International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue:3 Volume No.10 May - June – 2026
DOI:10.46647/ijetms.2026.v10i03.024 ISSN: 2581-4621

AI-Augmented Metagenomic Diagnostic for Early Detection of Emerging Microbial Pathogens

V.Manjula[1], Divya C P[2], Shilpa C[3], G. Satheesh Kumar[4], A. Lakshmannath[5]
[1,2,3] *Assistant Professor, Dept. of CSE, K.S Institute of Technology, Banglore, Karnataka, India.*
[4,5] *Assistant Professor, Dept. of CSE, City Engineering College, Banglore, Karnataka, India.*

Abstract
The integration of artificial intelligence (AI) and mNGS is dismembering these walls, enabling briskly, further scalable outbreak response. This review synthesizes how AI- driven analytics are transubstantiating mNGS operations, from genome assembly to sequence bracket, using advanced infrastructures similar as convolutional neural networks, intermittent neural networks, and mills. Beyond accelerating workflows, AI's capacity for pattern recognition outperforms traditional homology- grounded styles, easing the discovery of new viral families and tracing retired transmission chains through anomaly discovery. nevertheless, critical challenges remain, including limited training data, the interpretability of AI models, and resource- ferocious computational demands that risk widening an "AI peak" in global health. We estimate these obstacles and highlight forward- looking strategies, including allied literacy for sequestration- conserving data sharing and resolvable AI for perfecting trust and natural sapience. Looking ahead, we fantasize an "AI-first" paradigm for outbreak preparedness, anchored in integrated "Digital Immune Systems" for nonstop, global- scale surveillance. By framing the community between mNGS and AI as a transformative vault, this review underscores its eventuality to strengthen adaptability against unborn afflictions.
Keywords: Artificial Intelligence, Pathogen Discovery, mNGS (metagenomic Next-Generation Sequencing).

1. INTRODUCTION

A critical determinant in mitigating the impact of these events is the speed of the public health response. The rapid and accurate identification of the causative pathogen is the essential first step for initiating effective containment, guiding therapeutic development, and deploying public health interventions (Yimer et al., 2024). To identify pathogens, public health laboratories employ a variety of testing methods. Traditional assays include microscopy, culture-based analyses, and immunoassays that detect either pathogen antigens or host immune responses (Miller et al., 2013; Roux et al., 2021).

The adoption of nucleic acid amplification tests (NAATs), such as (Polymerase chain reaction) PCR, marked a significant advancement in speed and sensitivity but remains inherently targeted (Miller et al., 2013; Khan et al., 2024). Despite the availability of conventional testing approaches, many samples submitted to public health laboratories during outbreaks remain undiagnosed, leaving critical questions unanswered and exposing the limitations of standard diagnostic methods. This diagnostic gap has positioned metagenomic next-generation sequencing (mNGS) as a pivotal frontier for novel viral discovery.

Viral metagenomic next-generation sequencing, which enables the analysis of DNA and/or RNA from a sample (Roux et al., 2021), has emerged as a powerful tool for pathogen detection (Khan et al., 2024). By comprehensively interrogating nucleic acids in clinical and environmental samples, mNGS can identify known and novel viruses without prior knowledge of a causative agent (Roux et al., 2021; Khan et al., 2024; Mokili et al., 2012). This unbiased nature makes it indispensable for investigating unknown outbreaks.

1.1 Methods

1.1.1 Information sources and search strategy

The search strategy employed a combination of keywords and Medical Subject Headings (MeSH) terms related to the core concepts. The primary search string was: ("viral metagenomic*" OR

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue:3 Volume No.10 May - June – 2026
DOI:10.46647/ijetms.2026.v10i03.024 ISSN: 2581-4621

"metagenomic next-generation sequencing" OR "mNGS") AND ("artificial intelligence" OR "machine learning" OR "deep learning") AND ("outbreak investigation" OR "pathogen discovery" OR "pandemic preparedness"). This string was adapted to the syntax requirements of each database.pt).

The Methods sections should be brief, but they should include sufficient technical information to allow the experiments to be repeated by a qualified reader. Only new methods should be described in detail. Cite previously published procedures in References.

