

Artificial intelligence in modern virology: From prediction to experimental validation

Engku Nur Syafirah Engku Abd Rahman^{1*} and Nur Rina Alissa Razian²

¹ Department of Physiology, School of Medical Sciences, Universiti Sains Malaysia, Health Campus, 16150 Kubang Kerian, Kelantan Darul Naim. ² Department of Medical Microbiology and Parasitology, School of Medical Sciences, Universiti Sains Malaysia, Health Campus, 16150 Kubang Kerian, Kelantan Darul Naim

Abstract

Artificial intelligence (AI) is transforming virology, with applications spanning viral genomic surveillance, protein structure prediction, diagnostics, and antiviral drug discovery. The scale of contemporary viral datasets makes AI valuable for identifying patterns, generating hypotheses, and prioritizing experimental investigations. However, computational predictions should not be regarded as biological evidence. Model performance may be compromised by biased datasets, limited generalizability, inadequate external validation, and incomplete reporting, while predicted molecular interactions or antiviral activities may not translate into experimental effects. This Perspective argues that AI will have its greatest impact when integrated with rigorous biological validation rather than used as a substitute for experimental virology. Transparent reporting, representative datasets, independent evaluation, human oversight, and responsible governance will be essential for reliable application. The field should therefore move beyond prediction toward experimentally validated, reproducible, and biologically meaningful AI-assisted virological discovery.

Keywords: artificial intelligence, machine learning, virology, viral genomics, antiviral discovery, computational biology

1. Perspective

Virology is a rapidly evolving discipline at the intersection of fundamental biology, medicine, public health, and technological innovation. Viral emergence and re-emergence, genomic evolution, persistent infections, antiviral resistance, virus-associated cancers, zoonotic transmission, and the continuing development of novel diagnostic, preventive, and therapeutic approaches demonstrate the breadth and global importance of contemporary virology.

It is in this context that artificial intelligence (AI) has emerged as an increasingly important tool. The enormous diversity of viral genomes, rapid evolutionary dynamics, structural complexity of viral proteins, and growing availability of genomic, epidemiological, imaging, and clinical datasets create problems that are difficult to address using conventional approaches alone. AI and machine-learning methods can process complex datasets, identify patterns that may not be readily apparent to investigators, and generate predictions that accelerate hypothesis generation. Yet the growing enthusiasm surrounding AI in virology should be accompanied by a clear recognition of its limitations. The greatest value of AI will not come from replacing experimental virology, but from integrating computational prediction with biological validation.

*Correspondence:

Engku Nur Syafirah Engku Abd Rahman (engkunursyafirah@gmail.com)

Received: 10-08-2026, **Accepted:** 03-09-2026, **Published:** 28-09-2026

How to cite: Engku Abd Rahman ENS and Razian NRA (2026) Artificial intelligence in modern virology: From prediction to experimental validation. *Global Virology Reports*, 1(1): 3–6.

© The Author(s) (2026). This article is an open access article under the terms of the Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>)

Publisher's Note. IAA Academic Arena Ltd. remains neutral with regard to jurisdictional claims in published maps and institutional affiliations



2. From viral sequences to biological insight

One of the most promising applications of AI in virology is the analysis of viral genomic data. Large-scale sequencing has transformed surveillance of viral populations, allowing researchers to characterize viral diversity, identify emerging variants, and reconstruct transmission patterns. AI approaches can extend these capabilities by identifying sequence patterns associated with viral fitness, host adaptation, antigenic change, or other phenotypes of biological and epidemiological importance [1,2].

The emergence of protein language models represents a further development. By learning patterns within large collections of protein sequences, these models can provide information about evolutionary relationships and potentially predict functional consequences of mutations. Their integration into pathogen genomic surveillance could help prioritize viral variants for further investigation. Nevertheless, predictions derived from historical sequence data remain dependent on the quality, representativeness, and evolutionary context of the datasets used to train them. A model that performs well on previously observed viruses may not necessarily generalize to a novel viral lineage.

