iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** X **Month of publication:** October 2026

DOI: https://doi.org/10.22214/ijraset.2026.85452

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue X Oct 2026- Available at www.ijraset.com

Computational Botany: A Review of Machine Learning, IoT, and Bioinformatics in Plant Sciences

Nishchal Sharma[1], Dr. Harish Chand[2]
[1]*Department of Computer Science, Govt. College Kullu*
[2]*Department of Botany, Govt. College Kullu*

Abstract: ***The intersection of computer science and botany has catalyzed a paradigm shift in plant taxonomy, pathology, and agricultural management. This paper reviews the current literature on computational applications in plant sciences, focusing on three primary domains: Deep Learning for automated disease detection, Internet of Things (IoT) sensor networks for physiological monitoring, and high-performance computing in plant genomics. By synthesizing recent methodologies, this review identifies existing technological bottlenecks—such as dataset biases and processing latencies—and outlines future trajectories for integrating artificial intelligence into sustainable botanical research and precision agriculture.***
Keywords: Computational Botany, Precision Agriculture, Artificial Intelligence in Plant Sciences, Deep Learning & Computer Vision, Internet of Things (IoT), Bioinformatics & Transcriptomics, Edge Computing, Convolutional Neural Networks (CNNs), Plant Pathology & Disease Detection, High-Throughput Phenotyping, Plant Genomics, Microclimate Monitoring

I. INTRODUCTION

The study of plant sciences has historically relied on empirical observation, manual phenotyping, and labor-intensive biochemical assays. For centuries, botanical taxonomy, pathology, and ecological monitoring were fundamentally constrained by the physical limits of human processing capacity and manual data collection. However, the 21st century has witnessed an unprecedented explosion of biological data, driven by the pressing demands of global agricultural security and rapid ecological shifts resulting from climate change (Fiorani & Schurr, 2013[1]). This exponential increase in data complexity has catalyzed a foundational paradigm shift, necessitating the integration of advanced computer science methodologies into botanical research. Today, computational botany is not merely an auxiliary tool but a core driver of scientific discovery, bridging the gap between biological intricacy and scalable, data-driven modeling.

The convergence of these two disciplines is primarily anchored by three transformative technological pillars: Artificial Intelligence (AI), the Internet of Things (IoT), and high-performance Bioinformatics. Over the last decade, the integration of deep learning—specifically Convolutional Neural Networks (CNNs) and advanced computer vision architectures—has revolutionized high-throughput plant phenotyping and pathology (Kamilaris & Prenafeta-Boldú, 2018 [2]). Traditionally, diagnosing foliar diseases or assessing morphological traits required expert plant pathologists to physically examine specimens. Now, deep learning models, ranging from real-time object detection algorithms like YOLO (You Only Look Once) to deep residual networks (ResNets), can autonomously identify pathogens, assess nutrient deficiencies, and classify botanical species from digital images with accuracies frequently exceeding human expert baselines (Singh et al., 2016 [3]). Furthermore, the integration of edge computing allows these computationally heavy AI models to be deployed directly onto agricultural drones or mobile devices in remote, resource-constrained environments.

Simultaneously, the deployment of IoT architectures and Wireless Sensor Networks (WSNs) has radically transformed microclimate monitoring and physiological tracking. Modern botanical studies leverage distributed sensor nodes to capture high-frequency, real-time telemetry on variables such as soil moisture gradients, ambient thermodynamics, and photosynthetically active radiation (PAR) (Tzounis et al., 2017 [4]). These sensor arrays generate massive, continuous time-series datasets. When processed through cloud computing infrastructure and predictive machine learning algorithms, this data allows researchers to model plant stress responses dynamically and optimize precision agriculture frameworks.

Beyond macroscopic observation, computer science is indispensable at the molecular level. The advent of Next-Generation Sequencing (NGS) has generated terabytes of genomic, transcriptomic, and proteomic data for complex, highly polyploid plant species. High-performance computing and sophisticated bioinformatics pipelines, such as de Bruijn graph-based sequence assemblers, are now strictly required to map gene expression pathways and understand how plants adapt to abiotic stresses at a molecular level (Mochida & Shinozaki, 2010 [5]).

