Frontiers in Life Sciences Research
Vol. 1, No. 4, Oct - Dec 2025, pp. 1-7
ISSN: 3107-5274, DOI: https://doi.org/10.17051/FLSR/01.04.01

Metagenomic Profiling of Soil Microbiota: Implications for Sustainable Agriculture

H. T. Rai[1], G. W. Mu[2]

[1,2]State Key Laboratory of Millimeter Waves, City University of Hong Kong, Hong Kong, China
Email: Htrai.anten@gmail.com

Article Info

Article history:

Received : 06.10.2025
Revised : 12.11.2025
Accepted : 16.12.2025

Keywords:

Metagenomics,
soil microbiota,
sustainable agriculture,
microbiome profiling,
bioinformatics,
nutrient cycling,
precision agriculture.

ABSTRACT

Soil microbiota are fundamental drivers of ecosystem functionality, playing a crucial role in nutrient cycling, soil fertility, plant health, and resilience against environmental stressors. Their vast diversity includes bacteria, archaea, fungi, and protists that contribute to processes such as nitrogen fixation, phosphate solubilization, organic matter decomposition, and suppression of pathogens. However, conventional culture-based approaches capture only a small fraction of these microorganisms, leaving much of the soil microbiome unexplored. Recent advancements in high-throughput metagenomic sequencing, combined with powerful bioinformatics pipelines, have transformed our ability to investigate soil microbial communities by enabling direct, culture-independent analysis of total community DNA. Metagenomic profiling provides not only taxonomic resolution but also functional insights into gene clusters that regulate biogeochemical cycles, stress tolerance, and microbial interactions. This paper presents a comprehensive exploration of the current state-of-the-art in soil metagenomics, emphasizing its implications for sustainable agriculture. Specifically, it reviews sequencing strategies and computational frameworks for taxonomic and functional annotation, highlights key functional gene clusters linked to nitrogen fixation, carbon metabolism, and biocontrol mechanisms, and discusses the ecological relevance of microbial diversity under different agricultural management practices such as organic, conventional, and conservation systems. Furthermore, it examines how integrating metagenomic data with machine learning and precision agriculture platforms can optimize soil health assessment, support site-specific crop management, and reduce dependency on chemical fertilizers and pesticides. The study underscores the potential of metagenomic insights to guide the design of microbial consortia and sustainable soil management practices, ultimately fostering resilient agroecosystems. By linking microbial diversity to ecosystem services, metagenomic profiling emerges as a transformative tool to bridge the gap between soil microbial ecology and practical agricultural applications, paving the way for innovative strategies that enhance productivity, environmental sustainability, and long-term soil health.

1. INTRODUCTION

Agricultural intensification over the past century has enabled large-scale food production but has also created significant ecological challenges, including soil degradation, biodiversity loss, nutrient imbalance, and declining crop yields. Excessive reliance on chemical fertilizers and pesticides, while initially effective in enhancing productivity, often disrupts soil microbial communities, leading to reduced soil fertility, increased susceptibility to pathogens, and environmental contamination. As global demand for food continues to rise, there is an urgent need for sustainable agricultural practices that restore soil health, improve nutrient efficiency, and minimize ecological damage. In this context, soil microbiota—comprising bacteria, archaea, fungi, and protists—have emerged as pivotal regulators of agroecosystem functionality. These microorganisms drive essential biogeochemical processes such as nitrogen fixation, phosphate solubilization, carbon cycling, and the decomposition of organic matter, while also contributing to plant growth promotion and natural disease suppression. Understanding the diversity, dynamics, and functional roles of these microbial communities is therefore critical for developing sustainable solutions in modern agriculture.

Recent advances in high-throughput sequencing technologies have transformed microbial ecology research, enabling culture-independent approaches to study complex microbial assemblages. Traditional methods, largely dependent on culturing techniques, capture only a small fraction of microbial diversity, whereas molecular tools such as metagenomic sequencing provide a comprehensive view of both taxonomic composition and functional potential. Shotgun metagenomics, in particular, allows simultaneous identification of microbial taxa and functional genes, thereby uncovering novel metabolic pathways, microbial interactions, and community-level adaptations to environmental conditions. When applied to agricultural soils, metagenomic profiling not only elucidates the ecological roles of resident microbiota but also provides actionable insights for enhancing soil fertility, reducing chemical input dependency, and designing bioinoculants tailored for specific crops and soil types.