3. Structural prediction and the expanding viral protein landscape

AI-based protein structure prediction has created another important opportunity for virology. Tools such as AlphaFold have substantially expanded access to predicted three-dimensional structures and have been applied to viral proteins, protein–protein interactions, and structure-guided investigation of viral mechanisms [3]. Large-scale structural prediction has also revealed the remarkable diversity of viral proteins and provided potential functional clues for proteins that are difficult to characterize using sequence similarity alone [4].

For virologists, this creates opportunities ranging from molecular evolution and protein annotation to structural analysis and antiviral discovery. Predicted structures can facilitate hypothesis generation when experimental structures are unavailable and can help identify candidate interaction sites or molecular targets.

However, a predicted structure should not automatically be treated as an experimentally established structure. Viral proteins may undergo conformational changes, interact with host or viral partners, contain intrinsically disordered regions, or adopt context-dependent states that are difficult to capture computationally. Structural predictions are therefore most valuable when treated as hypotheses that can subsequently be tested using biochemical, biophysical, structural, or cellular approaches.

4. AI and antiviral discovery

Antiviral drug discovery is another area in which AI has considerable potential. Machine-learning approaches can assist target identification, virtual screening, drug repurposing, molecular optimization, and prediction of resistance-associated mutations [5,6]. Generative approaches may additionally facilitate the design of candidate molecules with desired physicochemical or biological properties.

Yet antiviral discovery also illustrates why enthusiasm for AI must be balanced with experimental discipline. Many computational studies terminate at docking scores, predicted binding affinities, or other in-silico endpoints. These outputs can be useful for prioritization but do not establish antiviral activity. AI-assisted drug discovery can be affected by small or heterogeneous datasets, assay dependence, limited interpretability, and insufficient prospective validation [6].

A more productive model is therefore an iterative computational–experimental cycle: AI generates or prioritizes hypotheses; experimental systems test those hypotheses; experimental results are incorporated into improved models; and the process is repeated. Such an approach places AI within the established scientific method rather than treating computational prediction as an endpoint.

5. Beyond prediction: AI-assisted diagnostics and surveillance

AI is also increasingly relevant to viral diagnostics and public health. Machine-learning models can support image interpretation, molecular diagnostic workflows, epidemiological modelling, outbreak detection, and genomic surveillance [1,7]. Integration of genomic, clinical, environmental, and epidemiological information could potentially provide earlier signals of emerging threats than any individual data stream can provide.

The challenge is ensuring that these systems remain reliable outside the environments in which they were developed. Differences in sequencing platforms, patient populations, viral prevalence, healthcare infrastructure, and data quality can substantially affect model performance. Consequently, external validation across geographical and epidemiological settings should be considered essential before AI systems are incorporated into consequential public health or clinical decisions.

6. The problem of data, bias, and reproducibility

AI systems are only as reliable as the data and assumptions underlying them. Viral datasets are not necessarily representative of global viral diversity. Sequencing and clinical data may be concentrated in countries and institutions with greater research capacity, while important populations and geographic regions remain underrepresented. Models trained on such datasets can consequently reproduce or amplify existing biases.

Reproducibility presents another challenge. Important details concerning training data, preprocessing, model architecture, hyperparameters, evaluation procedures, and software versions are sometimes incompletely reported. Transparent reporting and accessible benchmarking are therefore particularly important for AI-based virology research.

The field would benefit from well-curated, openly accessible datasets with standardized metadata, clearly defined benchmark tasks, independent test sets, and prospective evaluation. These resources would make it easier to distinguish genuine biological performance from model-specific artefacts or overfitting.

7. Responsible AI must become part of virology

The rapid development of AI also introduces ethical, security, and governance considerations. AI systems can produce inaccurate or misleading outputs, while excessive confidence in automated predictions may encourage inappropriate reliance on computational results. The World Health Organization has emphasized the importance of transparency, accountability, equity, human oversight, and responsible governance in the use of AI for health [8].