©IJRASET: All Rights are Reserved

Despite these remarkable advancements, the literature reveals critical bottlenecks that hinder the seamless integration of computer science and botany. Persistent challenges include the scarcity of geographically diverse, accurately annotated datasets, the "black box" nature of deep learning models that lack biological interpretability, and the steep learning curve preventing botanical researchers from deploying advanced computational frameworks.

This literature review aims to synthesize the current state-of-the-art at the intersection of these two fields. By systematically evaluating recent methodologies, this paper will map the trajectory of computational botany. The review is structured as follows: Section 2 critically examines the application of deep learning and computer vision in plant pathology and phenotyping. Section 3 explores the role of IoT and sensor networks in capturing real-time ecophysiological data. Section 4 delves into the computational algorithms driving modern bioinformatics and plant genomics. Finally, Section 5 discusses prevailing technological bottlenecks and outlines future research directions, emphasizing the need for interpretable algorithms and interdisciplinary frameworks to ensure global agricultural sustainability.

II. REVIEW OF LITERATURE

A. Deep Learning And Computer Vision In Plant Pathology

The identification and management of plant diseases remain critical to mitigating global crop yield losses, which currently range between 20% and 40% annually (Savary et al., 2019 [6]). Historically, diagnosing phytopathogens required destructive sampling and laboratory analysis, causing significant delays in treatment application. Deep learning (DL), a robust subset of machine learning, has effectively disrupted this paradigm by enabling non-destructive, rapid, and highly accurate automated visual diagnostics. Over the past decade, Computer Vision (CV) techniques—particularly Convolutional Neural Networks (CNNs)—have emerged as the dominant computational architecture for analyzing foliar lesions, pest infestations, and complex morphological abnormalities.

The efficacy of CNNs lies in their capacity to autonomously extract hierarchical spatial features, such as edges, textures, and color gradients, directly from raw pixel data, thereby eliminating the need for manual feature engineering by human experts. Early pioneering studies predominantly leveraged standardized, publicly available image repositories, most notably the PlantVillage dataset, to train deep architectures like VGG16, ResNet-50, and InceptionV3. For instance, Mohanty et al. (2016) [7] demonstrated that a fine-tuned CNN could classify 26 distinct diseases across 14 crop species with a diagnostic accuracy exceeding 99% under controlled conditions. Advancing beyond simple binary or multi-class image classification, contemporary literature focuses heavily on object detection algorithms such as YOLO (You Only Look Once) and Faster R-CNN. These models allow for the precise bounding-box localization of multiple, simultaneous disease symptoms on a single leaf or across overlapping crop canopies in agricultural fields (Barbedo, 2019 [8]).

A significant current trajectory in the literature is the migration of these computational models from centralized cloud servers to localized edge devices. To ensure diagnostic tools are accessible to field botanists and farmers, researchers are increasingly compressing and deploying lightweight neural networks, such as MobileNet and EfficientNet, directly onto mobile application frameworks and Unmanned Aerial Vehicles (UAVs). This shift toward edge computing mitigates processing latency and removes the strict requirement for high-bandwidth internet connectivity, enabling real-time, in-field phenotyping in remote regions (Chen et al., 2020 [9]).

Despite the exceptionally high accuracies reported in controlled studies, the literature consistently identifies critical bottlenecks that hinder seamless real-world deployment. The most prominent challenge is the "domain shift" problem: models trained exclusively on homogeneous datasets with uniform backgrounds frequently suffer severe performance degradation when exposed to the complex, cluttered, and variably illuminated environments of natural field conditions (Ferentinos, 2018 [10]). Furthermore, deep neural networks operate essentially as computational "black boxes." In botanical pathology, understanding the biological reasoning behind a model's prediction is vital for scientific validation. Consequently, recent literature emphasizes the integration of Explainable AI (XAI) techniques. Tools like Gradient-weighted Class Activation Mapping (Grad-CAM) are now being utilized to visually highlight the specific phenotypic symptoms—such as a necrotic halo or fungal sporulation—that triggered the network's diagnostic decision, thereby restoring biological interpretability to the computational process (Toda & Okura, 2019 [11]).