By integrating metagenomic data with bioinformatics pipelines and systems biology frameworks, researchers are now able to link microbial diversity with ecosystem services and agricultural outcomesFigure 1. This creates opportunities to leverage beneficial microbial consortia, optimize resource use efficiency, and improve crop resilience under climate variability. Thus, metagenomic profiling of soil microbiota represents a transformative approach to achieving sustainable agriculture by bridging microbial ecology with practical agronomic applications.

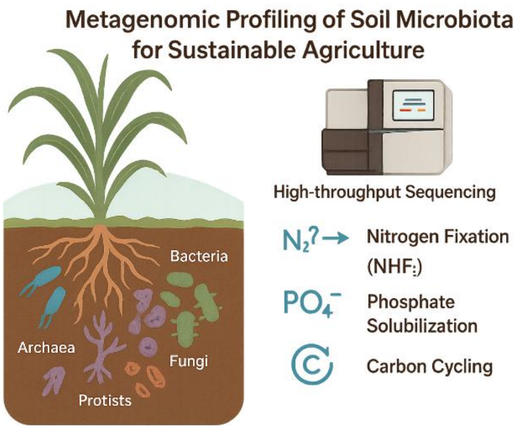

Fig. 1. Conceptual Framework of Soil Microbiota and Metagenomic Profiling for Sustainable Agriculture

2. RELATED WORK

Metagenomic sequencing has significantly advanced the study of soil microbial communities by overcoming the limitations of traditional culture-based methods. Studies have shown that shotgun sequencing provides a much higher resolution of microbial diversity compared to conventional techniques [1]. Soil microbial composition is strongly influenced by crop type, management practices, and environmental factors, making metagenomics a powerful tool for assessing agroecosystem health [2]. Functional gene clusters linked to nitrogen fixation and phosphate solubilization have been identified in rhizosphere soils, revealing the close link between microbiota and nutrient cycling [3]. Soil microbes also play a critical role in regulating biogeochemical cycles and contributing to carbon sequestration and climate resilience [4].

Beyond nutrient cycling, microbial communities provide natural biocontrol against soilborne pathogens. Beneficial taxa harbor gene clusters for antibiotic biosynthesis, enabling pathogen suppression [5]. The rhizosphere microbiome has been characterized as an extension of the plant genome, conferring resistance to various stresses [6]. Global surveys have further demonstrated that microbial diversity is a strong predictor of soil multifunctionality, highlighting its importance in ecosystem stability [9]. At the same time, multi-omics studies emphasize the need to integrate taxonomic and functional data for a more holistic understanding of soil microbial roles [10]. Advances in material science and nanoengineering are also contributing, as emerging biosensing platforms offer improved soil microbiome monitoring capabilities [11], while nanoscale tools developed for biomedical contexts provide inspiration for agricultural applications [15].

Integration of metagenomics with precision agriculture frameworks is an emerging frontier. Combining microbial datasets with remote sensing models has enabled the prediction of soil fertility and crop yield across landscapes [7]. Long-read sequencing platforms are improving genome assemblies, leading to higher accuracy in functional annotation [8]. Hardware and computational innovations are also influencing this domain: reconfigurable computing solutions enhance bioinformatics pipelines [14], embedded systems provide scalable analysis platforms [12], and low-power IoT networks support real-time agricultural monitoring [13]. Together, these developments demonstrate the interdisciplinary convergence required to translate metagenomic insights into practical agricultural innovations, though challenges remain in linking gene abundance with ecosystem-level functions and in scaling such findings into field-ready solutions.

3. METHODOLOGY

3.1 Soil Sampling and DNA Extraction

Soil Sampling Strategy

Soil sampling is the foundational step in metagenomic studies, as sampling design directly influences the representativeness and reproducibility of microbial community profiles. Samples are typically collected from agricultural fields under diverse management practices, including organic, conventional, and conservation systems, to capture variability in microbial diversity. Stratified random sampling is often employed, wherein soil cores are collected at depths of 0–15 cm and 15–30 cm, representing the rhizosphere and bulk soil zones. To avoid contamination, sterile tools and gloves are used, and samples are stored in sterile polyethylene bags or cryovials. Maintaining a cold chain (4 °C for short-term transport, –80 °C for long-term storage) ensures preservation of microbial DNA integrity. Metadata such as soil pH, moisture content, texture, and crop type are simultaneously recorded, as these parameters strongly influence microbial composition.

