (2025). Collaborative Optimization of Model Pruning and Knowledge Distillation for Efficient and Lightweight Multi-Behavior Recognition in Piglets. *Animals, 15*(11), 1563. https://doi.org/10.3390/ani15111563

Madiwal, A. S., Jha, R. B., & Barthakur, P. (2025). Edge AI and Iot for Real-Time Crop Disease Detection: A Survey of Trends, Architectures, and Challenges – *International Journal of Research and Innovation in Applied Science (IJRIAS).* https://rsisinternational.org/journals/ijrias/articles/edge-ai-and-iot-for-real-time-crop-disease-detection-a-survey-of-trends-architectures-and-challenges/

Mahato, S., & Neethirajan, S. (2025). Dairy DigiD: An Edge-Cloud Framework for Real-Time Cattle Biometrics and Health Classification. *AI, 6*(9), 196. https://doi.org/10.3390/ai6090196

Majeed, Y., Ojo, M. O., & Zahid, A. (2024). Standalone edge AI-based solution for Tomato diseases detection. *Smart Agricultural Technology, 9,* 100547. https://doi.org/10.1016/j.atech.2024.100547

Morales-García, J., Bueno-Crespo, A., Martínez-España, R., García, F. J., Ros, S., Fernández-Pedauyé, J., & Cecilia, J. M. (2023). SEPARATE: A tightly coupled, seamless IoT infrastructure for deploying AI algorithms in smart agriculture environments. *Internet of Things, 22,* 100734. https://doi.org/10.1016/j.iot.2023.100734

Morales-García, J., Bueno-Crespo, A., Martínez-España, R., Posadas, J.-L., Manzoni, P., & Cecilia, J. M. (2023). Evaluation of low-power devices for smart greenhouse development. *The Journal of Supercomputing, 79*(9), 10277–10299. https://doi.org/10.1007/s11227-023-05076-8

Navarro, E., Costa, N., & Pereira, A. (2020). A Systematic Review of IoT Solutions for Smart Farming. *Sensors, 20*(15), 4231. https://doi.org/10.3390/s20154231

Nyakuri, J. P., Nkundineza, C., Gatera, O., Nkurikiyeyezu, K., & Mwitende, G. (2025). AI and IoT-powered edge device optimized for crop pest and disease detection. *Scientific Reports, 15,* 22905. https://doi.org/10.1038/s41598-025-06452-5

Phukan, A. (2022). Hydroponics using IOT and Machine Learning. *International Research Journal of Engineering and Technology (IRJET), 9*(10). https://www.irjet.net/archives/V9/i10/IRJET-V9I1035.pdf

Proietti, M., Bianchi, F., Marini, A., Menculini, L., Termite, L., Garinei, A., Biondi, L., & Marconi, M. (2021). Edge Intelligence with Deep Learning in Greenhouse Management: *Proceedings of the 10th International Conference on Smart Cities and Green ICT Systems,* 180–187. https://doi.org/10.5220/0010451701800187

Qiao, H., Chen, Z., Ji, C., Gao, J., Xue, X., Zhou, S., Xu, H., & Huang, J. (2026). SmartEars: A practical framework for poultry respiratory monitoring via spectrogram-based audio classification and AI-assisted labeling.

Computers and Electronics in Agriculture, 241, 111241. https://doi.org/10.1016/j.compag.2025.111241

Ranjan, R., Sharrer, K., Tsukuda, S., & Good, C. (2023). MortCam: An Artificial Intelligence-aided fish mortality detection and alert system for recirculating aquaculture. *Aquacultural Engineering, 102*, 102341. https://doi.org/10.1016/j.aquaeng.2023.102341

Reddy, K. R. G., Thirunavukkarasu, K. S., & Sekhar, K. H. (2025). Edge AI and CNN for Real-Time Plant Disease Detection using IOT. *International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET), 14*(5). https://doi.org/10.15680/IJIRSET.2025.1405462

Salabi, L., & Manthila, P. (2025). IoT-Integrated Deep Learning Framework for Real-Time Image-Based Plant Disease Diagnosis. *National Journal of Signal and Image Processing,* 43–51. https://doi.org/10.17051/NJSIP/01.02.06

Salem, M. A., Perez, M. C., & Rabia, A. H. (2025). A TinyML Reinforcement Learning Approach for Energy-Efficient Light Control in Low-Cost Greenhouse Systems. *2025 Interdisciplinary Conference on Electrics and Computer (INTCEC),* 1–6. https://doi.org/10.1109/INTCEC65580.2025.11256135

Sanchez-Iborra, R., Zoubir, A., Hamdouchi, A., Idri, A., & Skarmeta, A. (2023). Intelligent and Efficient IoT Through the Cooperation of TinyML and Edge Computing. *Informatica,* 147–168. https://doi.org/10.15388/22-INFOR505

Silva, P. E. C. da, & Almeida, J. (2024). An Edge Computing-Based Solution for Real-Time Leaf Disease Classification using Thermal Imaging. *IEEE Geoscience and Remote Sensing Letters, 22,* 1–5. https://doi.org/10.1109/LGRS.2024.3456637