1.1.2 Selection process

The selection process adhered to a two-stage screening protocol to ensure rigor and minimize bias. First, two independent reviewers screened the titles and abstracts of all retrieved records against the predefined eligibility criteria. Any discrepancies or conflicts regarding inclusion at this stage were resolved through consensus discussion or, when necessary, arbitration by a third senior reviewer. Second, the full-text articles of all records deemed potentially relevant during the initial screening were retrieved and subjected to a comprehensive eligibility assessment by the same two reviewers. Final inclusion decisions were made based on strict application of the criteria.

1.1.1 Data extraction and synthesis

Data from included studies were extracted into a standardized form, capturing information on study objectives, AI/ML methodologies, sequencing platforms, key findings, and identified challenges. Given the diverse and rapidly evolving nature of the field, a narrative synthesis approach was employed. Findings were thematically organized to construct a coherent overview of current applications, comparative advantages, persistent challenges, and future directions of AI-powered viral metagenomics.

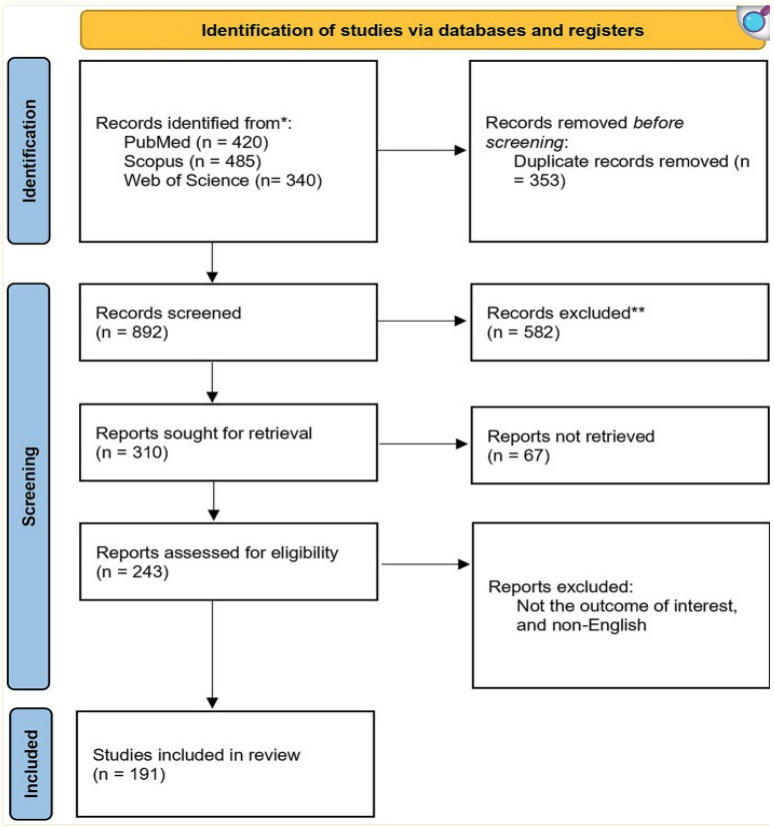

Figure 1. The results of this systematic selection process are detailed in the PRISMA-style flow diagram

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue:3 Volume No.10 May - June – 2026
DOI:10.46647/ijetms.2026.v10i03.024 ISSN: 2581-4621