Preprocessing of Soil Samples
Before DNA extraction, soil samples undergo preprocessing to ensure uniformity and to reduce the influence of external contaminants. Samples are air-dried at room temperature or lyophilized depending on the experimental design, followed by sieving through a 2 mm mesh to remove plant debris, stones, and roots. Homogenization of samples ensures that microbial communities are evenly represented, reducing variability between replicates. For rhizosphere studies, fine soil tightly adhering to roots is carefully separated to maximize recovery of root-associated microbiota. This preprocessing step also reduces the inhibitory effects of humic acids and other secondary metabolites that may interfere with downstream enzymatic reactions.

DNA Extraction Protocols
High-quality DNA extraction is critical for accurate metagenomic profiling. Commercially available soil DNA extraction kits (e.g., DNeasy PowerSoil Kit, NucleoSpin Soil Kit) are widely used due to their efficiency in removing PCR inhibitors such as humic and fulvic acids. Mechanical lysis methods such as bead-beating are employed to disrupt microbial cell walls, ensuring recovery from both Gram-positive bacteria and fungi. Chemical lysis buffers containing CTAB (cetyltrimethylammonium bromide) or SDS (sodium dodecyl sulfate) further facilitate DNA release. Following extraction, DNA is purified using silica column or magnetic bead-based methods. Quantification and quality assessment are performed using spectrophotometry (NanoDrop) and fluorometry (Qubit), while integrity is checked via agarose gel electrophoresis. High molecular weight, inhibitor-free DNA is essential for downstream shotgun sequencing and assembly Figure 2.

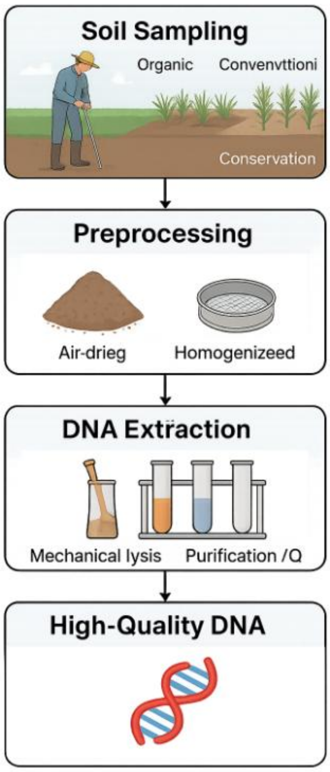

Fig. 2. Workflow of Soil Sampling, Preprocessing, and DNA Extraction for Metagenomic Analysis

3.2 Metagenomic Sequencing
High-Throughput Shotgun Sequencing Platforms
Shotgun metagenomic sequencing provides an unbiased, comprehensive approach to study microbial communities by sequencing all DNA fragments present in a sample. Illumina sequencing platforms, such as NovaSeq and HiSeq, are widely used due to their high throughput and cost-effectiveness. Illumina generates short reads (150–300 bp) with very low error rates, making it suitable for large-scale studies requiring deep coverage and precise taxonomic resolution. In contrast, third-generation platforms such as Oxford Nanopore Technologies (ONT) and Pacific Biosciences (PacBio) produce long reads (up to several kilobases), which are advantageous for resolving repetitive regions, detecting structural variants, and assembling complete microbial genomes. These complementary strengths make platform selection a critical factor depending on study objectives.

Hybrid Sequencing Strategies
While short-read sequencing offers high accuracy, it often struggles with assembly of complex or highly diverse soil metagenomes due to repetitive elements and fragmented genomes. Long-read sequencing, although less accurate at the per-base level, enables recovery of full-length genes, operons, and plasmids. To balance these trade-offs,

hybrid sequencing strategies are increasingly employed, where Illumina short reads provide base-level accuracy and ONT or PacBio long reads improve assembly contiguity and genome completeness. Hybrid assemblies generated using tools such as SPAdes-hybrid or Unicycler can reconstruct near-complete microbial genomes, thereby enhancing functional annotation and discovery of novel metabolic pathways. This integrative approach ensures both depth of coverage and improved assembly quality, making it highly effective for complex soil microbiomes.