Smink, M., Liu, H., Döpfer, D., & Lee, Y. J. (2024). Computer Vision on the Edge: Individual Cattle Identification in Real-time with ReadMyCow System. *2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV),* 7041–7050. https://doi.org/10.1109/WACV57701.2024.00690

Srinivasagan, R., El Sayed, M. S., Al-Rasheed, M. I., & Alzahrani, A. S. (2025). Edge intelligence for poultry welfare: Utilizing tiny machine learning neural network processors for vocalization analysis. *PLOS ONE, 20*(1), e0316920. https://doi.org/10.1371/journal.pone.0316920

Suhaimi, A. F., Mutiara, G. A., & Fahru Rizal, M. (2025). Leveraging Embedded Accelerometers and Machine Learning for Real-Time Sheep Behavior Classification in Precision Farming. *IEEE Access, 13,* 165006–165024. https://doi.org/10.1109/ACCESS.2025.3612098

Taueatsoala, K., Daniels, C., Ramsunar, A. J., Bronkhorst, P., & Ezugwu, A. E. (2026). *TinyML-Enabled IoT for Sustainable Precision Irrigation* (2601.13054). arXiv. https://doi.org/10.48550/arXiv.2601.13054

Tricco, A., Lillie, E., Zarin, W., O'Brien, K., Colquhoun, H., Levac, D., Moher, D., Peters, M., Weeks, L., Hempel, S., Akl, E., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M., Garritty, C., & Straus, S. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Annals of Internal Medicine, 169*. https://doi.org/10.7326/M18-0850

Tudevdagva, U., & Shimada, T. (2023). AI-Powered Crop Disease Detection Using Edge Computing for Smallholder Farming Systems. *National Journal of Smart Agriculture and Rural Innovation (NJSARI), 1*. https://doi.org/10.17051/NJSARI/01.01.05

Ugli, O. E. M., & Lee, C.-H. (2025). Deep Learning-Based Optimal Condition Monitoring System for Plant Growth in an Indoor Smart Hydroponic Greenhouse. *Symmetry, 17*(7), 1092. https://doi.org/10.3390/sym17071092

Ugwu, O. P.-C., Ogenyi, F. C., Alum, E. U., Eze, V. H. U., Basajja, M., Ugwu, J. N., Ugwu, C. N., Ejemot-Nwadiaro, R. I., Okon, M. B., Egba, S. I., & Ejim, U. D. (2025). Implementing artificial intelligence and machine learning algorithms for optimized crop management: A systematic review on data-driven approach to enhancing resource use and agricultural sustainability. *Cogent Food & Agriculture, 11*(1), 2569982. https://doi.org/10.1080/23311932.2025.2569982

UN DESA. (2022). World Population Prospects 2022: Summary of Results | Population Division. https://www.un.org/development/desa/pd/content/World-Population-Prospects-2022

Venkataramanan, V., Pimpale, M., Kapure, V., Mishra, P., Rokade, A., Bhushan, T., & Singh, J. (2025). A Hybrid IoT and Machine Learning Framework for Smart Greenhouse Automation in Sustainable Agriculture. *International Research Journal of Multidisciplinary Technovation,* 58–69. https://doi.org/10.54392/irjmt2545

Verma, A., Bammidi, A., & Sarade, S. J. (2025). AquaAura: An Edge AI Framework for Predictive Soil Moisture Management Using Public and Field-Validated Datasets. *Journal of Emerging Technologies and Innovative Research (JETIR), 12*(9). https://www.jetir.org/papers/JETIR2509505.pdf

Wu, Z., Yang, J., Zhang, H., & Fang, C. (2025). Enhanced Methodology and Experimental Research for Caged Chicken Counting Based on YOLOv8. *Animals, 15*(6), 853. https://doi.org/10.3390/ani15060853

Xu, Y., Chen, Q., Kong, S., Xing, L., Wang, Q., Cong, X., & Zhou, Y. (2022). Real-time object detection method of melon leaf diseases under complex background in greenhouse. *Journal of Real-Time Image Processing, 19*(5), 985–995. https://doi.org/10.1007/s11554-022-01239-7

Yousif, A. J., Zaidan, F. K., & Ibrahim, N. J. (2025). A Portable AI-Driven Edge Solution for Automated Plant Disease Detection. *Diyala Journal of Engineering Sciences, 18*, 124–135. https://doi.org/10.24237/djes.2025.18308

Yu, P., Teng, F., Zhu, W., Shen, C., Chen, Z., & Song, J.

(2025). Cloud–edge–device collaborative computing in smart agriculture: Architectures, applications, and future perspectives. *Frontiers in Plant Science, 16,* 1668545. https://doi.org/10.3389/fpls.2025.1668545

Zha, W., Li, H., Wu, G., Zhang, L., Pan, W., Gu, L., Jiao, J., & Zhang, Q. (2023). Research on the Recognition and Tracking of Group-Housed Pigs' Posture Based on Edge Computing. *Sensors, 23*(21), 8952. https://doi.org/10.3390/s23218952

Zhang, Q., & Kanjo, E. (2025). A Multicore and Edge TPU-Accelerated Multimodal TinyML System for Livestock Behavior Recognition. *IEEE Internet of Things Journal, 13*(1), 666–677. https://doi.org/10.1109/JIOT.2025.3624811