B. IoT and Sensor Networks for Microclimate Monitoring

Understanding the intricate relationship between a plant and its immediate environment is a cornerstone of botanical research. Historically, tracking ecophysiological responses required manual, point-in-time measurements that failed to capture the dynamic, diurnal fluctuations of natural ecosystems.

The advent of the Internet of Things (IoT) and Wireless Sensor Networks (WSNs) has fundamentally resolved this limitation, providing continuous, high-resolution telemetry that directly links environmental stimuli to physiological outcomes (Tzounis et al., 2017 [4]). By embedding connected computational nodes into agricultural fields and forest canopies, researchers can now conduct longitudinal studies on plant-environment interactions with unprecedented granularity.

The standard IoT architecture in computational botany consists of distributed edge nodes equipped with specialized transducers. For example, dielectric sensors measure soil volumetric water content, while micro-meteorological stations track ambient temperature, relative humidity, and Photosynthetically Active Radiation (PAR). This continuous data stream is crucial for modeling botanical outcomes such as stomatal conductance, transpiration rates, and photosynthetic efficiency (Navarro et al., 2020 [12]). When a WSN detects a rapid decline in soil moisture coupled with high vapor pressure deficit (VPD), cloud-based algorithms can predict the onset of drought-induced stomatal closure before visible wilting occurs. This predictive capability is transforming precision agriculture, shifting irrigation management from reactive scheduling to proactive, physiological need-based application.

A significant focus of recent literature is the transmission and aggregation of this environmental data over expansive, topographically complex terrains. Traditional Wi-Fi or Bluetooth protocols are insufficient for large-scale agricultural or forestry applications due to their limited range and high energy consumption. Consequently, researchers have widely adopted Long Range Wide Area Network (LoRaWAN) protocols. LoRaWAN facilitates the transmission of small data packets over several kilometres with minimal power requirements, making it ideal for mapping spatial variability in microclimates (Ray, 2017 [13]). This spatial mapping allows botanists to deeply analyze Genotype-by-Environment (GxE) interactions, revealing how localized microclimate variations affect the phenotypic expression of different plant cultivars across a single geographic site.

Despite the transformative potential of IoT in plant sciences, the literature highlights several persistent engineering challenges. Deploying computational hardware in harsh outdoor environments often leads to sensor drift, biofouling, and hardware degradation, compromising data integrity over time. Furthermore, powering remote WSN nodes remains a critical bottleneck. While current research is exploring energy-harvesting technologies—such as miniature photovoltaic panels and thermoelectric generators—power management algorithms that dynamically adjust sensor sampling rates based on battery life and environmental priority are essential (Oksanen et al., 2015 [14]). Lastly, the sheer volume of continuous time-series data generated by these networks necessitates robust cloud infrastructure and advanced machine learning algorithms, such as Long Short-Term Memory (LSTM) networks, to filter noise, impute missing values, and extract biologically meaningful trends from the raw telemetry.

C. Bioinformatics and Plant Genomics

While deep learning and IoT operate at the organismal and ecological levels, the intersection of computer science and botany is arguably most profound at the molecular level. Plant genomes are notoriously complex; they are frequently characterised by massive sizes, high degrees of heterozygosity, extensive repetitive elements, and complex polyploidy (Michael & Jackson, 2013 [15]). The advent of Next-Generation Sequencing (NGS) and third-generation long-read technologies has generated an exponential influx of biological data, fundamentally shifting the primary bottleneck in botanical research from physical DNA sequencing to computational data analysis and algorithmic processing.