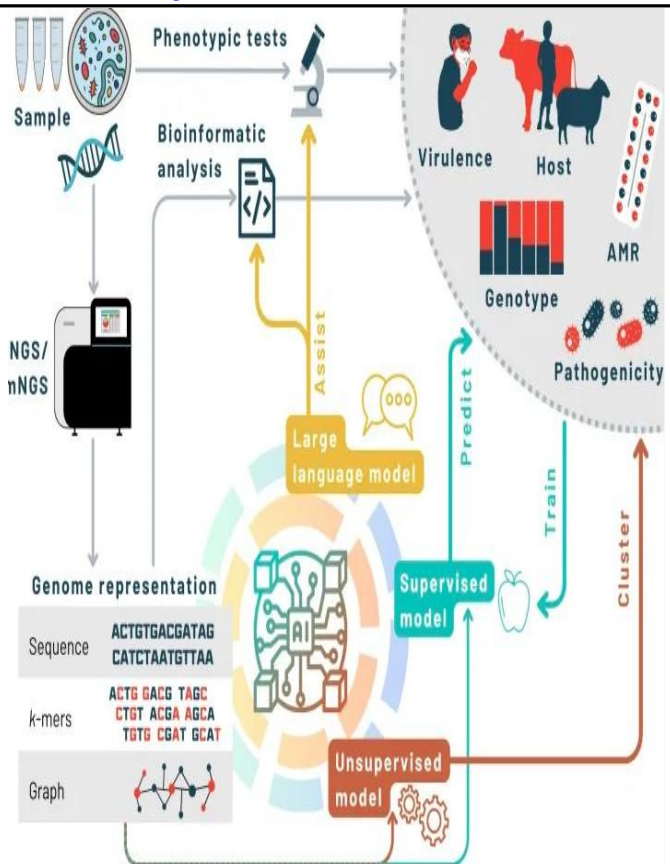

Figure 2. Emerging applications of AI in pathogen genomics.

2. RESULTS AND DISCUSSION

2.1 Metagenomic next generation sequencing technologies

Metagenomic next-generation sequencing platforms are now widely employed not only for targeted sequencing of specific genes or genomic regions but also for comprehensive, sequence-based association analyses that drive pathogen discovery and characterization (Harismendy et al., 2009). Their growing adoption is largely fueled by the urgent need for faster, more accurate, and versatile diagnostic tools in infectious disease management.

Beyond diagnostics, mNGS has opened entirely new avenues for research by allowing scientists to interrogate genetic information at an unprecedented scale and resolution, thereby advancing our understanding of microbial diversity, host–pathogen interactions, and evolutionary dynamics (Goodwin et al., 2016).

In contrast, the advent of third-generation platforms such as Oxford Nanopore Technologies (ONT) and Pacific Biosciences (PacBio) has revolutionized the field by enabling long-read sequencing, often spanning thousands of bases, with steadily improving accuracy (Han et al., 2024). While short reads (<300 bp) can lead to fragmented assemblies and may overlook structural variants, long reads provide the ability to resolve repetitive regions and capture full-length genes, thereby offering significant advantages in metagenomic applications (Han et al., 2024). Below we compare Illumina, ONT, and PacBio across key metrics for metagenomics (Table 1).

Table 1. Key artificial intelligence architectures and their applications in viral metagenomics.

Feature	Illumina	Oxford nanopore technologies	Pacific biosciences

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue:3 Volume No.10 May - June – 2026
DOI:10.46647/ijetms.2026.v10i03.024 ISSN: 2581-4621

Key technology	Sequencing -by-synthesis with reversible dye-terminators	Nanopore-based electronic sensing of DNA/RNA strands	Single Molecule, Real-Time (SMRT) sequencing with circular consensus sequencing (CCS)
Typical read length	50–300 bp (short)	10 kb– 100+ kb (ultra-long)	Long & Accurate (15–25 kb) (HiFi reads)
Through put and speed	Very High throughput (Terabases/ run). Run time: hours to days	Scalable & Real-Time. From portable (MinION) to high throughput (PromethION). Data is available immediately. Scalable (MinION: ~30 Gb; PromethION: ~13 Tb)	High (Revio: ~100–120 Gb per SMRT Cell)
Accuracy	Very High (>99.9%); substitution errors	Moderate-High (Raw: ~99%; with Q20+: >99.9%); indel errors in homopolymers	Extremely High (>99.9% with HiFi mode)
Best suited for mNGS application	Large-scale surveillance and detection: Sensitive pathogen identification in complex samples (e.g., microbiome, plasma). High-resolution SNP analysis for outbreak tracing.	Rapid Real-time pathogen detection and outbreak response in the field. Denovo assembly of unknown pathogens. Epigenetic modification detection	Gold-standard de novo genome assembly of novel viruses/bacteria. Resolving complex viral communities and strain variants. Full-length 16S/18S sequencing without amplification bias.