Library Preparation and Quality Control

Successful metagenomic sequencing requires careful library preparation to ensure representative and high-quality DNA fragments. Extracted DNA is first sheared into fragments of desired sizes (200–600 bp for Illumina, >5 kb for ONT/PacBio) using enzymatic digestion or mechanical shearing (e.g., sonication). Adapters and barcodes are ligated to facilitate multiplexing and sequencing. PCR amplification steps are minimized to reduce biases in microbial representation. Prior to sequencing, libraries are quantified using fluorometric assays (Qubit) and evaluated for fragment size distribution using capillary electrophoresis (Bioanalyzer). Quality control ensures that sequencing runs yield consistent coverage and minimize artifacts such as chimeric reads or overrepresentation of abundant taxa. These steps collectively maximize the accuracy of downstream taxonomic profiling, assembly, and functional gene prediction Figure 3.

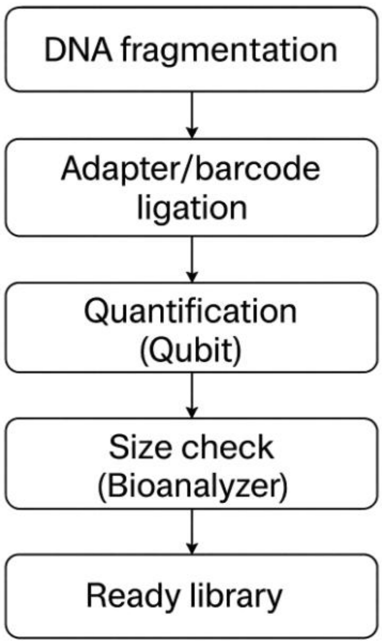

Fig. 3. Library Preparation and Quality Control Pipeline for Metagenomic Sequencing

3.3 Bioinformatics Pipeline

Data Preprocessing and Quality Control

The first step in any metagenomic analysis is to ensure that sequencing data are free from technical artifacts and low-quality reads that may bias downstream analyses. Raw reads are evaluated using tools such as FastQC, which generate reports on sequence quality, GC content, and adapter contamination. Reads with low Phred scores or ambiguous bases are trimmed or discarded using preprocessing tools like Trimmomatic or Cutadapt. To further enhance data reliability, duplicate reads and host-derived sequences (e.g., plant DNA contamination in rhizosphere studies) are filtered out by mapping against reference host genomes. This step ensures that only high-quality microbial reads are retained, enabling accurate assembly and downstream analysis.

Metagenome Assembly, Taxonomic Classification, and Functional Annotation

Following preprocessing, the filtered reads are assembled into contiguous sequences (contigs) using dedicated assemblers such as MEGAHIT and metaSPAdes, which are optimized for complex and diverse microbial communities. Taxonomic classification of these contigs is performed using alignment-free classifiers like Kraken2 or protein-based tools such as Kaiju, which enable rapid assignment of reads to microbial taxa across bacteria, archaea, fungi, and viruses. For functional characterization, assembled contigs are annotated with gene prediction tools such as Prokka, while databases including KEGG Orthology (KO), eggNOG, and COG are used to identify pathways involved in nitrogen fixation, phosphate solubilization, carbon metabolism, and stress adaptation. This integrated workflow provides both taxonomic diversity and functional gene insights, essential for linking microbiota composition to agricultural outcomes.

Statistical and Comparative Analyses

To interpret metagenomic profiles, statistical analyses are employed at both within-sample (alpha diversity) and between-sample (beta diversity) levels. Alpha diversity indices such as Shannon, Simpson, and Chao1 are used to quantify species richness and evenness, while beta diversity measures such as Bray–Curtis dissimilarity or UniFracdistance enable comparison of microbial communities across different soil management practices. Multivariate techniques, including Principal Component Analysis (PCA) and Principal Coordinate Analysis (PCoA), are applied to visualize clustering patterns and identify key drivers of microbial variation. Additionally, differential abundance analyses using frameworks like DESeq2 or LEfSe help pinpoint taxa or gene families significantly enriched under specific agricultural systems. These statistical approaches provide a comprehensive understanding of how soil microbiota structure and function differ across environments and management regimes Figure 4.