The de novo assembly of plant genomes relies entirely on advanced computer science algorithms, primarily rooted in graph theory. Short-read sequencing data (such as Illumina) produces millions of fragmented nucleotide sequences that must be computationally stitched together. To solve this, bio-informatics literature heavily features the application of de Bruijn graph algorithms, utilized in assemblers like SOAPdenovo and Velvet, which map overlaps between small k-mers to reconstruct the original sequence (Li et al., 2010 [16]). More recently, the integration of long-read sequencing (e.g., PacBio and Oxford Nanopore) has revived Overlap-Layout-Consensus (OLC) algorithms. These computational models are critical for spanning highly repetitive regions in plant genomes—such as the massive hexaploid wheat genome—allowing botanists to identify crucial structural variations and evolutionary adaptations that were previously computationally invisible.

Beyond static genomic sequences, computational botany heavily explores dynamic gene expression through transcriptomics (RNA-Seq). By computationally mapping millions of RNA transcripts to a reference genome, researchers can quantify how a plant alters its genetic expression in response to environmental stimuli, such as drought, soil salinity, or pathogen attacks. The RNA-Seq pipeline relies on highly optimized string-matching algorithms, such as the Burrows-Wheeler Transform used in alignment tools like STAR or Bowtie (Conesa et al., 2016 [17]). Furthermore, the downstream analysis of this data utilizes rigorous statistical modelling and machine learning to perform Differential Gene Expression (DGE) analysis. Through these computational pipelines, botanists can pinpoint the exact gene networks responsible for stress tolerance, directly informing molecular breeding programs and the development of climate-resilient crops.

Despite these computational achievements, the literature points to a rapidly approaching horizon of new challenges. The botanical community is currently transitioning from single reference genomes to "pan-genomics"—the computational representation of the entire genetic diversity within a species. Constructing and querying graphical pan-genomes requires vastly more efficient data structures to handle the memory-intensive graph alignments (Bayer et al., 2020 [18]). Additionally, there is a pressing need for machine learning frameworks capable of "multi-omics" integration: seamlessly fusing genomic, transcriptomic, and proteomic data with the real-time phenotypic and IoT data discussed in previous sections to create holistic, predictive models of plant biology.

III. CONCLUSION

The synthesis of computer science and botany represents a fundamental paradigm shift in how humanity understands, monitors, and manipulates plant life. As this literature review demonstrates, computational tools are no longer merely auxiliary accessories to traditional botanical research; they have become the foundational infrastructure required to navigate the complexities of modern biology. The transition from manual, empirical observation to high-throughput, data-driven analysis is rapidly accelerating across all scales of plant science.

At the organismal level, deep learning and computer vision architectures have revolutionised plant pathology, moving diagnostics from the laboratory directly to the field via edge computing. At the ecological level, the deployment of IoT sensor networks and LoRaWAN protocols has provided an unprecedented, real-time window into ecophysiology, enabling precision agriculture to proactively respond to micro climate fluctuations. Simultaneously, at the molecular level, advanced bio-informatics algorithms and graph theory are untangling the immense complexity of polyploid genomes and mapping dynamic transcriptomic responses to environmental stress.

However, realising the full potential of computational botany requires addressing several critical bottlenecks. The literature consistently highlights the need to overcome the "domain shift" in field-deployed machine learning, the demand for Explainable AI to restore biological interpretability to neural networks, and the necessity of robust computational infrastructure to support the incoming wave of multi-omics and pan-genomic data.

Ultimately, the future of global food security, climate change adaptation, and sustainable agriculture depends on the continued convergence of these disciplines. By fostering interdisciplinary collaboration and developing scalable, accessible computational frameworks, researchers can transform the exponential influx of biological data into actionable insights, ensuring the resilience of plant ecosystems in a rapidly changing world.