2.2 AI applications

The rapid expansion of metagenomic sequencing has generated unprecedented volumes of complex and heterogeneous data, necessitating advanced analytical frameworks beyond conventional bioinformatics (Pita-Galeana et al., 2025). AI, particularly ML and DL, has emerged as a transformative tool for extracting biologically meaningful insights from metagenomic datasets. Its applications span the entire analytical pipeline, from raw data preprocessing to functional inference and clinical translation.

One of the central challenges in viral metagenomics is accurate classification of sequences, especially when viral genomes exhibit high mutation rates or when reference databases are incomplete. Traditional bioinformatics pipelines rely on alignment-based methods (e.g., BLAST, Bowtie) or k-mer frequency approaches (Pita-Galeana et al., 2025; Wu et al., 2021). These approaches are limited when viral sequences diverge significantly from known references.

AI techniques have also been applied to biome source tracking: the task of identifying source microbial communities that contributed to a specimen (Zha et al., 2022). Rather than searching for fragments of genomes to explain a pathology, this task is concerned with grouping all of the data into likely communities of organisms. Shenhav et al. (2019) developed an unsupervised learning approach based on expectation maximization, which they used to predict the contribution of maternal microflora to infant microbiome, to identify evidence of food and soil contaminants in

DOI:10.46647/ijetms.2026.v10i03.024 ISSN: 2581-4621

longitudinal samples from a household, and to distinguish gut microbiota of critically ill patients from those of healthy adults. Deep learning approaches have been successfully applied to classify human microbiomes by the associated disease group with high accuracy and reduced computational requirements for prediction compared to existing approaches.

CONCLUSION

The integration of artificial intelligence with viral metagenomics marks a paradigm shift in outbreak response, moving us from reactive diagnostics to proactive pandemic preparedness. AI directly addresses the core bottleneck of mNGS—data complexity—by enabling rapid pathogen identification, novel virus discovery beyond traditional methods, and predictive modeling of outbreaks. While challenges of data scarcity, model interpretability, and equitable access remain, emerging solutions like explainable AI and federated learning provide a clear path forward. This powerful synergy is forging a new "AI-first" frontier in global health, paving the way for intelligent surveillance systems capable of defending against future viral threats.

REFERENCES

[1] Alwakeel M. M. (2025). AI-assisted real-time monitoring of infectious diseases in urban areas. Mathematics 13:1911. doi: 10.3390/math13121911.

[2] Alcolea-Medina A., Alder C., Snell L. B., Charalampous T., Aydin A., Nebbia G., et al. (2024). Unified metagenomic method for rapid detection of microorganisms in clinical samples. Commun. Med. 4:135. doi: 10.1038/s43856-024-00554-3.

[3] Aswini R., Saranya B., Gayathri K., Karthikeyan E. (2025). Revolutionizing infectious disease surveillance: multi-omics technologies and AI-driven integration. Decoding Infect Trans 3:100061. doi: 10.1016/j.dcit.2025.100061.

[4] Calvino G., Peconi C., Strafella C., Trastulli G., Megalizzi D., Andreucci S., et al. (2024). Federated learning: breaking down barriers in global genomic research. Genes 15:1650. doi: 10.3390/genes15121650.

[5] Cantón Cruz K. A., Durán Barrón M. A., Morales Lozada I. A., Mujica Sánchez M. A., Deloya Brito G. G., García Colín M. d. C., et al. (2025). Detection of SARS-CoV-2 using the Abbott™ PANBIO™ COVID-19 SELF-TEST rapid test in patients seen at INER. Biomedicine 13:1012. doi: 10.3390/biomedicines13051012.

[6] Chen Z., Grim C. J., Ramachandran P., Meng J. (2024). Advancing metagenome-assembled genome-based pathogen identification: unraveling the power of long-read assembly algorithms in Oxford nanopore sequencing. Microbiol. Spect. 12, e0011724–e0011724. doi: 10.1128/spectrum.00117-24.