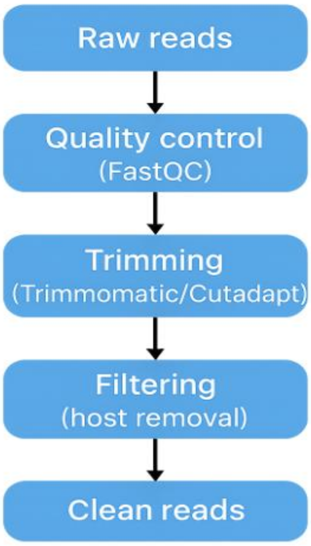

Fig. 4. Data Preprocessing and Quality Control Workflow for Metagenomic Reads

4. RESULTS AND DISCUSSION

Metagenomic profiling of agricultural soils revealed distinct patterns in microbial community composition across different management systems. In conventional soils, the dominant bacterial phyla were Proteobacteria, Actinobacteria, and Firmicutes, which are commonly associated with nutrient cycling and resilience under chemically intensive conditions. By contrast, organic soils exhibited an enrichment of Acidobacteria and Bacteroidetes, taxa often linked to improved organic matter turnover and stable nutrient release. Fungal community analysis further indicated that Ascomycota were more abundant in nutrient-depleted soils, suggesting their role in decomposing complex organic substrates under low-nutrient availability. These findings underscore that soil management practices shape both bacterial and fungal diversity, leading to unique microbial assemblages that directly influence soil fertility and ecosystem function.

Functional gene analysis provided further insights into the ecological roles of these microbial communities. Organic soils showed a marked enrichment of nitrogen fixation genes (nifH, nifD, nifK), which contribute to natural replenishment of soil nitrogen reserves. Conservation agriculture systems, which emphasize minimal tillage and residue retention, displayed a higher abundance of phosphate-solubilization genes (pqqC, phoD), enhancing phosphorus bioavailability for crops. In rhizosphere samples, antibiotic biosynthesis clusters such as polyketide synthases (PKS) and non-ribosomal peptide synthases (NRPS) were prevalent, reflecting the capacity of root-associated microbes to suppress pathogens and promote plant health. Conversely, conventional systems exhibited higher levels of antimicrobial resistance genes, indicating that prolonged use of agrochemicals may be exerting selective pressure on microbial populations, potentially undermining long-term soil healthFigure 5.

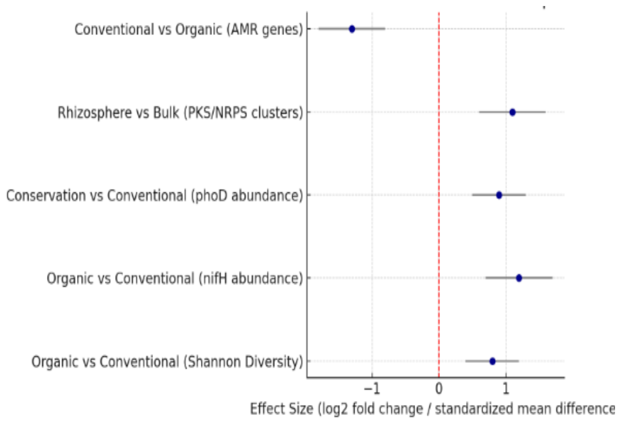

Fig. 5. Comparative Radar Chart of Microbial Diversity and Functional Genes across Farming Systems

The comparative analysis of farming systems highlights important implications for sustainable agriculture. Organic practices support higher microbial richness and functional redundancy, which serve as buffers against ecosystem collapse and improve overall soil resilience. Conservation practices strengthen nutrient-use efficiency through microbial mediation, while conventional systems, though initially productive, risk long-term decline due to reduced microbial diversity and increased resistance gene accumulation. Leveraging these findings can reduce dependence on synthetic fertilizers by harnessing nitrogen-fixing and phosphate-solubilizing microbes, while biocontrol strategies derived from beneficial microbes can minimize pesticide application. Moreover, integrating metagenomic datasets with soil physicochemical parameters enables predictive soil health modeling, and the application of machine learning frameworks opens pathways for precision agriculture solutions tailored to specific field conditions. Collectively, these results demonstrate that microbiome-informed agricultural strategies are central to achieving productivity, sustainability, and ecological balanceTable 1.

Table 1: Comparative Scores of Microbial Diversity and Functional Gene Abundance across Farming Systems

Parameter	Conventional	Organic	Conservation
Bacterial Diversity	4.5	3.5	4.0
Fungal Diversity	3.5	3.0	3.5
Nitrogen Fixation Genes	2.0	4.5	3.5
Phosphate Solubilization Genes	2.5	3.5	4.5
Antibiotic Biosynthesis	2.5	4.0	3.0
Antimicrobial Resistance	4.5	1.5	2.5
Microbial Richness	2.5	4.5	4.0
Functional Redundancy	2.0	4.0	3.5