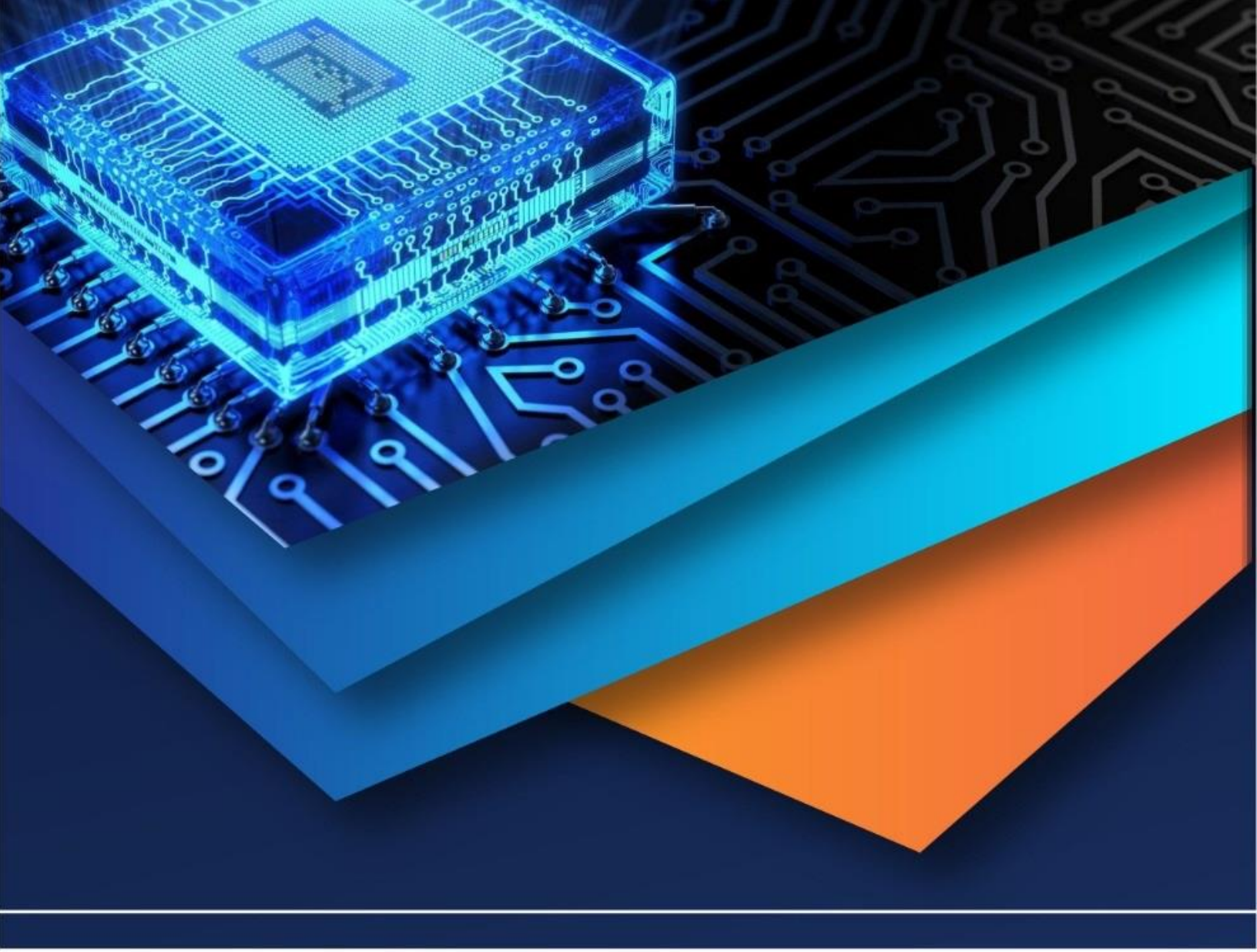
REFERENCES

[1] F. Fiorani and U. Schurr, “Future Scenarios for Plant Phenotyping,” Annu. Rev. Plant Biol., vol. 64, pp. 267–291, 2013, doi: 10.1146/annurev-arplant-050312-120137.
[2] A. Kamilaris and F. X. Prenafeta-Boldú, “Deep learning in agriculture: A survey,” Comput. Electron. Agric., vol. 147, pp. 70–90, Apr. 2018, doi: 10.1016/j.compag.2018.02.016.
[3] A. Singh, B. Ganapathysubramanian, A. K. Singh, and S. Sarkar, “Machine Learning for High-Throughput Stress Phenotyping in Plants,” Trends Plant Sci., vol. 21, no. 2, pp. 110–124, Feb. 2016, doi: 10.1016/j.tplants.2015.10.015.
[4] A. Tzounis, N. Katsoulas, T. Bartzanas, and C. Kittas, “Internet of Things in agriculture, recent advances and future challenges,” Biosyst. Eng., vol. 164, pp. 31–48, Dec. 2017, doi: 10.1016/j.biosystemseng.2017.09.007.
[5] “Genomics and Bioinformatics Resources for Crop Improvement | Plant and Cell Physiology | Oxford Academic.” Accessed: Oct. 27, 2026. [Online]. Available: https://academic.oup.com/pcp/article/51/4/497/1828103
[6] S. Savary, L. Willocquet, S. J. Pethybridge, P. Esker, N. McRoberts, and A. Nelson, “The global burden of pathogens and pests on major food crops,” Nat. Ecol. Evol., vol. 3, no. 3, pp. 430–439, Mar. 2019, doi: 10.1038/s41559-018-0793-y.
[7] S. P. Mohanty, D. P. Hughes, and M. Salathé, “Using Deep Learning for Image-Based Plant Disease Detection,” Front. Plant Sci., vol. 7, Sep. 2016, doi: 10.3389/fpls.2016.01419.
[8] J. G. Arnal Barbedo, “Plant disease identification from individual lesions and spots using deep learning,” Biosyst. Eng., vol. 180, pp. 96–107, Apr. 2019, doi: 10.1016/j.biosystemseng.2019.02.002.
[9] J. Chen, J. Chen, D. Zhang, Y. Sun, and Y. A. Nanehkaran, “Using deep transfer learning for image-based plant disease identification,” Comput. Electron. Agric., vol. 173, p. 105393, Jun. 2020, doi: 10.1016/j.compag.2020.105393.