These considerations are particularly relevant to virology because computational technologies can have dual-use implications. As AI becomes more capable of analyzing viral sequences, predicting protein function, modelling mutations, or designing biological molecules, appropriate safeguards and responsible scientific practices become increasingly important.

8. A path forward

AI should therefore be viewed neither as a solution to every problem in virology nor as a threat to conventional experimental research. Its greatest contribution is likely to emerge from integration. The virologist of the future may increasingly work across experimental and computational boundaries, using AI to identify patterns, prioritize experiments, explore molecular hypotheses, and interpret complex datasets while retaining biological reasoning and experimental validation as essential components of the research process.

The next stage of AI-enabled virology should consequently move from prediction to validation. Models should be evaluated prospectively, computational findings should be tested experimentally, uncertainty should be reported explicitly, and performance should be assessed across diverse viral systems and populations. Such an approach can transform AI from an attractive computational tool into a reliable component of the virology research pipeline.

For a field confronting rapidly evolving viruses and increasingly complex biological data, the objective should not be to make virology more artificial. It should be to make virology more predictive, integrative, reproducible, and experimentally informed.

Data availability

All relevant data are contained in the paper.

Ethical approval

Not applicable.

Authors' contribution

ENSEAR: Conceptualization, Methodology, Investigation, Writing – Original Draft Preparation, Writing – Review & Editing, Supervision; NRAR: Methodology, Investigation, Writing – Review & Editing.

Acknowledgments

None.

Funding

None.

Competing interests

The authors declare that they have no competing interests.

Informed consent statement

Not applicable.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

The authors utilized ChatGPT (OpenAI) exclusively as a tool to enhance the language, flow, and grammatical accuracy of this manuscript. All AI-assisted revisions were carefully vetted and modified by the authors, who retain complete responsibility for the final text and the validity of the reported research.

References

- [1] Lytras S, Lamb D., Ito J, Grove J, Yuan K, Sato K, et al. Pathogen genomic surveillance and the AI revolution. *J Virol* 2025;99:e01601-24. <https://doi.org/10.1128/jvi.01601-24>
- [2] Alavi SMA, Borgonovo F, Pourghazi F, Tande AJ, Berbari EF, O'Horo JC. Practical applications of artificial intelligence in infectious disease surveillance and control: opportunities, challenges, and ethical considerations. *Clin Microbiol Infect* 2026 (In press) <https://doi.org/10.1016/j.cmi.2026.07.052>
- [3] Akdel M, Pires DEV, Pardo EP, Jänes J, Zalevsky AO, Mészáros B, et al. A structural biology community assessment of AlphaFold2 applications. *Nat Struct Mol Biol* 2022;29:1056–67. <https://doi.org/10.1038/s41594-022-00849-w>
- [4] Nomburg J, Doherty EE, Price N, Bellieny-Rabelo D, Zhu YK, Doudna JA. Birth of protein folds and functions in the virome. *Nature* 2024;633:710–7. <https://doi.org/10.1038/s41586-024-07809-y>
- [5] Izmailyan R, Matevosyan M, Khachatryan H, Shavina A, Gevorgyan S, Ghazaryan A, et al. Discovery of new antiviral agents through artificial intelligence: in vitro and in vivo results. *Antiviral Res* 2024;222:105818. <https://doi.org/10.1016/j.antiviral.2024.105818>
- [6] Harutyunyan V, Matevosyan M, Brancale A, Zakaryan H. About artificial intelligence adoption in antiviral drug discovery: thoughts after ICAR 39. *Antiviral Res* 2026;255:106538. <https://doi.org/10.1016/j.antiviral.2026.106538>
- [7] Borham A, Kamal LT, Chun S. Artificial intelligence in epidemic watch: revolutionizing infectious diseases surveillance. *Front Digit Health* 2025;7:1692617. <https://doi.org/10.3389/fdgth.2025.1692617>
- [8] WHO. Ethics and governance of artificial intelligence for health: guidance on large multi-modal models. 2025 <https://www.who.int/publications/i/item/9789240084759> (accessed July 24, 2026).