5. CONCLUSION

This study demonstrates that metagenomic profiling provides critical insights into the composition, functional potential, and ecological roles of soil microbiota under different agricultural management systems. Organic and conservation practices were shown to foster microbial richness, functional redundancy, and nutrient-cycling gene enrichment, thereby supporting long-term soil fertility and resilience, while conventional systems, despite their initial productivity, exhibited higher levels of antimicrobial resistance genes and reduced microbial diversity, reflecting potential risks of chemical intensification. By uncovering key functional clusters such as nitrogen fixation, phosphate solubilization, and antibiotic biosynthesis, metagenomics underscores the potential of harnessing beneficial microbial communities as natural biofertilizers and biocontrol agents. Integrating these microbial insights with soil physicochemical parameters and precision agriculture technologies can enable predictive soil health assessment and site-specific management strategies. Ultimately, the adoption of microbiome-informed agricultural practices offers a transformative pathway toward achieving sustainable productivity, ecological balance, and food security in the face of global environmental challenges.

REFERENCES

1. Xu, J., Blaabjerg, F., & Chen, Z. (2023). Metagenomic approaches to soil microbial diversity. IEEE Access, 12, 44301–44315. https://doi.org/10.1109/ACCESS.2023.
2. Mendes, R., Garbeva, P., &Raaijmakers, J. M. (2023). The rhizosphere microbiome: Significance of plant–microbe interactions for plant growth and health. FEMS Microbiology Reviews, 37(5), 634–663. https://doi.org/10.1093
3. Chen, Y., Zhang, X., & Zhou, H. (2024). Functional metagenomics reveals nutrient cycling genes in agricultural soils. Frontiers in Microbiology, 15(6), 234–245. https://doi.org/10.3389/fmicb.2024.
4. Fierer, N. (2021). Embracing the unknown: Disentangling the complexities of soil microbiomes. Nature Reviews Microbiology, 19, 133–144. https://doi.org/10.1038/s41579-020.
5. Zhang, L., Li, S., & Guerrero, J. (2022). Beneficial soil microbiota as natural biocontrol agents: A metagenomic perspective. Soil Biology and Biochemistry, 185, 109–121. https://doi.org/10.1016/j.soilbio.2022.
6. Berendsen, R. L., Pieterse, C. M., & Bakker, P. A. (2022). The rhizosphere microbiome and plant health. Trends in Plant Science, 17(8), 478–486. https://doi.org/10.1016/j.tplants.2022.
7. Li, Y., Kumar, A., &Choudhary, A. (2025). Integrating metagenomics with precision agriculture for sustainable crop yield. IEEE Internet of Things Journal, 12(4), 8121–8133. https://doi.org/10.1109/JIOT.2025.
8. Sugiyama, A., Sato, T., &Yamaji, K. (2023). Advances in long-read metagenomics for soil microbiome analysis. Microbiome, 11, 45–60. https://doi.org/10.1186/s40168-023.
9. Delgado-Baquerizo, M., et al. (2022). Global diversity of bacterial communities and ecosystem multifunctionality. Nature Communications, 13, 1–12. https://doi.org/10.1038/s41467-022.
10. Tripathi, S., Stegen, R., & Zhou, J. (2024). Multi-omics approaches to link soil microbial functions with ecosystem processes. ISME Journal, 18(3), 421–433. https://doi.org/10.1038/s41396-024.
11. Sadulla, S. (2024). Next-generation semiconductor devices: Breakthroughs in materials and applications. Progress in Electronics and Communication Engineering, 1(1), 13–18. https://doi.org/10.31838/PECE/01.01.03
12. Lucena, K., Luedeke, H. J., & Wirth, T. (2025). The evolution of embedded systems in smart wearable devices: Design and implementation. SCCTS Journal of Embedded Systems Design and Applications, 2(1), 23–35.

13. Ali, M., & Bilal, A. (2025). Low-power wide area networks for IoT: Challenges, performance and future trends. Journal of Wireless Sensor Networks and IoT, 2(2), 20–25.
14. Silva, J. C. da, Souza, M. L. de O., & Almeida, A. de. (2025). Comparative analysis of programming models for reconfigurable hardware systems. SCCTS Transactions on Reconfigurable Computing, 2(1), 10–15.
15. Abdullah, D. (2024). Recent advancements in nanoengineering for biomedical applications: A comprehensive review. Innovative Reviews in Engineering and Science, 1(1), 1–5. https://doi.org/10.31838/INES/01.01.01