[10] K. P. Ferentinos, “Deep learning models for plant disease detection and diagnosis,” Comput. Electron. Agric., vol. 145, pp. 311–318, Feb. 2018, doi: 10.1016/j.compag.2018.01.009.
[11] “How Convolutional Neural Networks Diagnose Plant Disease | Plant Phenomics.” Accessed: Oct. 27, 2026. [Online]. Available: https://spj.science.org/doi/full/10.34133/2019/9237136
[12] E. Navarro, N. Costa, and A. Pereira, “A Systematic Review of IoT Solutions for Smart Farming,” Sensors, vol. 20, no. 15, p. 4231, Aug. 2020, doi: 10.3390/s20154231.
[13] P. P. Ray, “Internet of things for smart agriculture: Technologies, practices and future direction,” J. Ambient Intell. Smart Environ., vol. 9, no. 4, pp. 395–420, Jun. 2017, doi: 10.3233/AIS-170440.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

[14] M. O. Yerebakan and B. Hu, "Human–Robot Collaboration in Modern Agriculture: A Review of the Current Research Landscape," Adv. Intell. Syst., vol. 6, no. 7, p. 2300823, 2024, doi: 10.1002/aisy.202300823.
[15] T. P. Michael and S. Jackson, "The First 50 Plant Genomes," Plant Genome, vol. 6, no. 2, p. plantgenome2013.03.0001in, 2013, doi: 10.3835/plantgenome2013.03.0001in.
[16] R. Li et al., "De novo assembly of human genomes with massively parallel short read sequencing," Genome Res., vol. 20, no. 2, pp. 265–272, Feb. 2010, doi: 10.1101/gr.097261.109.
[17] A. Conesa et al., "A survey of best practices for RNA-seq data analysis," Genome Biol., vol. 17, no. 1, p. 13, Jan. 2016, doi: 10.1186/s13059-016-0881-8.
[18] P. E. Bayer, A. A. Golicz, A. Scheben, J. Batley, and D. Edwards, "Plant pan-genomes are the new reference," Nat. Plants, vol. 6, no. 8, pp. 914–920, Aug. 2020, doi: 10.1038/s41477-020-0733-0.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)